6-22-2018

The Effect of Uncertain and Weak Modal Words in 10-K Filings on Analyst Forecast Attributes

Myung Sub Kim
Myung Sub Kim, mkim013@fiu.edu

DOI: 10.25148/etd.FIDC006848

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

THE EFFECT OF UNCERTAIN AND WEAK MODAL WORDS IN 10-K FILINGS ON ANALYST FORECAST ATTRIBUTES

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BUSINESS ADMINISTRATION

by

Myung Sub Kim

2018

To: Dean Joanne Li
College of Business

This dissertation, written by Myung Sub Kim, and entitled The Effect of Uncertain and Weak Modal Words in 10-K Filings on Analyst Forecast Attributes, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Wen-Hsiu Chou

_____
Xiaochuan Huang

_____
Jonathan Milian

_____
Clark Wheatley

_____
Steve Lin, Major Professor

Date of Defense: June 22, 2018

The dissertation of Myung Sub Kim is approved.

_____
Dean Joanne Li
College of Business

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2018

DEDICATION

I dedicate this thesis to my parents, Sung-Il Kim and Soo-Ho Shin. Without their love, support, patience, and understanding, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

I thank God for His amazing grace and unchanging love to complete this degree. I wish to thank my wife, Okboon, whose love, support, endurance, and sacrifice enabled me to complete this journey. My two daughters, Lauren and Emily, bring us joy and happiness in times of difficulties and restlessness. I would like to thank the members of my committee for their time, support and guidance: Dr. Clark Wheatley, Dr. Jonathan Milian, Dr. Xiaochuan Huang, and Dr. Wen-Hsiu Chou. I wish to express my deepest gratitude to my advisor, Dr. Steve Lin, for his endless care and support, and excellent guidance through this process. He has been a great mentor and role model scholar during my Ph.D. studies. I would like to thank my professors in the School of Accounting at Florida International University for their encouragement and guidance throughout the last four years. Finally, I want to thank my fellow Ph.D. students in the School of Accounting at Florida International University for their comradeship and friendship.

ABSTRACT OF THE DISSERTATION

THE EFFECT OF UNCERTAIN AND WEAK MODAL WORDS IN 10-K FILINGS ON ANALYST FORECAST ATTRIBUTES

by

Myung Sub Kim

Florida International University, 2018

Miami, Florida

Professor Steve Lin, Major Professor

This study examines the determinants of the use of uncertain and weak modal words in 10-K filings and the effect of these words on analyst forecast attributes. I find that the use of uncertain and weak modal words in 10-K filings is positively (negatively) associated with firm size, volatility of business and operations (firm age and number of business segments). More importantly, after controlling for readability and management tone, I find that the use of uncertain and weak modal words in 10-K filings is associated with greater analyst following, lower forecast dispersion, greater forecast accuracy, and lower uncertainty in analysts' overall and common information environment. The results of this study provide more insights into why management uses uncertain and weak modal words in 10-K filings and how these words in 10-K filings affect analysts' behavior and their forecast outcomes.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

CHAPTER 1

INTRODUCTION

The SEC introduced the plain English[1] Rule 421(d) in 1998 to make financial disclosures easier to read and understand for investors, brokers, advisers, and others in the financial services industry. This rule encourages firms to use plain English not only in their prospectuses but also in all SEC documents and communication with shareholders (SEC 1998; Francis 2014; Loughran and McDonald 2014). This new rule has motivated many studies to investigate the effects of the readability of 10-K filings on the usefulness of financial disclosures and the behavior of primary users of these disclosures such as investors and analysts.

Many studies examine and find that disclosure quality significantly affects analyst following and analysts' forecast accuracy and dispersion. More specifically, disclosure quality is measured by various proxies including disclosure scores (Lang and Lundholm 1996; Healy, Hutton, and Palepu 1999), segment disclosures (Botosan and Harris 2000), intangible assets (Barth, Kasznik, and McNichols 2001), a firm's accounting choices (Hopkins, Houston, and Peters 2000; Bradshaw, Miller, and Serafeim 2008), tax law changes (Plumlee 2003), specific financial items that represent earnings quality (Hirst and Hopkins 1998; Hirst, Hopkins, and Wahlen 2004), global diversification (Duru and Reeb 2002), and readability (Li, 2008).

[1] *A Plain English Handbook* provided by SEC.gov explains how the issuer can prepare cleaner and more informative documents, and it defines plain English as creating a document that is visually and logically organized and easily understandable. For example, long sentences are fixed by short sentences; passive voice is fixed by active voice. Please find the handbook: https://www.sec.gov/pdf/handbook.pdf.

This dissertation aims to contribute to this line of research by examining the relationship between a firm's financial disclosure and some attributes of analysts' forecasts. The financial disclosure examined in this study is the use of uncertain and weak modal words in financial statements (i.e., 10-K filings). More specifically, this study investigates the impact of using vague language, such as uncertain and weak modal words, in financial statements on analyst behavior and forecast outcomes.

Both the SEC and influential media have often criticized the increasing use of complex language in 10-K filings (Schroeder 2002). Their main concern is that users of financial statements including investors (especially small investors) may not be able to fully understand complex financial information (SEC 1998; Schroeder 2002; Cox 2007). Given the increasing trend of complicated financial disclosures, it is important to examine whether financial analysts require greater knowledge and skills to interpret these complex disclosures or if they can understand and reflect the information contained in complex financial reports into their forecasts (Lehavy, Li, and Merkley 2011). Similar to Lehavy et al. (2011), this study examines the impact of the complexity of a firm's financial disclosures on the attributes of analysts' forecasts. Different from Lehavy et al. (2011) that investigates the impact of readability of financial statements on analysts' forecasts, this study examines the extent to which management's ambiguous language in 10-K filings impacts some attributes of analysts' forecasts after controlling for different measures of readability. More specifically, this study investigates the determinants of usage of ambiguous words in 10-K filings and the association between the use of uncertain and weak modal words in 10-K filings and analysts' subsequent forecast outcomes. To the best of my knowledge, this is the first study to examine this relation in the literature.

There are various ways to define and measure 'readability' in the literature, but in the context of financial disclosures, it is normally defined as the ability of primary users of financial information such as investors and analysts to assimilate valuation-relevant information from firm disclosures (Loughran and McDonald 2014). The Fog Index, measured by the average sentence length and percentage of complex words, is one of the most commonly used measures for readability in the accounting and finance literature (De Franco, Hope, Vyas, and Zhou 2015) although it is widely criticized for being poorly specified when used to evaluate financial documents. Li (2008) is the first study to apply the Fog index to examine the impact of the readability of 10-K filings on firm performance and earnings persistence. He finds that 10-K filings of firms with lower earnings have lower readability and 10-K filings with higher readability are more likely to have persistent positive earnings. Many studies have followed Li (2008) to examine the implications of annual report readability for investment efficiency (Biddle, Hilary and, Verdi 2009), investors' responses to information content of annual reports (You and Zhang 2009; Rennekamp, 2012), small versus large investors (Miller 2010), individual investors (Lawrence 2013), management forecasts (Guay, Samuels, and Talyor 2015), and credit rating and cost of debt capital (Bonsall and Miller 2014).

Despite the popularity of the Fog Index in the literature, there has been some criticism of how it identifies "complex' words. The Fog index defines words with more than two syllables as complex words, which has been criticized for being a poorly designed measure because many words like *corporation, agreement,* and *management* are common and reasonably comprehensive in the context of firms' business disclosures even though they have more than two syllables (Loughran and McDonal, 2014). Loughran and

McDonald (2014) argue that the file size of 10-K filings as a gross measure of 10-K readability is straightforward, easy to calculate, readily applicable to other readability research, and more importantly prone to fewer measurement errors. Consistent with these arguments, they find that the 10-K file size outperforms the Fog Index in explaining both unexpected earnings and analyst forecast dispersion. Many recent publications and working papers have used 10-K file size to measure 10-K readability (e.g., Bonsall and Miller 2014; Bratten, Gleason, Larocque, and Mills 2014; Li and Zhao 2014; Ertugrul, Lei, Qiu, and Wan 2017). Although 10-K file size is a reasonable proxy for information complexity of 10-K filings, it may not be suitable for shorter and unvaried disclosure such as press releases and earning conference calls (Loughran and McDonald 2016). Also, both the Fog index and 10-K file size measures do not consider other lexical features of 10-K filings, such as tone and choice of words, which may have significant effects on 10-K readability and be used by managers to strategically deliver valuation-relevant information to market participants. This study aims to further investigate this issue.

There is a growing body of textual analysis research examining the tone and sentiment of firms' written communication with investors (e.g., 10-K filings, earnings press releases, and investor message boards). However, Loughran and McDonald (2016) argue that a lot of textual analysis research (Davis, Piger, and Sedor 2012; Demers and Vega 2014; Frankel, Mayew, and Sun 2010) has focused on a simple positive and negative dichotomy of sentiment analysis. They also argue that positive sentiment appears to have lower power in these studies. Therefore, Loughran and McDonald (2016) suggest exploring other keywords like "*uncertain*," "*litigious*," "*strong modal*," and "*weak modal*" words to have additional means of parsing sentiment. One relevant example is to examine managers'

opportunistic word choice of uncertain and weak modal words to hide bad news during conference calls. However, this leads to questions as to when and why firms use these uncertain words, such as *approximate, contingent, uncertain,* and *indefinite*, and weak modal words, such as *might, could, possible,* and *maybe*, in their 10-K filings and conference calls. Management obfuscation hypothesis states that managers have incentives to obfuscate information when they experience or expect poor performance simply because more complex disclosures take a longer time to process, which will delay market reaction on poor performance (Bloomfield 2002). You and Zhang (2009) find that investors' underreaction to information contained in 10-Ks tends to be stronger for firms with more complex 10-Ks, measured by number of words. However, there is not much research on the incentive and the impacts of using vague language such as uncertain and weak modal words in 10-K filings. This study investigates this issue.

My first research question is, therefore, to investigate the determinants of the use of uncertain and weak modal words in 10-K filings. Li (2008) examines the determinants of annual report readability, measured by the Fog Index and length of the document. He finds that large firms, volatile business environment, firms with seasonal equity offering and merger-and-acquisition, and Delaware firms (firm age, firms with special items, and number of business segments) have less (more) readable 10-Ks. Lehavy et al. (2011) investigates the association of 10-K readability and analysts' forecast behaviors and find that less readable 10-Ks are positively associated with analyst following, forecast dispersion and error, and uncertainty in analysts' information environment.

Since no prior research has examined the determinants of use of uncertain and weak modal words, I follow both Li (2008) and Lehavy et al. (2011) to investigate the potential

determinants of use of uncertain and weak modal words. Using 37,442 observations from Loughran and McDonald's word lists, Compustat financial data, and SDC database for special event data, I predict and find that the use of uncertain and weak modal words in 10-K filings is positively associated with firm size, volatility of business and operations. I further predict and find that the use of uncertain and weak modal words is negatively associated with firm age and number of business segments.

The main research question is to examine the association between the use of uncertain and weak modal words in 10-K filings and some attributes of analysts' forecasts. To examine this relation, I use 25,673 observations from Loughran and McDonald's word lists, Compustat financial data, and I/B/E/S analysts' forecast data. First, I examine the relation between analyst following and the use of uncertain and weak modal words in 10-K filings. On the one hand, Lehavy et al. (2011) argue that more complex financial disclosures increase analyst following because increased cost of processing firm's disclosure increases the demand for analysts' services. Therefore, I expect a positive association between the use of uncertain and weak modal words in 10-K filings and analyst following. On the other hand, some prior literature argues that there are additional costs for analysts to cover firms with less readable written communication (Mikhail, Walther, and Willis 1999; Plumlee 2003; Hong and Kubik 2003). I expect that analysts may prefer less ambiguous 10-K filings to produce more reliable forecasts so that they can maintain their reputation. I expect that fewer analysts would pursue tasks of firms with more ambiguous words in 10-K filings. My finding is consistent with the prediction that increased use of uncertain and weak modal words in 10-Ks leads to more analyst following. I report a

positive and statistically significant relation between number of analysts and the use of ambiguous words in 10-Ks.

Second, I examine whether forecast dispersion is associated with the use of uncertain and weak modal words in 10-Ks. Syntactic complexity driven by uncertain and weak modal words in 10-K filings may lead to higher costs for analysts in processing and interpreting such disclosures. These words create the asymmetric distribution and interpretation of firm information among analysts, leading to more diverse explanations about the firm's disclosures among analysts who follow the same firm. Thus, I expect more dispersion in analysts' reports when 10-K filings contain more uncertain and weak modal words.

However, prior studies find that analyst earnings forecasts become more optimistic when the uncertainty in firms' information environment increases (Ackert and Athanassakos 1997; Das, Levine, and Sivaramakrishnan 1998; Huberts & Fuller 1995; Lim 2001). Prior literature argues that analysts are more likely to report their earnings forecast with optimistic bias because they want to maintain a good relationship with clients, and this phenomenon occurs more often when uncertainty in clients' information environment increases. I, therefore, expect that more uncertain and weak modal words in 10-K filings increase the uncertainty in firm's information environment; thus, it may cause analysts' forecasts to be consistently upwards and decrease earnings forecast dispersion among analysts. Moreover, analysts may make extra efforts to produce more accurate forecasts when facing more uncertain information environment. My finding is consistent with the idea that analyst forecast dispersion is negatively related to uncertain and weak modal words in 10-K filings, which supports the management-relations and analyst effort

hypotheses discussed in the hypotheses development section, indicating financial analysts collectively bias their forecasts upward or make extra efforts to produce earnings forecasts when facing more uncertain information environment.

Third, I examine the relation between forecast accuracy[2] (forecast error) and the use of uncertain and weak modal words in 10-Ks. Lehavy et al. (2011) report that less readable 10-Ks are associated with less accurate forecasts due to increased costs for research and information-processing. I predict ambiguous language in 10-K filings increases costs for research and information-processing and these costs may decrease the accuracy of the forecast. Thus, the use of more uncertain and weak modal words in firms' disclosures may decrease the accuracy of analyst forecasts.

However, there are two theories that can explain the positive association between uncertain and weak modal words in 10-K filings and accurate forecasts. First, analysts may make extra efforts to produce more accurate earnings forecasts. Second, managers may use uncertain and weak modal words to signal analysts and other market participants about the uncertainty of future earnings. My finding is consistent with the idea that forecast accuracy (forecast error) is negatively related to uncertain and weak modal words in 10-K filings which suggests that analysts effectively process 10-Ks with ambiguous words, and are able to produce more accurate forecasts. It is possible that analysts make extra efforts to produce their forecasts when 10-K filings contain more uncertain and weak modal words. Future research needs to use behavioral research methods, such as interviews with analysts, to enhance our understanding of analysts' behavior.

[2] Forecast accuracy is defined as the squared value of the difference between the reported earnings in I/B/E/S and the most recent analyst consensus forecast, scaled by stock price 90 days before the consensus earnings forecast.

Fourth, Barron, Byard, Kim, and Stevens (1998) measure uncertainty in private and common (publicly available) information environment using equations with the accuracy, the dispersion, and the number of the analyst.[3] Employing the Barron et al. (1998)'s measure, Lehavy et al. (2011) measure uncertainty in analysts' overall information environment by the sum of common and idiosyncratic (private) uncertainty among analysts. Lehavy et al. (2011) predict that analyst forecasts for firms with less readable reports will be associated with greater overall uncertainty. Thus, I predict that there will be higher uncertainty in the analysts' overall information environment when firms use more ambiguous words in their 10-K filings. Inconsistent with my prediction, I find that uncertainty in analysts' overall information environment is decreasing in uncertain and weak modal words in 10-Ks, suggesting that managers may use ambiguous language in 10-Ks to signal analysts and other market participants about the uncertainty of future earnings.

Finally, following Barron et al. (1998)'s measure, Lehavy et al. (2011) measure common analyst forecast uncertainty by the ratio of common uncertainty to total uncertainty among analysts. In short, it measures the degree to which analysts share a common belief. Lehavy et al. (2011) find a positive relation between common analyst forecast uncertainty and the Fog Index, suggesting publicly available information such as the 10-K becomes more important to analysts relative to private (idiosyncratic) information with more complex 10-K filings. Similarly, I predict that more ambiguous words in 10-K filings increase the degree to which analysts share a common belief such as 10-K filings. This is because more uncertain and weak modal words in 10-Ks make it difficult for

[3] The measures and equations in Barron et al. (1998) will be discussed in research design.

analysts to process private information; thus, they may increase the importance of publicly available information such as 10-Ks. Inconsistent with my prediction, I find that uncertainty in analysts' common information environment decreases for firms with increased use of uncertain and weak modal words in firms' 10-K filings. Again, managers may use ambiguous language in 10-Ks to signal analysts and other market participants about the uncertainty of future earnings.

For a robustness test, I replace *File_Size* with the Fog Index (*Fog*) to examine if the effect of uncertain and weak modal words in 10-Ks hold for analyst forecast attributes with a different measure of readability (*Fog*). I employ the Fog Index data from Feng Li's website[4] for the sample period of 2000 – 2011. The evidence indicates that the Fog replacement does not change the overall results. Interestingly, I also find that the effect of the use of ambiguous language in 10-Ks on analysts' forecast attributes becomes stronger when *Fog* replaces *File_Size* (except *Analyst Following*). One limitation of this additional analysis is inconsistent sample period[5]. Future studies will match the sample period.

This study differs from Li (2008) and Lehavy et al. (2011) in three ways. First, I examine the determinants of the use of uncertain and weak modal words while Li (2008) examine the determinants of readability. Second, I examine the effect of the use of uncertain and weak modal words in 10-K filings on analyst forecast attributes while Lehavy et al. (2011) examine the association between annual report readability and analyst forecast attributes. Finally, this study examines whether the use of uncertain and weak modal words in 10-Ks provides significant incremental valuation-relevant information for financial

[4] Please check the data from http://webuser.bus.umich.edu/feng/.

[5] Main test sample period is 2000-2016.

analysts beyond 10-K readability, measured by the Fog index and 10-K file size, and tone management.

Main contributions of this study are threefold. First, this is the first study to examine the characteristics of firms that use uncertain and weak modal words in their 10-K documents. Therefore, this study provides insight into firms' opportunistic word choices in their 10-K filings under firm-specific financial conditions. Second, this is the first study to examine how financial analysts perceive information contained in uncertain and weak modal words in 10-K filings and reflect this information into their decision making. Third, this study is relevant to users of financial disclosures because it provides insight into how vague words in a clients' 10-K impact analyst behavior and forecasts. This is especially important given analysts play a key role as intermediaries between accounting information and investors. Hence, it is important to investigate the extent to which financial analysts interpret the information contained in uncertain and weak modal words and reflect this information into their efforts and research outputs that would directly affect the decision making of market participants.

This dissertation is organized as follows. Chapter 2 provides a summary of the literature review. Chapter 3 discusses hypothesis development. Chapter 4 describes the research design. Chapter 5 describes the data and sample. Chapter 6 presents the empirical results, and Chapter 7 concludes.

CHAPTER 2

LITERATURE REVIEW

This study is related to three areas of accounting and finance research. They are the role of readability, tone management, and the use of uncertain and weak modal words (ambiguous language) in market-based accounting and finance research.

2.1 Readability of Financial Disclosures

The first area of research examines the association between annual report/analyst report readability, mostly measured by the Fog index, and earnings quality/analyst forecast attributes. A seminal paper by Li (2008) examines the impact of annual report readability on firm performance and earnings persistence. He finds that annual reports of firms with lower earnings have lower readability while annual reports with higher readability are more likely to have persistent positive earnings. Many finance and accounting studies use the Fog Index to investigate the readability of annual reports in relation to earnings persistence (Li 2008), investment efficiency (Biddle et al. 2009), and timeliness of price adjustment (Callen, Khan, and Lu 2009). However, only a few studies examine the implications of annual report readability for analyst behavior and the readability of analyst reports. For example, Lehavy et al. (2011) examine the association between readability of 10-Ks and financial analysts' behavior and their financial reports' attributes. Using 10-K filings from SEC's EDGAR database for 1995-2006, they find that analyst reports of firms with less readable 10-K filings are linked with more analyst following and are more informative. Moreover, analyst reports of firms with more complex 10-K filings are associated with

higher forecast dispersion, less accurate analysts' forecasts, and greater overall uncertainty. De Franco et al. (2015) use a large text database with analysts' reports from 2002 to 2009 and examine the importance of the readability of analyst reports. They also find that "high-ability analysts[6]" produce more readable reports and that trading volume reaction increases with more readable analyst reports.

However, many studies argue that the widely used the Fog index, measured by the number of complex words (measured by the syllables) and the average length of sentences, is poorly specified when applied to business documents (e.g., Jones and Shoemaker 1994; Loughran and McDonald 2014, 2016). Also, Loughran and McDonald (2014, 2016) point out that the Fog index does not consider writing style. For example, the value of the Fog index will be identical even though we randomly reorder words in the original sentence to make it incomprehensive. Loughran and McDonald (2014) provide evidence that 10-K file size, as a comprehensive proxy for 10-K readability, appears to outperform the Fog Index when explaining unexpected earnings and analyst forecast dispersion. Hence, it has been increasingly popular to use 10-K file size for measuring annual report readability in accounting and finance research. For example, Ertugrul et al. (2017) find that larger 10-K file size (i.e., a less readable 10-K) decreases loan maturity and increases the probability of collateral requirement. This shows that banks increase their level of monitoring severity for firms with less readable annual reports. This is consistent with the argument made by Loughran and McDonald (2014) that low readability is related to firms' intention to obfuscate mandated earnings-relevant information by burying it in longer documents. In

[6] De Franco et al. (2015) define that "high-ability analysts" are more experienced, issue more timely earnings forecasts and more frequent forecast revisions. They are also better ranked by Institutional Investor magazine, and issue more consistent earnings forecasts and recommendations.

sum, previous studies provide evidence indicating that readability is related to the informativeness of annual reports, which affects market participants' behavior although there is a debate on how readability should be measured.

2.2 Management Tone in Financial Disclosures

Another relevant research area is to investigate the tone (positive and negative) of the words in the firm's disclosures (e.g., earnings announcements, 10-K, 10-Q, MD&A, and conference calls) and how investors react to tone management. Loughran and McDonald (2011) develop positive and negative words lists. For example, the positive words in the list are *achieve, attain, efficient, improve, profitable*, and *upturn*; negative words in the list are *loss, impairment, against, decline, negatives, restated, litigation,* and *misstatement*. Some studies find that managers have incentives to manage their optimistic or pessimistic language at earnings announcements especially for information that is more sensitive to stock prices (e.g. Feldman, Govindaraj, Livnat, and Segal 2009; Davis and Tama-Sweet 2012; Davis et al. 2012; Huang, Teoh, and Zhang 2014; Dermers and Vega 2014). For example, Feldman et al. (2009) find management's tone change in the MD&A section of 10-Q and 10-K is significantly related to short-window market reactions and excess returns drift around the SEC filing. This study shows that investors seem to consider this nonfinancial information (e.g., the change of management tone) in addition to quantitative financial information from MD&A disclosures when making their investment decisions. Davis and Tama-Sweet (2012) find that fast-growing firms and firms that precisely meet or just beat analysts' earnings forecasts report less pessimistic language in their earnings press releases because managers have a greater incentive to report

strategically when disclosed information is more sensitive to the stock returns. Using 23,000 quarterly earnings press releases between 1998 and 2003, Davis et al. (2012) measure net optimistic language using DICTION software program (counts optimistic and pessimistic words). Their evidence suggests that the managers' use of net optimistic language is positively related to expected future firm performance (e.g., higher ROA) and a significant market reaction. Li (2008) suggests that conference calls may provide a better platform to investigate the relation between linguistic information content and firm performance because unscripted question and answer sessions may provide a full examination of firm disclosure between managers and analysts. Many studies examine and find that the tone of the words used during conference calls affects conference call return and trading volume (e.g., Price, Doran, Peterson, and Bliss 2012; Davis, Matsumoto, and Zhang 2015; Druz, Wagner, and Zeckhauser 2016; Milian and Smith 2017). Price et al. (2012) employ computer-based content analysis to examine the question and answer session in a conference call and find that conference call discussion tone has highly significant explanatory power for initial reaction CARs (Cumulative Abnormal Returns), the post-earnings-announcement-drift, and abnormal trading volume. Also, if uncertainty exists in the firm's cash dividend payout decision, the tone of conference call has more explanatory power for abnormal returns and trading volume. Using the effect of manager-specific factors such as previous career experience (e.g., charitable organization involvement, etc.) and education backgrounds (e.g., MBA, LAW degree, etc.), Davis et al. (2015) find the tone of conference calls is significantly associated with manager-specific factors. Also, they find some evidence that the choice of managers' language influences the investors' interpretation of the firm's financial performance. Milian and Smith (2017)

examine 16,609 conference calls of S&P 500 companies and find that the amount of praise by analysts on earnings conference calls is positively related to the earnings surprise and a more significant extent the earnings announcement stock return. These studies, however, are criticized for using an overly simplified positive (optimistic) and negative (pessimistic) dichotomy of sentiment analysis. Loughran and McDonald (2016) argue that tests for positive sentiment appear to have a lower test power because positive sentiment becomes ambiguous when negative information is filled with positive words.[7] Moreover, these studies do not consider how the tone used in the earnings press releases affects analyst forecast attributes. This study differs from the above studies in that this study examines how the tone (negative and positive) in 10-K filings influences analyst's behavior and forecast outcomes. Overall, studies on tone management find that the tone of the language chosen by management in the earnings press releases is informative for investors although there are some measurement issues with interpreting positive words in financial disclosures.

2.3 Uncertain and Weak Modal Words in Financial Disclosures

The third related accounting and finance research area is the use of uncertain and weak modal words in financial disclosures. In addition to readability and tone, ambiguous language in annual reports can be a source of firm risk because it may increase (decrease) informational risk (investors' ability to comprehend financial reports). Loughran and

[7] Loughran and McDonald (2016) show an example case, "a careful manager might use 90% positive words in dismissing an employee." which explains one critical issue with positive and negative tone textual analysis. The manager intended to dismiss the employee but textual analysis will capture more positive words than negative words in the manager's comment.

McDonald (2011) create word lists to reflect ambiguous words in the financial context. For example, the list of uncertain words such as *approximate, assume, contingent, depend, and indefinite*, expresses imprecision; the list of weak modal words indicates a lack of confidence, and examples are *might, could, maybe, depending, and possible*. Recent studies find that ambiguous texts of corporate disclosures affect valuation uncertainty. For example, Loughran and McDonald (2011) find a positive relation between the use of uncertain and weak modal words in 10-K filings and the stock return volatility. They perform event studies for report excess returns, abnormal volume, and post-event return volatility and find a significantly negative relation between event period excess returns and the use of uncertain and weak modal words in 10-K filings. Moreover, they find that the use of uncertain and weak modal words in 10-K filings is positively associated with the abnormal trading volume during the event window. There is a positive relation between stock return volatility in the year after 10-K filings and a proportion of uncertain and weak modal words in 10-K filings. Their paper documents that ambiguous words are linked with market reactions around the 10-K filings, abnormal trading volume, and stock return volatility after 10-K filings. Loughran and McDonald (2013) find a positive link between the ambiguous language in S-1 initial public offering (IPO) filings and first-day returns, absolute price revisions, and subsequence volatility. Using a sample of 1,887 completed U.S. IPOs during 1997-2010, they find associations between uncertain words in the Form S-1 and first-day returns, offer price revisions, and volatility. Unlike prior literature which measures firm age, sales, and IPO gross proceeds as ex-ante uncertainty proxies, they use S-1's tone as a direct proxy for measuring ex-ante uncertainty about an IPO's valuation. Ertugrul et al. (2017) find that more ambiguous words in annual reports are related to

stricter loan contracts and higher stock price crash risk. They find that the frequencies of uncertain and weak modal words in annual reports are positively associated with stricter loan contracts. This finding indicates that the ambiguous words of 10-K filings contain relevant information in assessing firm's risk level and have an effect on both price and nonprice loan terms. Also, they find that high frequency of uncertain and weak modal words in financial disclosures increases the likelihood of a stock price crash risk. This finding is consistent with their prediction that 10-K filings with more ambiguous language are related to higher cost of capital. Overall, their results provide significant evidence that the readability and ambiguous language of 10-K filings are associated with a firm's information-concealing behavior that increases its information risk and cost of capital. Moreover, Loughran and McDonald (2016) suggest, uncertain and weak modal words used in annual reports and earnings press releases are additional means of parsing sentiment, which warrants an excellent future area of research. This study aims to examine the effect of uncertain and weak modal words in 10-K filings on financial analysts' behavior and analyst forecast outcomes.

CHAPTER 3

HYPOTHESIS DEVELOPMENT

3.1 Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

Li (2008) examines the determinants of annual report readability, measured by the Fog Index and the length of the document. He regresses the Fog Index and the length of annual reports on potential determinants such as firm size, market-to-book, firm age, special items, volatility of business and operations, the complexity of operations, financial complexity, firm events, and incorporation state. Li (2008) finds that larger firms, firms with more volatile business, firms with merger-and-acquisition (M&A) transactions, and firms incorporated in Delaware state are positively related to the Fog Index (i.e., listed items are associated with less readable 10-K reports). However, his finding suggests that firm age, firms with special items, firms with geographic segments, and firms that are issuing new equity are negatively associated with the Fog Index (i.e., listed items are associated with more readable 10-K reports).

This study examines the determinants of the use of uncertain and weak modal words in 10-K filings. To my knowledge, there is no prior study that examines this issue. Thus, I follow Li (2008)'s method to examine the determinants of the use of these words in 10-K filings. First, I predict that the use of uncertain and weak modal words in 10-K filings is positively associated with firm size, market-to-book ratio, volatility of business and operations. I expect that larger firms normally face more complex and uncertain business environment and therefore are more likely to use uncertain and weak modal words in their 10-Ks. I expect that growth firms (i.e., firms with the higher market-to-book) may also face

a more uncertain business environment and therefore their financial reports are likely to include more ambiguous words. I also predict that firms facing a volatile business operating environment tend to use more uncertain and weak modal words in their financial disclosures. Second, I predict that the use of uncertain and weak modal words in 10-Ks is negatively associated with firm age, special items, and firm events such as merger-and-acquisition (M&A) and seasoned equity offering (SEO). I predict that older firms may have less information asymmetry and information uncertainty. I also expect firms with more negative special items are more likely to use uncertain and weak modal words in their 10-K filings. I expect less ambiguous words in annual reports if a firm expects M&A or SEO near future. Following Li (2008), I include a Delaware dummy to examine whether firms incorporated in Delaware state use more uncertain and weak modal words in their 10-K filings because Daines (2001) argues that firms in Delaware follow different laws and regulations from similar firms in other states.

3.2 The Relation between Uncertain and Weak Modal Words in 10-K Filings and Analyst Forecast Attributes

3.2.1 Analyst Following

I have no prediction between the use of uncertain and weak modal words in 10-K filings and analyst following because prior literature provides rather mixed evidence. Some literature argues that less readable (more complex) financial disclosures lead to increased analyst following (e.g., Lehavy et al. 2011). In general, the cost of processing complex accounting information is higher for users of financial statements. However, users of financial information have different levels of abilities to interpret and process the

information (Indjejikian 1991; Ball 1992). Therefore, financial analysts with their private analysis of firms can make a profit from these differences (Schipper 1991). The greater the cost to users of processing firm's disclosure, the more profit the analysts make. If analysts intend to provide their services to meet this increased demand for processing more complicated financial disclosures, then firms with complicated financial disclosures will have more analyst following (Lehavy et al. 2011). Therefore, I expect a positive association between the use of uncertain and weak modal words in 10-K filings and analyst following.

However, there are additional costs for analysts to cover firms with less readable written communication; such costs include the direct costs of processing information provided by management, research costs, and the adverse impact on analysts' reputation from inaccurate forecasts and recommendations (Mikhail et al. 1999; Plumlee 2003; Hong and Kubik 2003). Prior literature also finds that the cost of potential manipulations attributed to the less readable written communication may discourage analyst following (Li 2008; Lang, Lins, and Miller 2004). Only "high-ability" analysts issue more consistent earnings forecast and produce more readable forecast reports (De Franco et al. 2015). Thus, I expect that analysts may prefer to work with more transparent and less ambiguous 10-K filings to produce more reliable earnings forecasts and maintain their reputation. I expect that fewer analysts would pursue tasks of firms with more ambiguous words in 10-K filings.

Due to the mixed results from the prior literature, I predict (in a null hypothesis form),

H1: ***There is no association between the use of uncertain and weak modal words in firm's 10-K filings and analyst following.***

3.2.2 Forecast Dispersion and Accuracy

I also examine how the use of the uncertain and weak modal words in 10-K filings affects analyst earnings forecast dispersion and accuracy as well as uncertainty in firm idiosyncratic (private) or common (public) information environments. Syntactic complexity driven by uncertain and weak modal words in 10-K filings may lead to higher costs for analysts in processing and interpreting such disclosures. Moreover, it influences the dispersion and accuracy of earnings forecasts and the uncertainty in the information environment. Less readable written communication increases analysts' information processing cost. As a result, analysts are more likely to have more diverse perceptions and interpretations of firm disclosures (Shipper 1991; Mikhail et al. 1999; Plumlee 2003; Hong and Kubik 2003). Lehavy et al. (2011) find that less readable 10-K filings are associated with higher analyst forecast dispersion. Since uncertain and weak modal words are ambiguous, the use of these words could increase analyst forecast dispersion. These words create the asymmetric distribution and interpretation of firm information among analysts, leading to more diverse explanations about the firm's disclosures among analysts who follow the same firm. Thus, I expect more dispersion in analysts' reports when 10-K filings contain more uncertain and weak modal words. However, prior studies find that analyst earnings forecasts become more optimistic when the uncertainty in the information environment increases (Ackert and Athanassakos 1997; Das et al. 1998; Huberts and Fuller 1995; Lim 2001). These findings are explained by two theories. The first is the management-relations hypothesis. Lim (2001) finds that analysts may report their earnings forecasts with optimistic bias by expecting favorable treatments and better private information from the client firms. He expects that this action becomes stronger when the

uncertainty in information environment increases. The other theory focuses on the reputational concerns of individual analysts. Ackert and Athanassakos (1997) suggest that reputational concerns related to optimistic opinion will be smaller when the uncertainty in firm's information environment is higher. This is because optimistic earnings forecast is scrutinized more easily when the uncertainty in firm's disclosures is low and with minute differences among analyst forecasts. Both theories predict that analysts' forecasts may collectively become more optimistic when firms' information environment becomes more uncertain, which may reduce forecast dispersion among analysts. Two other theories can also explain a negative association between the use of uncertain and weak modal words in 10-Ks and forecast dispersion. First, analysts may make extra efforts to produce more accurate earnings forecasts, which in turn reduces forecast dispersion (Chen and Matsumoto 2006; Bradley, Gokkaya, and Liu 2017). Second, managers may use uncertain and weak modal words in 10-Ks to signal analysts and market participants about the uncertainty of future earnings.

Due to the mixed prediction from prior literature, I predict (in a null hypothesis form),

H2: ***There is no association between the use of uncertain and weak modal words in firm's 10-K filings and the dispersion in analyst earnings forecasts.***

Lehavy et al. (2011) predict and find that less readable 10-K filings are associated with less accurate analyst earnings forecasts. They argue that less readable financial disclosures increase costs for research and information-processing, which may decrease forecast accuracy. If the ambiguous language in 10-K filings increases these costs, then I predict that the use of more uncertain and weak modal words in a firm's disclosures may

decrease the accuracy of analyst consensus forecast after 10-K filings. However, at least two theories that can explain the positive association between uncertain and weak words in 10-K filings and accurate analyst earnings forecasts. First, analysts may make extra efforts to produce more accurate earnings forecasts when facing more uncertain information environment, which in turn reduces forecast error. Second, managers may use uncertain and weak modal words in 10-Ks to signal analysts and market participants about the uncertainty of future earnings, which in turn increases accuracy in analyst earnings forecasts.

Due to the mixed prediction, I predict (in a null hypothesis form),

H3: ***There is no association between the use of uncertain and weak modal words in firm's 10-K filings and accuracy in analyst earnings forecasts.***

3.2.3 Uncertainty in Analysts' Overall and Common Information Environment

Barron et al. (1998) measure uncertainty in private and common (publicly available) information environment using equations with the accuracy, the dispersion, and the number of analysts.[8] Employing Barron et al. (1998)'s measure, Lehavy et al. (2011) measure uncertainty in analysts' overall information environment by the sum of common and idiosyncratic (private) uncertainty among analysts. Lehavy et al. (2011) predict that analyst forecasts for firms with less readable reports will be associated with greater overall uncertainty. There will be higher uncertainty in the analysts' overall information environment when firms use more ambiguous words in their 10-K filings.

Therefore, I predict (in an alternative form),

[8] The measures and equations in Barron et al. (1998) will be discussed in research design.

H4: ***The use of uncertain and weak modal words in firm's 10-K filings may increase in the uncertainty associated with analysts' overall information environment.***

Furthermore, Lehavy et al. (2011) find uncertainty in analysts' common (public) information environment is increasing in less readable 10-K filings. Following Barron et al. (1998)'s measure, they measure common analyst forecast uncertainty by the proportion of common uncertainty to total uncertainty among analysts (i.e., the ratio of uncertainty in analyst' common information environment to uncertainty in analysts' overall information environment). In other words, it measures the degree to which analysts share a common belief. Lehavy et al. (2011) find a positive relation between common analyst forecast uncertainty and the Fog Index, suggesting publicly available information such as the 10-K becomes more important to analysts relative to private information with more complex 10-K filings. Similarly, I predict that more ambiguous words in 10-K filings may increase the degree to which analysts share a common belief such as 10-K filings. This is because more uncertain and weak modal words in 10-Ks make it difficult for analysts to process private information; thus, they increase the importance of publicly available information such as 10-Ks. Therefore, I predict (in an alternative form),

H5: ***The use of uncertain and weak modal words in firm's 10-K filings may increase in the uncertainty associated with information common to all analysts.***

CHAPTER 4

RESEARCH DESIGN

4.1 Determinants of Uncertain and Weak Modal Words in 10-K Filings

I employ Li (2008)'s model to examine the determinants of the use of uncertain and weak modal words in 10-K filings as follows.

$$Uncertain_{i,t}/Weak\_Modal_{i,t} = \beta_0 + \beta_1 Firm\_Size_{i,t} + \beta_2 MTB_{i,t} + \beta_3 Firm\_Age_{i,t} + \beta_4 Special\_Items_{i,t} + \beta_5 Std\_Ret_{i,t} + \beta_6 Std\_Earn_{i,t} + \beta_7 Segments_{i,t} + \beta_8 M\&A_{i,t} + \beta_9 SEO_{i,t} + \beta_{10} Delaware_{i,t} + \varepsilon_{i,t} \quad (1)$$

In the above model, I have two dependent variables - *Uncertain* and *Weak_Modal*. *Uncertain* (*Weak_Modal*) is the proportion (percentage) of uncertain (weak modal) words to the total words in 10-K filings as defined in Loughran and McDonald (2011). *Firm_Size* is a proxy for firm's operational and business environment. It is defined as the logarithm of the market value of equity at the fiscal year-end and included to explain how the size of a firm influences the use of uncertain and weak modal words in firms' 10-K filings. I expect larger firms to use more ambiguous language in their financial disclosures because they normally face more uncertain and complex business environment. *MTB* (market-to-book) is a proxy for potential growth and investment opportunities of firms. It is the ratio of the market value of equity plus book value of liability to the book value of total assets at the fiscal year-end. I expect high *MTB* firms to use more uncertain and weak modal words in their 10-K filings than do low *MTB* firms. Older firms are more familiar to users and have

less information asymmetry. Thus I expect these firms to use less ambiguous language in 10-K filings. *Firm_Age* in model (1) is the number of years a firm appears in the CRSP monthly stock return database. *Special_Items* is the amount of special items divided by the book value of assets. I predict that firms with more negative special items[9] probably experience more unusual events which may lead to more ambiguous word usage in their financial disclosures. I measure firm-specific stock return and earnings volatility for business or operation volatility. *Std_Ret* is the standard deviation of the monthly stock returns in the prior year and *Std_Earn* is the standard deviation of the operating earnings during the past five fiscal years[10]. *Segments* is a proxy for complexity of operations using the logarithm of the number of business segments. I create two dummy variables to measure special firm events such as merger-and-acquisition (*M&A*) and Seasoned Equity Offering (*SEO*). I use the SDC Platinum M&A database for the *M&A* and the SDC Global New Issues database for the *SEO*. *M&A* equals 1 for firms that engage in M&A as an acquirer in a specific firm-year and 0 otherwise. *SEO* equals 1 for firms that have the seasoned equity offering in a specific firm-year and 0 otherwise. *Delaware* is a dummy variable equal to 1 if firms are incorporated in the state of Delaware and 0 otherwise. According to Daines (2001), firms that are incorporated in Delaware follow different corporate laws and investor protections, have more takeover bids, and are valued higher than similar firms in other states. Thus, I include this dummy to examine if firms in Delaware have more ambiguous language in their 10-K filings.

[9] Special items are more likely to be negative than positive due to accounting standards.

[10] I only include observations that have at least 3 years data.

4.2 The Effect of the Use of Uncertain and Weak Modal Words in 10-K Filings on Analyst Forecast Attributes

H1 predicts the relation between analyst following and the use of uncertain and weak modal words in 10-K filings. I estimate the following regression for H1.

$$\begin{aligned} \#Analysts_{i,t} = {} & \beta_0 + \beta_1 Uncertain_{i,t} + \beta_2 Weak\_Modal_{i,t} + \beta_3 File\_Size_{i,t} \\ & + \beta_4 Pos\_Tone_{i,t} + \beta_5 Neg\_Tone_{i,t} + \beta_6 Firm\_Size_{i,t} + \beta_7 Growth_{i,t} \\ & + \beta_8 Segments_{i,t} + \beta_9 \%\_Inst_{i,t} + \beta_{10} 10K\_News_{i,t} + \beta_{11} Adv_{i,t} \\ & + \beta_{12} R\&D_{i,t} + \beta_{13} Std\_Ret_{i,t} \\ & + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Following prior literature (e.g., O'Brien and Bhushan 1990; Brennan and Subrahmanyam 1995; Lehavy et al. 2011), *#Analysts* is the number of analysts included in the first I/B/E/S consensus earnings forecast after 10-K filings. My variables of interest are *Uncertain* and *Weak_Modal*. *Uncertain* (*Weak_Modal*) is the proportion of uncertain (weak modal) words to the total words in 10-K filings as defined in Loughran and McDonald (2011). Due to a high correlation between *Uncertain* and *Weak_Modal*, I separate the two variables and run two different regressions so that I can capture any incremental effect of each variable. To investigate an incremental effect of uncertain and weak modal words in 10-K filings over readability and tone management on analyst following, I include variables for the 10-K file size (*File_Size*), positive (*Pos_Tone*) and negative (*Neg_Tone*) tone. The 10-K file size (*File_Size*) is a simple readability proxy that outperforms the Fog Index in that it effectively measures how managers convey valuation-relevant information to analysts and investors (Loughran and McDonald 2014). Management's use of positive and negative tone is relevant in this analysis because this study examines how management tone influences

analysts' behavior and forecast attributes. As defined by Loughran and McDonald (2011), *Pos_Tone (Neg_Tone)* is the proportion of positive words (negative words) to the total number of words in 10-Ks.

In model (2), I follow Lehavy et al. (2011) with respect to control variables. *Firm_Size* is a proxy for firm's operational and business environment. It is the logarithm of the market value of equity at the fiscal year-end. This variable explains how firm size influences the use of uncertain and weak modal words in 10-K filings. I expect that larger firms use more words of an ambiguous nature in their financial disclosures. *Growth* is the difference in sales volume between the current year and prior year divided by prior year sales volume. I control for sales growth (*Growth)* because Barth et al. (2001) argue that high-growth firms may increase analyst following due to investor interest and the potential for future investment opportunities. They also argue that, due to the uncertain business operational environment, high-growth firms may lead an analyst to less accurate and more dispersed earnings forecast. I measure the complexity of operations using the logarithm of the number of business segments (*Segments*) from the Compustat Segment File (Bradshaw et al. 2008). Following (Bhushan 1989, Brennan and Subrahmanyam 1995, and Frankel, Kothari, and Weber 2006), I include *%_Inst* to examine the level of institutional holdings. These studies find a positive relation between institutional ownership and analyst following and information content of forecasts. Also, Institutional ownership may increase analyst forecast accuracy and decrease forecast dispersion due to the enhanced information environments. *%_Inst* is the percentage of a firm's shares that are held by institutional investors from the most recent quarter before 10-K filing from the 13F disclosures. *10K _News* is defined as two-day event window for market-adjusted return to control the

informativeness of the 10-K filing. Like Barth et al. (2001), I include *Adv* and *R&D* control variables. *Adv (R&D)* is advertising expense (research and development expense) divided by operating expense. I also measure firm-specific stock return for business or operation volatility. *Std_Ret* is the standard deviation of the monthly stock returns in the prior year.

H2 predicts the relation between analyst forecast dispersion and the use of uncertain and weak modal words in 10-K filings. I estimate the following regression for H2.

$$\begin{aligned} Dispersion_{i,t} = \beta_0 + \beta_1 Uncertain_{i,t} + \beta_2 Weak\_Modal_{i,t} + \beta_3 File\_Size_{i,t} \\ + \beta_4 Pos\_Tone_{i,t} + \beta_5 Neg\_Tone_{i,t} + \beta_6 Firm\_Size_{i,t} + \beta_7 Growth_{i,t} \\ + \beta_8 Segments_{i,t} + \beta_9 \%\_Inst_{i,t} + \beta_{10} 10K\_News_{i,t} + \beta_{11} Adv_{i,t} \\ + \beta_{12} R\&D_{i,t} + \beta_{13} Std\_Ret_{i,t} + \beta_{14} \#Analysts_{i,t} \\ + \varepsilon_{i,t} \end{aligned} \quad (3)$$

Dispersion is the standard deviation of the individual analyst forecasts in the first analyst consensus earnings forecast after the 10-K report, scaled by share price 90 days before the consensus earnings forecast. In model (3), the variables of interest and control variables, except *#Analysts*, are the same as those used in regression model (2). I include *#Analysts* to examine whether the number of analysts influences the results.

H3 predicts the relation between analyst forecast accuracy and the use of uncertain and weak modal words in 10-K filings. I estimate the following regression for H3.

$$Accuracy_{i,t} = \beta_0 + \beta_1 Uncertain_{i,t} + \beta_2 Weak\_Modal_{i,t} + \beta_3 File\_Size_{i,t} + \beta_4 Pos\_Tone_{i,t} + \beta_5 Neg\_Tone_{i,t} + \beta_6 Firm\_Size_{i,t} + \beta_7 Growth_{i,t} + \beta_8 Segments_{i,t} + \beta_9 \%\_Inst_{i,t} + \beta_{10} 10K\_News_{i,t} + \beta_{11} Adv_{i,t} + \beta_{12} R\&D_{i,t} + \beta_{13} Std\_Ret_{i,t} + \beta_{14} \#Analysts_{i,t} + \varepsilon_{i,t} \quad (4)$$

Accuracy is the squared value of the difference between the reported earnings in I/B/E/S and the most recent analyst consensus forecast, scaled by stock price 90 days before the consensus earnings forecast. In model (4), all the variables of interest and control variables are the same as those in the regression model (3).

For H4 and H5, I follow Barron et al. (1998) to measure uncertainty in analyst private and common (public) information environment using the following equations.

$$Uncertainty\_Overall = \left(1 - \frac{1}{\#Analysts}\right) * Dispersion + Accuracy \quad (5)$$

$$Uncertainty\_Common = \frac{Accuracy - \frac{Dispersion}{\#Analysts}}{Uncertainty\_Overall} \quad (6)$$

These measures combine the accuracy, the dispersion, and the number of analyst forecasts and enable me to directly measure how uncertain and weak modal words in 10-Ks are related to analysts' private and common information environment. *Uncertainty_Overall* is the sum of uncertainty related to analysts' private information and uncertainty related to common (public) information to all analysts. *Uncertainty_Common* is the ratio of uncertainty in analysts' common information environment to uncertainty in analysts'

overall information environment, and it measures whether or not the public information (e.g., 10-Ks) becomes more important to analysts in comparison with private information. It measures how much average analysts' beliefs reflect common versus private information. For example, if *Dispersion* in the above equations nears zero (i.e., no disagreement among analysts), then total uncertainty is only associated with analysts' common information (i.e., *Uncertainty_Common* approaches 1).

H4 predicts the relation between uncertainty in analyst's overall information environment and the use of uncertain and weak modal words in 10-K filings. I estimate the following regression for H4.

$$
\begin{aligned}
Uncertainty\_Overall_{i,t} &= \beta_0 + \beta_1 Uncertain_{i,t} + \beta_2 Weak\_Modal_{i,t} + \beta_3 File\_Size_{i,t} \\
&+ \beta_4 Pos\_Tone_{i,t} + \beta_5 Neg\_Tone_{i,t} + \beta_6 Firm\_Size_{i,t} + \beta_7 Growth_{i,t} \\
&+ \beta_8 Segments_{i,t} + \beta_9 \%\_Inst_{i,t} + \beta_{10} 10K\_News_{i,t} + \beta_{11} Adv_{i,t} \\
&+ \beta_{12} R\&D_{i,t} + \beta_{13} Std\_Ret_{i,t} + \beta_{14} \#Analysts_{i,t} \\
&+ \varepsilon_{i,t} \qquad (7)
\end{aligned}
$$

In model (7), all the variables of interest and control variables are the same as those in the regression model (3).

H5 predicts the relation between uncertainty in analysts' common information environment and the use of uncertain and weak modal words in 10-K filings. I estimate the following regression for H5.

$$Uncertainty\_Common_{i,t}$$

$$= \beta_0 + \beta_1 Uncertain_{i,t} + \beta_2 Weak\_Modal_{i,t} + \beta_3 File\_Size_{i,t}$$
$$+ \beta_4 Pos\_Tone_{i,t} + \beta_5 Neg\_Tone_{i,t} + \beta_6 Firm\_Size_{i,t} + \beta_7 Growth_{i,t}$$
$$+ \beta_8 Segments_{i,t} + \beta_9 \%\_Inst_{i,t} + \beta_{10} 10K\_News_{i,t} + \beta_{11} Adv_{i,t}$$
$$+ \beta_{12} R\&D_{i,t} + \beta_{13} Std\_Ret_{i,t} + \beta_{14} \#Analysts_{i,t}$$
$$+ \varepsilon_{i,t} \quad (8)$$

In model (8), all the variables of interest and control variables are the same as those in the regression model (3). All multivariate regression analyses are controlled for year and industry fixed effect (two-digit SIC code).

For the additional tests in a later chapter in this study, I follow Li (2008) to measure the readability of 10-K filings. The Fog Index, developed by the computational linguistics literature, indicates the number of years of formal education required for a reader of average intelligence would need to read the document once and understand it. Specifically, the Fog Index is measured as follows:

$FOG_{i,t}$= (*average words per sentence + percent of complex words*) × *0.4* (9)

Where a complex word is defined as a word with three or more syllables. For example, the Fog Index greater than or equal to 18 means unreadable; the Fog Index between 14 and 18 means difficult; the Fog Index between 12 and 14 means ideal; the Fog Index between 10 and 12 means acceptable; and the Fog Index between 8 and 10 indicates childish language.

CHAPTER 5

DATA COLLECTION AND SAMPLE SELECTION

The sample in this study consists of U.S. firms. The sample period is 2000-2016 because this study employs the uncertain and weak modal word lists from Loughran and McDonald (2011), and they have updated the word lists up to 2016.

Panel A in Table 1 presents the sample selection procedure for the analysis of the determinants of the use of uncertain and weak modal words in 10-Ks. I obtain uncertain and weak modal words from the sentimental word lists in Loughran and McDonald (2011), and this yields a sample of 133,745 observations. I exclude regulated utilities (SIC code 4900-4999) and financial (SIC code 6000-6999) firms because they have a different operating and regulatory environment. This yields a sample of 84,861 observations. I obtain financial data from Compustat, business segment data from Compustat Segments data, stock return data from CRSP, M&A data from SDC Platinum M&A database, and SEO data from SDC Global New Issue database. I merge Loughran and McDonald's sentiment word lists with these datasets by CIK, ticker, and fiscal year. This yields a sample of 37,442 observations.

Panel B of Table 1 presents the sample selection procedure to examine the relation between the use of uncertain and weak modal words in 10-Ks and analysts' behavior and forecast outcomes. I obtain 10-K file size and the sentiment word lists such as positive, negative, uncertain, and weak modal, from Loughran and McDonald's database. This yields a sample of 133,745 observations. I exclude regulated utilities (SIC code 4900-4999) and financial (SIC code 6000-6999) firms because they have a different operating and

regulatory environment. This yields a sample of 84,861 observations. I retrieve financial data from Compustat, analyst data from I/B/E/S, stock return data from CRSP, and institutional holdings data from Thomson-Reuters 13f Holdings. I merge these financial and analysts' forecast related data with Loughran and McDonald sentiment word lists by CIK and fiscal year. This procedure yields a sample of 42,627 observations. The final sample size for multivariate regressions depends on the number of observation of dependent variables. These five dependent variables are the number of analysts (*#Analysts*), analysts' forecast dispersion (*Dispersion*), analysts' forecast accuracy (*Accuracy*), the uncertainty in analysts' overall information environment (*Uncertainty_Overall*), and the uncertainty in analysts' common (public) information environment (*Uncertainty_Common*). This yields a sample of 25,673. In detail, I have 25,673 observations for the number of analysts, 19,003 (15,213) observations for analysts' forecast dispersion (accuracy), and 14,308 (14,306) observations for uncertainty in analysts' overall (common) information environment.

CHAPTER 6

RESULTS

6.1 Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

Descriptive Statistics

Following Li (2008), Table 2 presents descriptive statistics for variables used in this study. I have two dependent variables – uncertain (*Uncertain*) and weak modal (*Weak_Modal*) words in 10-K filings. The mean (median) of *Uncertain* is 0.0131 (0.0131), and the standard deviation is 0.0031, indicating that on average 1.3% of words used in 10-Ks are uncertain words. The mean (median) of *Weak_Modal* is 0.0056 (0.0054), and the standard deviation is 0.0020, indicating on average about 0.6% of words used in 10-Ks are weak modal words. My study also includes independent variables that explain the determinants of the use of uncertain and weak modal words in 10-K filings. These variables are firm size (*Firm_Size*), market-to-book ratio (*MTB*), firm age (*Age*), special items (*Special_Items*), return and earnings volatility (*Std_Ret* and *Std_Earn*), number of business segments (*Segments*), merger-and-acquisition (*M&A*), seasoned equity offering (*SEO*), and firms incorporated in the state of Delaware (*Delaware*). The mean (median) of *Firm_Size* is 5.9224 (5.9018). The mean (median) of *MTB* is 2.0583 (1.5466). The mean (median) of *Age* is 17.5012 (13.1710). The mean (median) of *Special_Items* is -0.0224 (-0.0009)[11]. The mean (median) of *Std_Ret* is 0.1483 (0.1228). The mean (median) of *Std_Earn* is 0.1504 (0.0608). The mean (median) of *Segments* is 1.0460 (0.6931). The mean (median) of *M&A*

[11] Special items are scaled by book value of assets.

is 0.2846 (0.0000). The mean (median) of *SEO* is 0.1393 (0.0000). The mean (median) of *Delaware* is 0.0028 (0.0000). Overall, these results are compatible with those of Li (2008).

Univariate Analysis

Table 3 reports the correlation analysis. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. Panel A discusses the determinants of the use of uncertain words in 10-K filings. Panel A reports that the use of uncertain words in 10-K filings is positively correlated with *Firm_Size* (coefficient = 0.0544, $p<0.05$), *MTB* (coefficient = 0.1101, $p<0.05$), *Std_Earn* (coefficient = 0.0969, $p<0.05$), *M&A* (coefficient = 0.0347, $p<0.05$), and *SEO* (coefficient = 0.0772, $p<0.05$). It reports a negative relation between the use of uncertain words in 10-K filings and *Age* (coefficient = -0.1181, $p<0.05$), *Std_Ret* (coefficient = -0.0154, $p<0.05$), *Segments* (coefficient = -0.1387, $p<0.05$), and *Delaware* (coefficient = -0.0264, $p<0.05$). Overall, the results in Panel B are consistent with those of Panel A except for the relation between the use of weak modal words and *Std_Ret* and *M&A* which have the opposite sign on coefficients.

Multivariate Analysis

The first column in Table 4 reports uncertain words (*Uncertain*) as the dependent variable and the second column presents weak modal words (*Weak_Modal*) as the dependent variable. In the first column, I find a positive relation between uncertain words in 10-Ks and these variables – *Firm_Size* (coefficient = 0.0001, t = 13.34), *Std_Ret* (coefficient = 0.0008, t = 4.71), *Std_Earn* (coefficient = 0.0005, t = 10.03), and *M&A* (coefficient = 0.0001, t = 2.78). I find a negative relation between uncertain words in 10-Ks and these variables – *Age* (coefficient = -0.0000, t = -28.35), *Segments* (coefficient = -

0.0007, *t* = -19.68), and *Delaware* (coefficient = -0.0006, *t* = -2.30). The results are overall consistent with my predictions except for *M&A*. Unlike my prediction, the firms engaged in *M&A* are more likely to report uncertain words in their 10-Ks. The second column reports overall consistent results with the first column, however, I find a positive relation between weak modal words in 10-Ks and *MTB* (coefficient = 0.0001, *t* = 10.87) and *Special_Items* (coefficient = 0.0002, *t* = 1.87); I find a negative relation between weak modal words in 10-Ks and *M&A* (coefficient = 0.0000, t = -1.73). These findings are different from the first column.

Overall, this analysis finds that firm size, firm age, volatility of price returns and earnings, and number of business segments are consistently associated with the use of both uncertain and weak modal words in firms' financial reports. However, MTB, special items, and M&A differently impact the use of either uncertain or weak modal words in firms' 10-K filings.

6.2 The Association between the Use of Uncertain and Weak Modal Words in 10-K Filings and Analyst Forecast Attributes

Descriptive Statistics

Table 5 presents descriptive statistics for 25,673 firm-year observations. For multivariate analysis, I have five dependent variables - *#Analysts, Dispersion, Accuracy, Uncertainty_Overall*[12]*, and Uncertainty_Common*[13]. The mean (median) number of

[12] The sum of uncertainty related to analysts' private information and uncertainty related to common (public) information to all analysts following Barron et al. (1998).

[13] The ratio of common uncertainty to total uncertainty among analysts following Barron et al. (1998).

analysts (*#Analysts*) per firm-year observation is 7.5121 (5.0000). The standard deviation of *#Analysts* is 6.5571. The mean (median) of forecast dispersion (*Dispersion*) is 0.0143 (0.0041). The mean (median) of squared forecast error (*Accuracy*) is 0.0402 (0.0024). The standard deviation for *Dispersion* and *Accuracy* is 0.0308 and 0.1567, respectively. The mean (median) of *Uncertainty_Overall* and *Uncertainty_Common* are 0.0445 (0.0076) and 0.3259 (0.3056), respectively. This indicates that about 4 (33) percent of analyst uncertainty about future earnings following the 10-K filing is based on overall information among analysts (publicly available information). The standard deviation for *Uncertainty_Overall* and *Uncertainty_Common* are 0.1413 and 0.4299, respectively.

I have five independent variables - *Uncertain, Weak_Modal, File_Size, Pos_Tone*, and *Neg_Tone*. The mean (median) of *Uncertain* is 0.0134 (0.0135), indicating on average about 1.3% of words used in 10-Ks are uncertain words. The mean (median) of *Weak_Modal* is 0.0058 (0.0056), indicating on average about 0.6% of words used in 10-Ks are weak modal words. The mean of 10-K file size (*File_Size*) is 12.7344 (383,427 in megabytes); the median of *File_Size* is 12.6999 (327,702 in megabytes). The mean (median) of positive tone (*Pos_Tone*) is 0.0074 (0.0073), indicating on average about 0.7% of words used in 10-Ks have positive tone. The mean (median) of negative tone (*Neg_Tone*) is 0.0170 (0.0169), indicating on average about 1.7% of words used in 10-Ks have negative tone. Table 5 also provides descriptive statistics on control variables. The mean (median) size of sample firm (*Firm_Size*) is 6.4350 (6.3575). The mean (median) of sales growth rate (*Growth*) is 0.1565 (0.0801). The mean (median) number of the business segment (*Segments*) is 1.0578 (0.6931), and the mean (median) of the percent of institutional ownership (*%_Inst*) is 0.6263 (0.6825). The mean (median) of *10K_News* is

0.0355 (0.0211). The mean (median) ratio of advertisement expense to operating expense (*Adv*) is 0.0128 (0.0000). The mean (median) ratio of research and development expense (*R&D*) to operating expense is 0.1114 (0.0171). The mean (median) of price return volatility (*Std_Ret*) is 0.1407 (0.1184).

6.2.1 Analyst following

Univariate Analysis

Table 6 reports the correlation analysis. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. I find a positive correlation between analyst following (*#Analysts*) and these variables – *File_Size* (coefficient = 0.2613, $p<0.05$), *Neg_Tone* (coefficient = 0.0298, $p<0.05$), *Firm_Size* (coefficient = 0.7225, $p<0.05$), *Segments* (coefficient = 0.1148, $p<0.05$), *%_Inst* (coefficient = 0.3904, $p<0.05$), and *Adv* (coefficient = 0.0792, $p<0.05$). Also, I report a negative correlation between *#Analysts* and these variables – *Weak_Modal* (coefficient = -0.0157, $p<0.05$), *Pos_Tone* (coefficient = -0.0206, $p<0.05$), *Growth* (coefficient = -0.0165, $p<0.05$), *10K_News* (coefficient = -0.1591, $p<0.05$), *R&D* (coefficient = -0.0312, $p<0.05$), and *Std_Ret* (coefficient = -0.2516, $p<0.05$). I find that uncertain words (*Uncertain*) and weak modal words (*Weak_Modal*) are highly and positively correlated with each other (coefficient = 0.7414, $p<0.05$). I find that the correlation coefficients among the independent variables are generally moderate.

Multivariate Analysis

In Table 7, I predict that analyst following is affected by the level of uncertain and weak modal words in 10-K filings. To investigate an incremental effect of uncertain and weak modal words in 10-K filings over readability and tone management on analyst

following, I include variables for the 10-K file size (*File_Size*), positive (*Pos_Tone*) and negative (*Neg_Tone*) tone. In this multivariate analysis, the variables of interest are *Uncertain* and *Weak_Modal*. Due to a high correlation between these two variables, I separate the two variables and run two different regressions thereby capturing any incremental effect of each variable. I follow Lehavy et al. (2011) to control for other factors that can affect analyst following.

Main test variable: *Uncertain*

Models 1 – 4 in Table 7 include uncertain words as the main test variable and model 5 – 8 have weak modal words as the main test variable. From models 2 – 4, I find a positive and significant relation between analyst following and the use of uncertain words in 10-K filings. Model 1 tests the regression without including uncertain words and indicates analyst following is positively associated with *File_Size*, *Pos_Tone,* and *Neg_Tone.* Model 2 is the most comprehensive model. This model includes all test and control variables. It examines the incremental effect of uncertain words in a 10-K filing in addition to readability (*File_Size*) and tone management (*Pos_Tone* and *Neg_Tone*) on analyst following, indicating that more analysts follow firms with more uncertain words, less readable contexts, and more tone management in 10-K filings. Model 3 excludes positive and negative tone variables, and Model 4 excludes *File_Size* from Model 2 accordingly, but the results are consistent with Model 2. Interestingly, Model 3 shows the largest economic magnitude (coefficient = 28.5131, t = 17.61) for the *Uncertain* variable, suggesting the relation between analyst following and the use of uncertain words becomes even stronger when the model is controlled by *File_Size* and other control variables.

Main test variable: *Weak_Modal*

Models 5 - 8 in Table 7 examine the relation between analyst following and the effect of weak modal words in 10-K filings. Overall, the results are similar to those found in Models 1 – 4. I find a significantly positive relation between analyst following and the use of weak modal words in 10-K filings. Model 6 contains all test and control variables. Model 7 reports the largest economic magnitude (coefficient = 47.2850, t = 18.59) for *Weak_Modal* variable, suggesting the relation between analyst following and the use of weak modal words becomes stronger when the model is controlled by *File_Size* (readability) and other control variables.

The regression results on control variables overall are consistent with the prior literature. For example, analyst following is positively associated with firm size, advertisement and R&D expenses (e.g., Barth et al. 2001), and stock return volatility (e.g., Bhushan, 1989). Consistent with Lehavy et al. (2011), I find a negative relation between analyst following and number of business segments. Overall, I find that more analysts follow firms with more use of uncertain and weak modal words in firms' 10-K filings.

6.2.2 Forecast Dispersion

Univariate Analysis

Table 8 reports the correlation analysis. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. The univariate analysis finds that analyst forecast dispersion (*Dispersion*) is positively correlated with these variables – *Uncertain* (coefficient = 0.0207, p<0.05), *Weak_Modal* (coefficient = 0.1175, p<0.05), *File_Size* (coefficient = 0.0805, p<0.05), *Pos_Tone* (coefficient = 0.0395, p<0.05), *Neg_Tone* (coefficient = 0.1362, p<0.05), *Growth* (coefficient = 0.0254, p<0.05),

10K_News (coefficient = 0.2194, p<0.05), *R&D* (coefficient = 0.1913, p<0.05), and *Std_Ret* (coefficient = 0.4039, p<0.05). Also, I find a negative correlation between *Dispersion* and these variables – *Firm_Size* (coefficient = -0.3924, p<0.05), *Segments* (coefficient = -0.1239, p<0.05), *%_Inst* (coefficient = -0.3409, p<0.05), and *#Analysts* (coefficient = -0.1493, p<0.05). I find that uncertain words and weak modal words are highly and positively correlated with each other (coefficient = 0.7405, p<0.05). I find that the correlation coefficients between the independent variables are generally moderate.

Multivariate Analysis

In Table 9, I predict that forecast dispersion is affected by the level of uncertain and weak modal words in 10-K filings. To investigate the incremental effect of uncertain and weak modal words in 10-K filings over readability and tone management on forecast dispersion, I include variables such as *File_Size, Pos_Tone*, and *Neg_Tone*. The variables of interest are *Uncertain* and *Weak_modal*. Due to a high correlation between these two variables, I also separate the two variables and run two different regressions. I follow Lehavy et al. (2011) to control other factors that can affect forecast dispersion.

Main test variable: *Uncertain*

Models 1 – 4 in Table 9 include uncertain words as the main test variable and models 5 – 8 include weak modal words as the main test variable. From models 2 – 4, I find a negative and statistically significant relation between analyst forecast dispersion (*Dispersion*) and the use of uncertain words (*Uncertain*) in 10-K filings. Model 1 tests the regression without *Uncertain* and indicates forecast dispersion is positively associated with *File_Size* and *Pos_Tone*. Model 2 includes all test and control variables. It examines the incremental effect of uncertain words in 10-K filings over *File_Size, Pos_Tone,* and

Neg_Tone on forecast dispersion, indicating that *Dispersion* is associated with the use of less uncertain words, less readable contexts (i.e., greater 10-K file size), and more positive tone in firms'10-K filings. Model 3 excludes tone management, and Model 4 excludes *File_Size* accordingly, but the results are overall consistent with Model 2. Model 4 reports the largest economic magnitude (coefficient = -0.6781, t = -9.20) for *Uncertain*, suggesting the relation between forecast dispersion and the use of uncertain words becomes stronger when the model is controlled by tone management and other control variables.

Main test variable: *Weak_Modal*

Models 5 - 8 in Table 9 examine the relation between forecast dispersion and the effect of weak modal words in firms' 10-K filings. Overall, the results report a negative and significant relation between forecast dispersion and the use of weak modal words in 10-K filings. Model 6 is the most comprehensive model and reports positive (negative) relation between *Dispersion* and 10-K file size and tone variables (weak modal words). Model 8 reports the largest economic magnitude (coefficient = -1.7907, t = - 13.18) for *Weak_Modal*, suggesting the relation between analyst forecast dispersion and the use of uncertain words becomes stronger when the model includes tone management and other control variables.

The regression results on control variables are consistent with prior literature except for *Growth.* Consistent with Lehavy et al. (2011), analyst forecast dispersion is positively associated with number of business segments, 10-K news, advertisement and R&D expenses, and stock return volatility. There is a positive relation between forecast dispersion and analyst following (*#Analyst*) and a negative relation between forecast dispersion and firm size, sales growth, and % of institutional ownership.

Overall the results show that analyst forecast dispersion is negatively related to uncertain and weak modal words in 10-K filings, supporting the management-relations, analyst extra effort, and signaling theories. The management-relations hypothesis predicts that financial analysts collectively bias their forecasts upward with more uncertain information environment, which in turn reduces forecast dispersion. The analyst extra effort hypothesis predicts that analysts make extra efforts to produce earnings forecasts when facing more uncertain information environment. Finally, the signaling theory predicts that uncertain and weak modal words in 10-K filings signal analysts about the uncertainty of future earnings. Both theories also predict to reduce forecast dispersion.

6.2.3 Forecast Accuracy

Univariate Analysis

The correlation analysis is reported in Table 10. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. The univariate analysis finds negative and statistically significant correlations between forecast accuracy (forecast error) and *Uncertain* (coefficient = -0.0475, $p<0.05$) and *Weak_Modal* (coefficient = -0.0259, $p<0.05$), indicating that forecast error may decrease when firm's management uses more ambiguous language in their 10-Ks. *Accuracy* is positively correlated with *File_Size* (i.e., less readable 10-Ks) (coefficient = 0.0804, $p<0.05$), *Neg_Tone* (coefficient = 0.0515, $p<0.05$), *10K_News* (coefficient = 0.0882, $p<0.05$), *and Std_Ret* (coefficient = 0.1989, $p<0.05$). I also find a negative correlation between forecast error (*Accuracy*) and these variables - *Pos_Tone* (coefficient = -0.0575, $p<0.05$), *Firm_Size* (coefficient = -0.1589, $p<0.05$), *%_Inst* (coefficient = -0.1285, $p<0.05$), and *#Analysts* (coefficient = -0.0686, $p<0.05$). I find that uncertain and weak modal words are positively

correlated with each other (coefficient = 0.7343, $p<0.05$) and that the correlation coefficients between the independent variables are generally moderate.

Multivariate Analysis

I predict that forecast accuracy (forecast error) is affected by the level of uncertain and weak modal words in 10-K filings. To investigate an incremental effect of uncertain and weak modal words in 10-K filings over readability and tone management on forecast accuracy, I include variables such as *File_Size*, *Pos_Tone*, and *Neg_Tone*. The variables of interest in this analysis are *Uncertain* and *Weak_Modal*. Due to a high correlation between *Uncertain* and *Weak_Modal*, I separate the two variables and run two different regressions so that I can capture any incremental effect of each variable. I follow Lehavy et al. (2011) to control other factors that can affect forecast accuracy (forecast error).

Main test variable: *Uncertain*

Models 1 – 4 in Table 11 include uncertain words as the main test variable and models 5 – 8 have weak modal words as the main test variable. From models 2 – 4, I find a negative and significant relation between forecast accuracy (forecast error) and the use of uncertain words (*Uncertain*) in 10-K filings. Model 1 tests the regression without *Uncertain* and reports forecast error are positively (negatively) associated with *File_Size* (*Pos_Tone*). Model 2 includes all test and control variables. It examines the incremental effect of uncertain words in a 10-K filing in addition to readability (*File_Size*) and tone management (*Pos_Tone* and *Neg_Tone*) on forecast accuracy (forecast error). The results show that analysts have fewer forecast error for firms that use more uncertain words in 10-K filings. Model 3 excludes tone management and Model 4 excludes *File_Size* accordingly, but the results are overall consistent with Model 2. Model 4 reports the largest economic

magnitude (coefficient =-3.6533, *t* = -7.49) for *Uncertain*, suggesting the relation between forecast error and the use of uncertain words becomes stronger when the model includes tone management and other control variables.

Main test variable: *Weak_Modal*

Models 5 - 8 in Table 11 examine the relation between forecast accuracy (forecast error) and the effect of weak modal words in firms' 10-K filings. Overall, the results are similar to Models 1 – 4 and report a negative and statistically significant relation between forecast error and the use of weak modal words in 10-K filings. Model 6 contains all test and control variables, and it reports that forecast error may decrease when firms use more weak modal words in 10-Ks. Also, it reports that forecast error may increase when firm's10-K becomes (contains) more complex (more negative tone). Model 8 reports a negative relation between *Accuracy* and *Weak_Modal* and has the largest economic magnitude (coefficient =-9.4055, *t* = -10.37) for *Weak_Modal*, suggesting the relation between analyst forecast error and the use of weak modal words becomes stronger when the model includes tone management and other control variables.

The regression results on control variables are partially consistent with prior literature. Consistent with Lehavy et al. (2011), analyst forecast accuracy (forecast error) is positively associated with number of business segments, and stock return volatility. I also find that analyst following (*#Analyst*) is associated with increased forecast error. However, forecast accuracy (forecast error) is negatively associated with firm size, sales growth, % of institutional ownership and R&D expense.

Overall the results provide evidence that forecast accuracy (forecast error) is negatively related to uncertain and weak modal words in 10-K filings, supporting the

analyst extra effort and signaling theories. The analyst extra effort hypothesis predicts that analysts make extra efforts to produce more accurate earnings forecasts when facing more uncertain information environment. Finally, the signaling theory predicts that uncertain and weak modal words in 10-K filings signal analysts about the uncertainty of future earnings, which in turn improves forecast accuracy (i.e., decreased forecast error).

6.2.4 Uncertainty in Analysts' Overall Information Environment

Univariate Analysis

Table 12 reports the correlation analysis. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. The univariate analysis shows uncertainty in analysts' overall information environment (*Uncertainty_Overall*) is positively correlated with *File_Size* (coefficient = 0.1048, p<0.05), *Neg_Tone* (coefficient = 0.0655, p<0.05), *10K_News* (coefficient = 0.1078, p<0.05), and *Std_Ret* (coefficient = 0.2364, p<0.05). I find a negative relation between *Uncertainty_Overall* and these variables – *Uncertain* (coefficient = -0.0430, p<0.05), *Pos_Tone* (coefficient = -0.0566, p<0.05), *Firm_Size* (coefficient = -0.1827, p<0.05), *%_Inst* (coefficient = -0.1378, p<0.05), and *#Analysts* (coefficient = -0.0620, p<0.05). Also, I find that uncertain and weak modal words are highly and positively correlated with each other (coefficient = 0.7331, p <0.05). I find that the correlation coefficients between the independent variables are generally moderate.

Multivariate Analysis

I predict that uncertainty in analysts' overall information environment is affected by the level of uncertain and weak modal words in 10-Ks. To investigate an incremental effect of uncertain and weak modal words in 10-K filings over readability and tone

management on analysts' overall information environment, I include variables for *File_Size*, *Pos_Tone*, and *Neg_Tone*. The variables of interest are *Uncertain* and *Weak_Modal*. Due to a high correlation between these two variables, I separate them and run two different regressions so that I can capture any incremental effect of each variable. I follow Lehavy et al. (2011) to control for other factors that can affect uncertainty in overall information among analysts.

Main test variable: *Uncertain*

Models 1 – 4 in Table 13 include uncertain words as the main test variable and models 5 – 8 have weak modal words as the main test variable. From models 2 – 4, I find a negative and significant relation between uncertainty in analysts' overall information environment (*Uncertainty_Overall*) and the use of uncertain words (*Uncertain*) in 10-K filings. Model 1 tests the regression without uncertain words and reports that uncertainty in analysts' overall information environment (*Uncertainty_Overall*) is positively associated with *File_Size* (i.e., less readable 10-K). This result is consistent with Lehavy et al. (2011). I find a negative relation between *Uncertainty_Overall* and *Pos_Tone*. Model 2 includes all test and control variables. It examines the incremental effect of uncertain words in a 10-K filing in addition to readability (*File_Size*) and tone management (*Pos_Tone* and *Neg_Tone*) on uncertainty in overall information among analysts, indicating that less uncertainty in analysts' overall information environment is associated with firms having more uncertain words in their 10-Ks. Model 3 excludes *Pos_Tone* and *Neg_Tone*, and Model 4 excludes *File_Size* accordingly, but the results are overall consistent with Model 2. Model 4 shows the largest economic effect (coefficient = -3.5870, t = -8.04) on *Uncertain*, suggesting the relation between uncertainty in analysts' overall

information environment and the use of uncertain words becomes stronger when the model includes tone management and other control variables.

Main test variable: *Weak_Modal*

Models 5 - 8 in Table 13 examine the relation between uncertainty in analysts' overall information environment (*Uncertainty_Overall*) and the effect of weak modal words (*Weak_Modal*) in firms' 10-K filings. Overall, the results present negative and statistically significant relation between *Uncertainty_Overall* and *Weak_Modal* in 10-K filings. Model 6 contains all test and control variables. It shows a positive (negative) relation between *Uncertainty_Overall* and *Neg_Tone* (*Weak_Modal*). Among models 6 - 8, model 8 indicates the largest coefficient (-8.7673, t = -10.55) on *Weak_Modal*, suggesting the relation between uncertainty in analysts' overall information environment and the use of weak modal words becomes stronger when the model includes tone management and other control variables.

For control variables in this analysis, uncertainty in overall information among analysts is positively associated with number of business segments, stock return volatility, and number of analysts. I report a negative relation between uncertainty in overall information among analysts and variables such as firm size, sales growth, % of institutional ownership and R&D expense.

Overall, Table 13 reports that uncertainty in analysts' overall information environment is decreasing in *Uncertain* and *Weak_Modal*, suggesting that uncertainty in the analysts' overall information environment decreases with firms' use of more ambiguous language in 10-K filings. This finding is different from my prediction which suggests that managers may use ambiguous words to signal analysts and investors about

the uncertainty of future earnings. Future research should investigate the extent to which uncertain and weak model words used in 10-Ks affect analysts' overall information environment using behavioral research methods including interviews with analysts.

6.2.5 Uncertainty in Analysts' Common Information Environment

Univariate Analysis

Table 14 reports the correlation analysis. The correlation coefficients below (above) the diagonal line are Pearson (Spearman) correlation coefficients. The univariate analysis finds that uncertainty in analysts' common information environment (*Uncertainty_Common*) is positively correlated with *File_Size* (coefficient = 0.0600, $p<0.05$), *Firm_Size* (coefficient = 0.0942, $p<0.05$), *Segments* (coefficient = 0.0810, $p<0.05$), *%_Inst* (coefficient = 0.1093, $p<0.05$), *Std_Ret* (coefficient = 0.0192, $p<0.05$), *and #Analysts* (coefficient = 0.1070, $p<0.05$). I find a negative relation between *Uncertainty_Common* and these variables – *Uncertain* (coefficient = -0.0704, $p<0.05$), *Weak_Modal* (coefficient = -0.0947, $p<0.05$), *Pos_Tone* (coefficient = -0.0710, $p<0.05$), *Neg_Tone* (Coefficient = -0.0181, $p<0.05$), *10K_News* (coefficient =-0.0184, $p<0.05$), *and R&D* (coefficient = -0.1151, $p<0.05$). Also, I find that *Uncertain* and *Weak_Modal* are highly correlated with each other (coefficient = 0.7330, $p<0.05$). I find that the correlation coefficients between the independent variables are generally moderate.

Multivariate Analysis

I predict that uncertainty in analysts' common information environment is affected by the level of uncertain and weak modal words in 10-K filings. To investigate an incremental effect of uncertain and weak modal words in 10-K filings over readability and tone management on uncertainty in public information among analysts, I include *File_Size*,

Pos_Tone, and *Neg_Tone*. The variables of interest are *Uncertain* and *Weak_Modal*. Due to a high correlation between these variables, I separate them and run two different regressions so that I can capture any incremental effect of each variable. I follow Lehavy et al. (2011) to control other factors that can affect uncertainty in analysts' common information environment.

Main test variable: *Uncertain*

Models 1 – 4 in Table 15 include uncertain words as the main test variable and models 5 – 8 have weak modal words as the main test variable. From models 2 – 4, I find overall a negative relation between *Uncertainty_Common* and *Uncertain*. Model 1 tests the regression without uncertain words. I find *Uncertainty_Common* is positively associated with *File_Size* (i.e., less readable 10-K) and negatively associated with *Pos_Tone*. Model 2 contains all test and control variables. It examines the incremental effect of uncertain words in 10-K filings in addition to readability (*File_Size*) and tone management (*Pos_Tone* and *Neg_Tone*) on uncertainty in analysts' common information environment. I find a negative but insignificant effect on *Uncertain.* Model 3 excludes tone variables, and Model 4 excludes the *File_Size* variable accordingly. Among models 2 - 4, model 4 reports the largest economic effect (coefficient = -5.4532, t = -3.87) on *Uncertain*, indicating that a negative and significant relation between uncertainty in analysts' common information environment and the use of uncertain words becomes stronger when the model includes tone management and other control variables.

Main test variable: *Weak_Modal*

Models 5 - 8 in Table 15 examine the relation between uncertainty in analysts' common information environment (*Uncertainty_Common*) and the effect of weak modal

words (*Weak_Modal*) in firms'10-K filings. Overall, the results report a negative and statistically significant relation between *Uncertainty_Common* and *Weak_Modal*. Model 6 includes all test and control variables. It reports a positive (negative) relation between *Uncertainty_Common* and *File_Size* (*Weak_Modal*). Among models 6 - 8, model 8 reports the largest economic magnitude (coefficient = -9.0535, t = -3.44) for *Weak_Modal*, suggesting the relation between uncertainty in analysts' common information environment and the use of weak modal words becomes stronger when the model includes tone management and other control variables. Consistent with Lehavy et al. (2011), *Uncertainty_Common* is positively (negatively) associated with these control variables - *%_Inst* and *Std_Ret* (*R&D*). I also find a positive relation between *Uncertainty_Common* and *Segments* and *#Analysts*.

Overall, this analysis finds that uncertainty in analysts' common information environment decreases for firms with increased use of uncertain and weak modal words in firms' annual financial disclosures. Again, this finding is different from my prediction, which indicates that managers may use ambiguous words in 10-K filings to signal analysts and investors about the uncertainty of future earnings. Future research should investigate the extent to which the use of uncertain and weak modal words in 10-Ks affects the uncertainty in analysts' common information environment using behavioral research methods including interviews with analysts.

6.3 Additional Tests

For a robustness test, I replace *File_Size* with the Fog Index (*Fog*) to examine if the effect of uncertain and weak modal words in 10-Ks hold for analyst forecast attributes with

a different measure of readability (*Fog*). I employ the Fog Index data from Feng Li's website[14] for the sample period of 2000 – 2011. Results are reported in Appendix 3.

6.3.1 Analyst following: Fog Replacement

Table 16 reports the correlation analysis. The univariate analysis reports overall consistent results with Table 6, but the correlation between analyst following and the Fog index (*Fog*) (coefficient = 0.0442, p<0.05) in Table 16 is weaker than the correlation between analyst following and *File_Size* (coefficient =0.2613, p<0.05) in Table 6. I find that the correlation coefficients between the independent variables are generally moderate.

Table 17 reports multivariate analysis with *Fog* as the readability measure. Overall, it provides consistent results with Table 7. In Table 17, the coefficients of uncertain and weak modal words in Model 2 (coefficient = 8.2556, *t* = 4.35) and Model 5 (coefficient = 28.8694, *t* = 7.99) become weaker than those of Model 2 and 6 in Table 7 when *File_Size* is replaced with *Fog*. The results for control variables are consistent with those of Table 7.

6.3.2 Analyst Forecast Dispersion: Fog Replacement

Table 18 reports the correlation analysis. The univariate analysis reports overall consistent results with Table 8, but the correlation between forecast dispersion (*Dispersion*) and the Fog index (*Fog*) (coefficient = 0.0085, p<0.05) in Table 18 is weaker than the correlation between forecast dispersion and *File_Size* (coefficient = 0.0805, p<0.05) in Table 8. I find that the correlation coefficients between the independent variables are generally moderate.

Table 19 reports multivariate analysis with *Fog* as the readability measure. Overall it reports consistent results with Table 9. In Table 19, the coefficients of uncertain and

[14] Please check the data from http://webuser.bus.umich.edu/feng/.

weak modal words in Model 2 (coefficient = -0.5170, *t* = -5.91) and Model 5 (coefficient = -1.5384, *t* = -9.23) become stronger than those of Model 2 and 6 in Table 9 when *File_Size* is replaced with *Fog*. This indicates the incremental effect of the use of uncertain and weak modal words in 10-K filings on forecast dispersion is increasing when the Fog index used for the readability measure. The results for control variables are consistent with those of Table 9.

6.3.3 Analyst Forecast Accuracy: Fog Replacement

Table 20 reports the correlation analysis. The univariate analysis reports overall consistent results with Table 10, but the correlation between forecast accuracy (forecast error) and the Fog index (*Fog*) (coefficient = 0.0108, $p<0.05$) in Table 20 is weaker than the correlation between analyst accuracy and *File_Size* (coefficient = 0.0804, $p<0.05$) in Table 10. I find that the correlation coefficients between the independent variables are generally moderate.

Table 21 reports multivariate analysis with *Fog* as the readability measure. Overall, results are consistent results with Table 11. In Table 21, the coefficients of uncertain and weak modal words in Model 2 (coefficient = -3.3171, *t* = -5.59) and Model 5 (coefficient = -9.9156, *t* = -8.74) are stronger than those of Model 2 and 6 in Table 11 when *File_Size* is replaced with *Fog*. This indicates the incremental effect of the use of uncertain and weak modal words in 10-K filings on forecast accuracy is increased when the Fog index is used for the readability measure. The results of control variables are consistent with those of Table 11.

6.3.4 Uncertainty in Analysts' Overall Information Environment: Fog Replacement

Table 22 reports the correlation analysis. The univariate analysis reports overall consistent results with Table 12, but the correlation between *Common_Overall* and the Fog index (*Fog*) (coefficient = 0.0104, $p<0.05$) in Table 22 is weaker than the correlation between analyst following and *File_Size* (coefficient = 0.1048, $p<0.05$) in Table 12. I find that the correlation coefficients between the independent variables are generally moderate.

Table 23 replaces *File_Size* with *Fog* for a readability proxy. Overall results are consistent with Table 13. In Table 23, the coefficients of uncertain and weak modal words in Model 2 (coefficient = -3.0118, t = -5.61) and Model 5 (coefficient = -8.9429, t = -8.69) are stronger than those of Model 2 and 6 in Table 13 when *File_Size* is replaced with *Fog*. This indicates the incremental effect of the use of uncertain and weak modal words in 10-K filings on uncertainty in analysts' overall information is increasing when the Fog index is used for the readability proxy. The results of control variables are consistent with those of Table 13.

6.3.5 Uncertainty in Analysts' Common Information Environment: Fog Replacement

Table 24 reports the correlation analysis. The univariate analysis reports overall consistent results with Table 14. In Table 24, the correlation between *Uncertainty_Common* and *Fog* is negative and significant (coefficient = -0.0305, $p<0.05$), but in Table 14 the coefficient of *File_Size* (coefficient = 0.0600, $p<0.05$) reports positive and significant correlation. I find that the correlation coefficients between the independent variables are generally moderate.

Table 25 reports multivariate analysis with *Fog* as a readability measure. Overall, it reports consistent results with Table 15. In Table 25, the coefficients of uncertain and

weak modal words in Model 2 (coefficient = -8.5482, *t* = -4.61) and Model 5 (coefficient = -16.4710, *t* = - 4.63) are stronger than those of Model 2 and 6 in Table 15 when *File_Size* is replaced with *Fog*. This indicates the incremental effect of the use of uncertain and weak modal words in 10-K filings on uncertainty in analysts' common information environment is increasing when the Fog index is used for the readability measure. The results of control variables are consistent with those of Table 15.

This additional analysis reports that the Fog replacement does not change overall results. Interestingly, it also reports that the effect of the use of ambiguous language in 10-Ks on analysts' forecast attributes becomes stronger when *Fog* replaces *File_Size* (except *Analyst Following*). One limitation of this additional analysis is inconsistent sample period[15]. Future studies will match the sample period.

[15] Main test sample period is 2000-2016.

CHAPTER 7

SUMMARY AND CONCLUSION

There is a growing body of literature demonstrating that language sentiment and tone embedded in financial disclosures and earnings press releases are informative for users of financial information. This study aims to contribute to this literature by investigating the determinants of the use of uncertain and weak modal words in 10-K filings and the extent to which the use of these words affects analyst forecast attributes. These are important issues because the use of ambiguous language such as uncertain and weak modal words in 10-K filings can increase informational risk and decrease investors' ability to comprehend financial reports. Recent studies find evidence supporting that ambiguous language of corporate disclosures affects valuation uncertainty and cost of debt. This is the first study to examine the characteristics of firms that use uncertain and weak modal words in their 10-K filings. This is also the first study to examine how financial analysts perceive information contained in uncertain and weak modal words in 10-K filings and reflect this information into their decision making.

This study first examines the determinants of the use of uncertain and weak modal words in 10-K filings, and I find that the use of uncertain and weak modal words in 10-K filings is positively associated with firm size and volatility of stock return and earnings. I further find that the use of these words is negatively associated with firm age and number of business segments. More importantly, this study investigates the extent to which the use of uncertain and weak modal words in 10-K filings affects analysts' behavior and forecast outcomes. First, I examine the relation between analyst following and the use of uncertain and weak modal words in 10-K filings. I find a positive and significant relation between

number of analysts and the use of ambiguous language in 10-Ks. Second, I examine whether forecast dispersion is associated with the use of uncertain and weak modal words in 10-Ks. I find that analyst forecast dispersion is negatively related to uncertain and weak modal words in 10-K filings. The result indicates that uncertain and weak modal words in 10-K filings contain valuable information about firms' future earnings so that analysts produce less dispersed forecasts among them. This finding is consistent with the notion that financial analysts collectively bias their forecasts upward with more uncertain information environment (i.e. the management-relations hypothesis). This finding is also consistent with the notion that analysts make extra efforts to produce forecasts when facing more uncertain information environment (i.e. the analyst extra effort hypothesis). Finally, this finding is also consistent with the notion that uncertain and weak modal words signal analysts and other market participations about firms' future earnings (i.e. the signaling theory). Analysts are able to use this information to produce more accurate forecasts, which in turn reduces forecast dispersion among analysts. Third, I examine the relation between forecast accuracy and the use of uncertain and weak modal words in 10-Ks. The finding provides evidence that use of uncertain and weak modal words in 10-K filings relates to more accurate analyst forecasts. The result suggests that analysts effectively process information contains uncertain and weak modal words in 10-K filings, which in turn produce more accurate forecasts. This finding is also consistent with the analyst extra effort and the signaling theories as described in this dissertation. Fourth, I predict that higher uncertainty in the analysts' overall information environment increases the use of uncertain and weak modal words in firms' 10-K filings. Inconsistent with my prediction, I find that uncertainty in the analysts' overall information environment decreases with firms' use of

more ambiguous language in 10-Ks. Finally, I predict that use of ambiguous words in 10-K filings increases the degree to which analysts share a common belief such as analysts' perception about the information contained in 10-K filings. Inconsistent with prediction, I find that increased use of uncertain and weak modal words in 10-K filings is associated with less uncertainty in analysts' common information environment. Both findings suggest that managers may use uncertain and weak modal words in 10-Ks to signal the market participants about their future earnings, which in turn reduces uncertainty in analysts' overall and common information environment. Future research should use behavioral research methods such as interviews with analysts to enhance our understanding of the effect of uncertain information environment on analysts' behavior.

The above results and findings are robust even after controlling for readability, measured by the FOG Index and file size, and management tone. Overall, I find that uncertain and weak modal words in 10-K filings provide significant incremental information content regarding analyst forecast attributes beyond readability and management tone. Also, I provide evidence that analysts can effectively process the information contained in uncertain and weak modal words in 10-Ks to produce their forecasts.

As a limitation, some of my findings are not consistent with my predictions due to potential measurement errors in variables. Therefore, future research is to survey analysts to further understand how they interpret and process financial disclosures with ambiguous language such as uncertain and weak modal words.

REFERENCES

Ackert, Lucy F., and George Athanassakos, 1997, Prior uncertainty, analyst bias and subsequent abnormal returns, *Journal of Financial Research* 20, 253-273.

Ball, R. 1992. The earnings price anomaly. *Journal of Accounting and Economics* 15 _2–3_: 319–345.

Barron, O. E., D. Byard, O. Kim, and D. E. Stevens. 1998. Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review* 73 _4_: 421–433.

Barth, M. E., R. Kasznik, and M. F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39 _1_: 1–34.

Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11_2–3_: 255–274.

Biddle, G.; G. Hilary; and R. Verdi. "How Does Financial Reporting Quality Relate to Investment Efficiency?" *Journal of Accounting and Economics* 48 (2009): 112–31.

Bloomfield, Robert J., 2002, The "incomplete revelation hypothesis" and financial reporting, Accounting Horizons 16, 233–243.

Bonsall, S. B., and B. P. Miller. "The Impact of Narrative Disclosure Readability on Bond Ratings and Rating Agency Disagreement." Working paper, Ohio State University, 2014.

Botosan, C. A., and M. S. Harris. 2000. Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. *Journal of Accounting Research* 38 _2_: 329–353.

Bradley, D. J., Gokkaya, S., & Liu, X. (2017). Before an analyst becomes an analyst: Does industry experience matter? Journal of Finance 72-2 (2017): 751-792.

Bradshaw, M. T., G. S. Miller, and G. Serafeim. 2008. Accounting method heterogeneity and analysts' forecasts. Working paper, Harvard University and University of Michigan.

Bratten, B.; C. A. Gleason; S. Larocque; and L. F. Mills. "Forecasting Tax Expense: New Evidence from Analysts." Working paper, University of Notre Dame, 2014.

Brennan, and A. Subrahmanyam. 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics* 38 _3_: 361–381.

Callen, J. L., M. Khan, and H. Lu. 2009. Accounting quality, stock price delay and future stock returns. Working paper, University of Toronto and Massachusetts Institute of Technology.

Chen, S., Matsumoto, D. 2006. Favorable versus unfavorable recommendations: The impact on analyst access to management-provided information. Journal of Accounting Research 44, 657-689.

Cox, C. 2007. Closing remarks to the Second Annual Corporate Governance Summit. Delivered at the USC Marshall School of Business, Los Angeles, CA, March 23. Available at http://www.sec.gov/news/speech/2007/spch032307cc.htm.

Daines, R., 2001. Does Delaware law improve firm value? *Journal of Financial Economics* 62, 525–558.

Das, S., Levine, C., Sivaramakrishnan, K., 1998. Earnings predictability and bias in analysts' earnings forecasts. *The Accounting Review* 73, 277–294.

Davis, A. K., W. Ge, D. Matsumoto, and J. L. Zhang. "The Effect of Manager-specific Optimism on the Tone of Earnings Conference Calls." *Review of Accounting Studies*, 20, (2015), pp. 639–673.

Davis, A. K., J. M. Piger and L. M. Sedor. "Beyond the Numbers: managers' use of optimistic and pessimistic tone in earnings press release." *Contemporary Accounting Research* 29 (2012): 845-868.

Davis, A., and I. Tama-Sweet. 2012. Managers' use of language across alternative disclosure outlets:Earnings press releases versus MD&A. *Contemporary Accounting Research* 29: 804–837.

De Franco,G.; O.Hope; D.Vyas; and Y.Zhou. "Analyst Report Readability." *Contemporary Accounting Research* 32 (2015): 76–104.

Demers, E., & Vega, C. (2014). Understanding the role of managerial optimism and uncertainty in the price formation process: evidence from the textual content of earnings announcements. Working paper, Univeristy of Virginia.

Druz, Marina and Wagner, Alexander F. and Zeckhauser, Richard J., Reading Managerial Tone: How Analysts and the Market Respond to Conference Calls (January 22, 2016). HKS Working Paper No. 16-004. Available at SSRN: https://ssrn.com/abstract=2782672.

Duru, A., and D. M. Reeb. 2002. International diversification and analysts' forecast accuracy and bias. *The Accounting Review* 77 _2_: 415–433.

Ertugrul, M., J. Lei, J. Qiu, and C. Wan. "Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing." *Journal of Financial and Quantitative Analysis 52* (2017): No.2: 811–836.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2009. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* 15: 915–953.

Francis, T. 2014, September 14. To be clear, SEC reviewers want filings in plain English, period. Wall Street Journal. Available at: http://www.wsj.com/articles/to-be-clear-secreviewers-want-filings-in-plain-english-period-1410555347.

Frankel, R., S. P. Kothari, and J. Weber. 2006. Determinants of the informativeness of analyst research. *Journal of Accounting and Economics* 41 _1–2_: 29–54.

Frankel, R., W. Mayew, and Y. Sun. 2010. Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies* 15: 220–242.

Guay, W.; D. Samuels; and D. Taylor. "Guiding Through the Fog: Financial Statement Complexity and Voluntary Disclosure." Working paper, University of Pennsylvania, 2015.

Healy, P. M., A. P. Hutton, and K. G. Palepu. 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research* 16 _3_: 485–520.

Hirst, D. E., and P. E. Hopkins. 1998. Comprehensive income reporting and analysts' valuation judgments. *Journal of Accounting Research* 36: 47–75.

Hirst, D. E., P. E. Hopkins, and J. M. Wahlen. 2004. Fair values, income measurement, and bank analysts' risk and valuation judgments. *The Accounting Review* 79 _2_: 453–472.

Hong H., and J. Kubik. "Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts Authors." *The Journal of Finance* 58 (2003): Issue 1: 313–351.

Hopkins, P. E., R. W. Houston, and M. F. Peters. 2000. Purchase, pooling, and equity analysts' valuation judgments. *The Accounting Review* 75 _3_: 257–281.

Huang, X.; S. H. Teoh; and Y. Zhang. "Tone Management." *The Accounting Review* 89 (2014):1083–113.

Huberts, L., Fuller, R., 1995. Predictability bias in the U.S. equity market. *Financial Analysts Journal* 51, 12–28.

Indjejikian, R., 1991, The impact of costly information interpretation on firm disclosure decisions, *Journal of Accounting Research* 29, 277-301.

Jones, Michael J., and Paul A. Shoemaker, 1994, Accounting narratives: A review of empirical studies of content and readability, *Journal of Accounting Literature* 13, 142–184.

Lang, M.H., K. V. Lins, and D. P. Miller. 2004. Concentrated control, analyst following, and valuation: Do analysts matter most when investors are protected least? *Journal of Accounting Research* 42 _3_: 589–623.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 _4_: 467–492.

Lawrence, A. "Individual Investors and Financial Disclosure." *Journal of Accounting & Economics* 56 (2013): 130–47.

Lehavy, R.; F. Li; and K. Merkley. "The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts." *The Accounting Review* 86 (2011): 1087–115.

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

Li, J., and X. Zhao. "Complexity and Information Content of Financial Disclosures: Evidence from Evolution of Uncertainty Following 10-K Filings." Working paper, University of Texas at Dallas, 2014.

Lim, T. (2001). Rationality and Analysts' Forecast Bias. *Journal of Finance*, LVI (1), 369–385.

Loughran, T., and B. McDonald. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66 (2011): 35–65.

Loughran, T., and B. McDonald. "IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language." *Journal of Financial Economics* 109 (2013): 307–26.

Loughran, T., and B. McDonald. "Measuring Readability in Financial Disclosures." *Journal of Finance* 69 (2014): 1643–71.

Loughran, T., and B. McDonald. "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research* (2016) 54(4): 1187-1230.

Mikhail, M. B., B. R. Walther, and R. H. Willis. 1999. Does forecast accuracy matter to security analysts? *The Accounting Review* 74 _2_: 185–200.

Milian, J. and A. Smith. 2017. "An Investigation of Analysts' Praise of Management During Earnings Conference Calls. *Journal of Behavioral Finance* Volume 18, 2017 - Issue 1.

Miller, B. P. “The Effects of Reporting Complexity on Small and Large Investor Trading.” *The Accounting Review* 85 (2010): 2107–43.

O’Brien and R. Bhushan. 1990. Analyst following and institutional ownership. *Journal of Accounting Research* 28: 55–76.

Plumlee, M (*2003*) The Effect of Information Complexity on Analysts' Use of That Information. *The Accounting Review*: January 2003, Vol. 78, No. 1, pp. 275-296.

Price, S., Doran, J., Peterson, D., & Bliss, B. (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. *Journal of Banking & Finance, 36*(4), 992–1011.

Rennekamp, Kristina, 2012, “Processing fluency and investors’ reactions to disclosure readability.” *Journal of Accounting Research* 50, 1319–1354.

Schipper, K. 1991. Analysts’ forecasts. *Accounting Horizons* 5 (4): 105-121.

Schroeder, M. 2002. SEC proposes rules to improve disclosure by public companies. *Wall Street Journal* _May 1_: C5.

SEC, 1998. A Plain English Handbook: How to Create Clear SEC Disclosure Documents. U.S. Securities and Exchange Commission, Washington, DC.

You, Haifeng, and Xiao-jun Zhang. "Financial reporting complexity and investor underreaction to 10-K information." *Review of Accounting Studies* 14.4 (2009): 559-586.

TABLE 1

Sample Selection

Panel A: Sample Selection for the Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

	Observations
All firms in Loughran and McDonald Word Lists firm-years 2000-2016	133,745
Less: observations in regulated utilities (SIC 4900-4999) and financial (SIC 6000-6999) firms	(48,884)
	84,861
Less: firm-years with insufficient financial and special events data	(47,419)
Number of observations	37,442

Panel B: Sample Selection for the Association between the Use of Uncertain and Weak Modal Words in 10-K Filings and Analyst Forecast Attributes

	Observations
All firms in Loughran and McDonald Word Lists firm-years 2000-2016	133,745
Less: observations in regulated utilities (SIC 4900-4999) and financial (SIC 6000-6999) firms	(48,884)
	84,861
Less: firm-years with insufficient Financial data	(42,234)
	42,627
Less: unmatched firm-years between I/B/E/S data and financial data for five dependent variables	(16,954)
Number of observations	25,673

TABLE 2

Descriptive Statistics for the Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

Variable	N	Mean	Median	Std. Dev
Dependent				
Uncertain	37,442	0.0131	0.0131	0.0031
Weak_Modal	37,442	0.0056	0.0054	0.0020
Independent				
Firm_Size	37,442	5.9224	5.9018	2.0292
MTB	37,442	2.0583	1.5466	1.5606
Age	37,442	17.5012	13.1710	15.2377
Special_Items	37,442	-0.0224	-0.0009	0.0759
Std_Ret	37,442	0.1483	0.1228	0.0946
Std_Earn	37,442	0.1504	0.0608	0.2844
Segments	37,442	1.0460	0.6931	0.4277
M&A	37,442	0.2846	0.0000	0.4512
SEO	37,442	0.1393	0.0000	0.3463
Delaware	37,442	0.0028	0.0000	0.0529

Variables are defined in Appendix 1.

TABLE 3

Correlation Matrix for the Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

Panel A: Uncertain Words

		1	2	3	4	5	6	7	8	9	10	11
1	*Uncertain*		0.0556*	0.1323*	-0.1230*	-0.0271*	0.0117*	0.1447*	-0.1386*	0.0332*	0.0757*	-0.0282*
2	*Firm_Size*	0.0544*		0.3743*	0.2016*	-0.0147*	-0.4539*	-0.3723*	0.2295*	0.3027*	0.1292*	-0.0085
3	*MTB*	0.1101*	0.2193*		-0.0913*	0.1015*	-0.0619*	0.1608*	-0.1420*	0.0533*	0.0720*	0.0018
4	*Age*	-0.1181*	0.2822*	-0.1217*		0.0318*	-0.3428*	-0.3654*	0.2824*	0.0787*	-0.0093	0.0055
5	*Special_Items*	-0.0036	0.1480*	0.0411*	0.0758*		-0.0850*	-0.1048*	-0.0657*	-0.0472*	0.0071	0.0048
6	*Std_Ret*	-0.0154*	-0.4055*	0.0842*	-0.2757*	-0.2101*		0.5339*	-0.2077*	-0.2176*	-0.0154*	-0.0078
7	*Std_Earn*	0.0969*	-0.1990*	0.2615*	-0.2289*	-0.1767*	0.3410*		-0.2723*	-0.1976*	0.0022	-0.0031
8	*Segments*	-0.1387*	0.2438*	-0.1818*	0.3165*	0.0165*	-0.1781*	-0.1901*		0.1172*	0.0107*	0.0014
9	*M&A*	0.0347*	0.3084*	-0.0275*	0.0953*	0.0319*	-0.1858*	-0.1103*	0.1199*		0.0521*	-0.0066
10	*SEO*	0.0772*	0.1215*	0.0596*	0.0231*	0.0394*	-0.0159*	0.0265*	0.0136*	0.0521*		-0.0038
11	*Delaware*	-0.0264*	-0.0012	0.0015	0.0190*	0.0026	-0.0079	0.0007	0.0064	-0.0066	-0.0038	

Panels A and B report the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 1.

TABLE 3 (Continued)

Panel B: Weak Modal Words

		1	2	3	4	5	6	7	8	9	10	11
1	*Weak_Modal*		0.0009	0.2060*	-0.2741*	-0.0292*	0.1432*	0.3076*	-0.2444*	-0.0132*	0.0794*	-0.0187*
2	*Firm_Size*	0.0005		0.3743*	0.2016*	-0.0147*	-0.4539*	-0.3723*	0.2295*	0.3027*	0.1292*	-0.0085
3	*MTB*	0.2240*	0.2193*		-0.0913*	0.1015*	-0.0619*	0.1608*	-0.1420*	0.0533*	0.0720*	0.0018
4	*Age*	-0.2574*	0.2822*	-0.1217*		0.0318*	-0.3428*	-0.3654*	0.2824*	0.0787*	-0.0093	0.0055
5	*Special_Items*	-0.0208*	0.1480*	0.0411*	0.0758*		-0.0850*	-0.1048*	-0.0657*	-0.0472*	0.0071	0.0048
6	*Std_Ret*	0.1022*	-0.4055*	0.0842*	-0.2757*	-0.2101*		0.5339*	-0.2077*	-0.2176*	-0.0154*	-0.0078
7	*Std_Earn*	0.2326*	-0.1990*	0.2615*	-0.2289*	-0.1767*	0.3410*		-0.2723*	-0.1976*	0.0022	-0.0031
8	*Segments*	-0.2489*	0.2438*	-0.1818*	0.3165*	0.0165*	-0.1781*	-0.1901*		0.1172*	0.0107*	0.0014
9	*M&A*	-0.0193*	0.3084*	-0.0275*	0.0953*	0.0319*	-0.1858*	-0.1103*	0.1199*		0.0521*	-0.0066
10	*SEO*	0.0899*	0.1215*	0.0596*	0.0231*	0.0394*	-0.0159*	0.0265*	0.0136*	0.0521*		-0.0038
11	*Delaware*	-0.0153*	-0.0012	0.0015	0.0190*	0.0026	-0.0079	0.0007	0.0064	-0.0066	-0.0038	

Panels A and B report the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 1.

TABLE 4

Multivariate Analysis for the Determinants of the Use of Uncertain and Weak Modal Words in the 10-K Filings

		Dependent variable	
Independent variable	**Prediction**	***Uncertain***	***Weak_Modal***
Firm_Size	+	0.0001***	0.0001***
		(13.34)	(15.53)
MTB	+	0.0000	0.0001***
		(0.44)	(10.87)
Age	-	-0.0000***	-0.0000***
		(-28.35)	(-49.23)
Special_Items	-	-0.0001	0.0002*
		(-0.28)	(1.87)
Std_Ret	+	0.0008***	0.0017***
		(4.71)	(15.90)
Std_Earn	+	0.0005***	0.0007***
		(10.03)	(22.82)
Segments	-	-0.0007***	-0.0006***
		(-19.68)	(-30.27)
M&A	-	0.0001***	-0.0000*
		(2.78)	(-1.73)
SEO	-	0.0000	0.0000
		(0.63)	(1.29)
Delaware	?	-0.0006***	0.0001
		(-2.30)	(0.33)
Intercept		0.0096***	0.0031***
		(33.01)	(18.38)
Year/Ind Fixed Effects		Yes	Yes
n		37,442	37,442
Adj. R^2		0.2869	0.4180

The *t*-statistics in parentheses *, **, *** stands for $p < 0.1$, $p<0.05$, $p<0.01$, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 1.

TABLE 5

Descriptive Statistics for the Association between the Use of Uncertain and Weak Modal Words in 10-K Filings and Analyst Forecast Attributes

Variable	n	Mean	Median	Std. Dev.
Dependent				
#Analysts	25,673	7.5121	5.0000	6.5571
Dispersion	19,003	0.0143	0.0041	0.0308
Accuracy	15,213	0.0402	0.0024	0.1567
Uncertainty_Overall	14,308	0.0445	0.0076	0.1413
Uncertainty_Common	14,306	0.3259	0.3056	0.4299
Independent				
Uncertain	25,673	0.0134	0.0135	0.0030
Weak_Modal	25,673	0.0058	0.0056	0.0019
File_Size	25,673	12.7344	12.6999	0.4589
Pos_Tone	25,673	0.0074	0.0073	0.0017
Neg_Tone	25,673	0.0170	0.0169	0.0037
Control				
Firm_Size	25,673	6.4350	6.3575	1.8124
Growth	25,673	0.1565	0.0801	0.4815
Segments	25,673	1.0578	0.6931	0.4350
%_Inst	25,673	0.6263	0.6825	0.2642
10K_News	25,673	0.0355	0.0211	0.0418
Adv	25,673	0.0128	0.0000	0.0334
R&D	25,673	0.1114	0.0171	0.1913
Std_Ret	25,673	0.1407	0.1184	0.0848

Variables are defined in Appendix 2.

TABLE 6

Correlation Matrix for the Association between Analyst Following and the Use of Uncertain and Weak Modal Words in 10-K Filings

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*#Analysts*		0.0239*	0.0023	0.2899*	0.0056	0.0368*	0.7296*	0.0687*	0.1067*	0.4890*	-0.1812*	0.0542*	-0.0099	-0.2915*
2	*Uncertain*	0.0046		0.7400*	-0.2660*	0.2592*	0.2579*	-0.0581*	0.0319*	-0.1688*	0.0041	0.0248*	0.0291*	0.2246*	0.0391*
3	*Weak_Modal*	-0.0157*	0.7414*		0.0386*	0.2545*	0.4410*	-0.1610*	0.0701*	-0.3020*	-0.0788*	0.1035*	0.0371*	0.3733*	0.1890*
4	*File_Size*	0.2613*	-0.3052*	0.0318*		-0.0914*	0.3240*	0.2810*	0.0135*	0.1266*	0.1544*	-0.0201*	0.0335*	0.0644*	-0.0175*
5	*Pos_Tone*	-0.0206*	0.2445*	0.2798*	-0.0851*		0.0833*	-0.0117	0.0044	-0.1008*	-0.0442*	-0.0086	0.0102	0.3978*	0.0426*
6	*Neg_Tone*	0.0298*	0.2617*	0.4543*	0.3288*	0.0624*		-0.0962*	-0.0831*	-0.1136*	-0.0240*	0.0884*	0.0533*	0.2849*	0.1949*
7	*Firm_Size*	0.7225*	-0.0540*	-0.1507*	0.2783*	-0.0058	-0.0802*		0.1315*	0.2754*	0.5535*	-0.2756*	0.0221*	-0.0871*	-0.5015*
8	*Growth*	-0.0165*	0.0421*	0.1189*	0.0363*	0.0662*	-0.0098	0.0202*		-0.0723*	0.0804*	-0.0149*	0.0003	0.0487*	-0.0067
9	*Segments*	0.1148*	-0.1689*	-0.3021*	0.1321*	-0.1080*	-0.1081*	0.2898*	-0.0936*		0.1362*	-0.1186*	-0.0417*	-0.2106*	-0.2352*
10	*%_Inst*	0.3904*	0.0084	-0.0822*	0.1545*	-0.0445*	-0.0222*	0.5461*	-0.0248*	0.1497*		-0.1922*	0.0084	-0.1209*	-0.3281*
11	*10K_News*	-0.1591*	0.0234*	0.0898*	-0.0211*	0.0117	0.0800*	-0.2668*	0.0268*	-0.1021*	-0.2200*		-0.0111	0.0649*	0.3778*
12	*Adv*	0.0792*	0.0096	0.0364*	0.0543*	-0.0013	0.0442*	0.0593*	0.0077	-0.0117	-0.0104	0.0146*		-0.0290*	-0.0357*
13	*R&D*	-0.0312*	0.1790*	0.4060*	0.0909*	0.4447*	0.2096*	-0.1124*	0.1909*	-0.2867*	-0.1510*	0.0929*	-0.0872*		0.2098*
14	*Std_Ret*	-0.2516*	0.0014	0.1373*	-0.0174*	0.0627*	0.1578*	-0.4468*	0.0909*	-0.2033*	-0.3657*	0.3904*	-0.0054	0.2512*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 7

Multivariate Analysis for the Association between Analyst Following and the Use of Uncertain and Weak Modal Words in 10-K Filings

DV: *#Analysts*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Uncertain	+		24.5022***	28.5131***	6.7185***				
			(14.23)	(17.61)	(4.56)				
Weak_Modal	+						41.2014***	47.2850***	33.6183***
							(15.01)	(18.59)	(12.35)
File_Size	+	0.1377***	0.2228***	0.2538***		0.1377***	0.16222***	0.1795***	
		(14.21)	(19.62)	(24.83)		(14.21)	(16.58)	(19.87)	
Pos_Tone	+	10.6225***	5.3943**		2.8300	10.6225***	5.4277**		0.1571
		(4.09)	(2.06)		(1.08)	(4.09)	(2.08)		(0.06)
Neg_Tone	+	13.9095***	8.2424***		18.7498***	13.9095***	7.0756***		14.5946***
		(11.76)	(6.63)		(16.58)	(11.76)	(5.60)		(12.32)
Firm_Size	+	0.3494***	0.3465***	0.3446***	0.3607***	0.3494***	0.3526***	0.3519***	0.3638***
		(120.42)	(119.61)	(120.1)	(127.56)	(120.42)	(121.72)	(122.29)	(128.47)
Growth	+	-0.0646***	-0.0721***	-0.0756***	-0.0646***	-0.0646***	-0.0763***	-0.0799***	-0.0721***
		(-8.05)	(-9.00)	(-9.44)	(-8.01)	(-8.05)	(-9.5)	(-9.98)	(-8.93)
Segments	-	-0.1541***	-0.1477***	-0.1504***	-0.1326***	-0.1541***	-0.1313***	-0.1311***	-0.1162***
		(-15.87)	(-15.25)	(-15.54)	(-13.64)	(-15.87)	(-13.42)	(-13.4)	(-11.87)
%_Inst	+	0.6341***	0.6185***	0.6222***	0.6323***	0.6341***	0.6229***	0.6264***	0.6274***
		(35.21)	(34.42)	(34.63)	(34.95)	(35.21)	(34.71)	(34.92)	(34.78)
10-K_News	+	0.0031	-0.0233	-0.0154	0.0081	0.0031	-0.0256	-0.0199	-0.0083
		(0.03)	(-0.24)	(-0.16)	(0.08)	(0.03)	(-0.26)	(-0.2)	(-0.08)
Adv	+	0.8001***	0.7236***	0.7595***	0.8269***	0.8001***	0.6981***	0.7237***	0.7636***
		(6.69)	(6.06)	(6.36)	(6.88)	(6.69)	(5.85)	(6.06)	(6.37)
R&D	+	0.4146***	0.3767***	0.3977***	0.4398***	0.4146***	0.3127***	0.3231***	0.3663***
		(14.52)	(13.19)	(14.46)	(15.38)	(14.52)	(10.7)	(11.41)	(12.54)

TABLE 7 (Continued)

DV: *#Analysts*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Std_Ret	+	0.5631***	0.5176***	0.5562***	0.6350***	0.5631***	0.5261***	0.5593***	0.6155***
		(9.50)	(8.75)	(9.44)	(10.71)	(9.50)	(8.91)	(9.51)	(10.41)
Intercept		-3.3613***	-4.5274***	-4.8040***	-1.8063***	-3.3613***	-3.6967***	-3.8052***	-1.7992***
		(-20.62)	(-24.89)	(-27.67)	(-15.23)	(-20.62)	(-22.56)	(-24.12)	(-15.26)
Year/Ind. Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n		25,673	25,673	25,673	25,673	25,673	25,673	25,673	25,673
Adjusted R-Squared		0.6043	0.6074	0.6067	0.6015	0.6043	0.6078	0.6073	0.6036

The *t*-statistics in parentheses. *, **, *** stands for $p < 0.1$, $p<0.05$, $p<0.01$, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 8

Correlation Matrix for the Association between Analyst Forecast Dispersion and the Use of Uncertain and Weak Modal Words in 10-K Filings

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 *Dispersion*		0.0666*	0.1778*	0.1003*	-0.0221*	0.1804*	-0.4635*	-0.1842*	-0.1361*	-0.2993*	0.2421*	-0.1101*	0.0935*	0.4705*	-0.2203*
2 *Uncertain*	0.0207*		0.7384*	-0.3136*	0.2586*	0.2538*	-0.1133*	0.0367*	-0.1867*	-0.0263*	0.0387*	0.0247*	0.2273*	0.0679*	-0.0255*
3 *Weak_Modal*	0.1175*	0.7405*		-0.0086	0.2528*	0.4312*	-0.2360*	0.0829*	-0.3268*	-0.1098*	0.1214*	0.0427*	0.3825*	0.2251*	-0.0570*
4 *File_Size*	0.0805*	-0.3527*	-0.0142		-0.1063*	0.2857*	0.2500*	0.0008	0.1248*	0.1004*	-0.012	0.0249*	0.0707*	-0.0122	0.2489*
5 *Pos_Tone*	0.0395*	0.2410*	0.2784*	-0.0985*		0.0958*	-0.0406*	0.0048	-0.1083*	-0.0756*	-0.0118	0.0170*	0.4199*	0.0493*	-0.0386*
6 *Neg_Tone*	0.1362*	0.2578*	0.4430*	0.2908*	0.0685*		-0.1493*	-0.0785*	-0.1261*	-0.0614*	0.1010*	0.0671*	0.3088*	0.2134*	-0.0132
7 *Firm_Size*	-0.3924*	-0.1058*	-0.2217*	0.2464*	-0.0333*	-0.1235*		0.0925*	0.2780*	0.4524*	-0.2634*	0.0179*	-0.1057*	-0.5209*	0.6867*
8 *Growth*	0.0254*	0.0423*	0.1260*	0.0271*	0.0782*	-0.0053	-0.0159*		-0.0932*	0.0565*	0.0018	0.0106	0.0598*	0.0079	0.0339*
9 *Segments*	-0.1239*	-0.1869*	-0.3257*	0.1299*	-0.1158*	-0.1198*	0.2881*	-0.1081*		0.1167*	-0.1196*	-0.0598*	-0.2160*	-0.2339*	0.0996*
10 *%_Inst*	-0.3409*	-0.0231*	-0.1208*	0.0998*	-0.0773*	-0.0625*	0.4551*	-0.0476*	0.1309*		-0.1582*	0.0032	-0.1380*	-0.2972*	0.3769*
11 *10K_News*	0.2194*	0.0407*	0.1129*	-0.0150*	0.0126	0.0981*	-0.2562*	0.0407*	-0.1047*	-0.1947*		-0.0180*	0.0581*	0.3691*	-0.1586*
12 *Adv*	0.0008	0.0153*	0.0507*	0.0417*	-0.0062	0.0617*	0.0372*	0.0132	-0.0386*	-0.0292*	0.0106		-0.0147*	-0.0332*	0.0487*
13 *R&D*	0.1913*	0.1820*	0.4193*	0.0924*	0.4649*	0.2207*	-0.1421*	0.2025*	-0.2950*	-0.1718*	0.0982*	-0.0835*		0.2135*	-0.0365*
14 *Std_Ret*	0.4039*	0.0277*	0.1728*	-0.010	0.0721*	0.1795*	-0.4669*	0.1120*	-0.2069*	-0.3517*	0.3884*	0.0021	0.2621*		-0.2901*
15 *#Analysts*	-0.1493*	-0.0329*	-0.0580*	0.2334*	-0.0544*	-0.0014	0.6949*	-0.0381*	0.1037*	0.3048*	-0.1441*	0.0631*	-0.0527*	-0.2457*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 9

Multivariate Analysis for the Association between Analyst Forecast Dispersion and the Use of Uncertain and Weak Modal Words in 10-K Filings

DV: *Dispersion*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Uncertain	?		-0.1629*	-0.1050	-0.6781***				
			(-1.85)	(-1.26)	(-9.20)				
Weak_Modal	?						-1.4781***	-1.2375***	-1.7907***
							(-10.71)	(-9.56)	(-13.18)
File_Size	+	0.0067***	0.0061***	0.0063***		0.0067***	0.0057***	0.0062***	
		(13.94)	(10.62)	(12.00)		(13.94)	(11.63)	(13.53)	
Pos_Tone	+	0.2586***	0.2903**		0.2148	0.2586***	0.4201***		0.2287*
		(2.00)	(2.22)		(1.64)	(2.00)	(3.23)		(1.77)
Neg_Tone	+	0.0506	0.0859		0.3481***	0.0506	0.2760***		0.5171***
		(0.85)	(1.37)		(6.03)	(0.85)	(4.37)		(8.64)
Firm_Size	-	-0.0076***	-0.0076***	-0.0076***	-0.0073***	-0.0076***	-0.0078***	-0.0078***	-0.0075***
		(-39.58)	(-39.53)	(-39.94)	(-38.30)	(-39.58)	(-40.63)	(-40.95)	(-39.32)
Growth	-	-0.0008**	-0.0007*	-0.0008**	-0.0005	-0.0008**	-0.0004	-0.0005	-0.0002
		(-2.02)	(-1.89)	(-2.00)	(-1.34)	(-2.02)	(-0.99)	(-1.34)	(-0.55)
Segments	+	0.0025***	0.0025***	0.0025***	0.0029***	0.0025***	0.0018***	0.0018***	0.0023***
		(5.22)	(5.13)	(5.09)	(6.06)	(5.22)	(3.61)	(3.64)	(4.81)
%_Inst	-	-0.0193***	-0.0192***	-0.0193***	-0.0190***	-0.0193***	-0.0190***	-0.0190***	-0.0189***
		(-20.98)	(-20.89)	(-20.96)	(-20.60)	(-20.98)	(-20.7)	(-20.72)	(-20.58)
10K_News	+	0.0303***	0.0306***	0.0306***	0.0314***	0.0303***	0.0319***	0.0321***	0.0323***
		(5.91)	(5.96)	(5.96)	(6.10)	(5.91)	(6.23)	(6.27)	(6.28)
Adv	+	0.0197***	0.0202***	0.0206***	0.0231***	0.0197***	0.0237***	0.0250***	0.0263***
		(3.34)	(3.42)	(3.51)	(3.91)	(3.34)	(4.03)	(4.24)	(4.46)
R&D	+	0.0076***	0.0078***	0.0087***	0.0096***	0.0076***	0.0113***	0.0123***	0.0133***
		(5.36)	(5.51)	(6.43)	(6.81)	(5.36)	(7.78)	(8.78)	(9.23)

TABLE 9 (Continued)

DV: *Dispersion*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Std_Ret	+	0.0829***	0.0832***	0.0837***	0.0865***	0.0829***	0.0844***	0.0860***	0.0876***
		(25.80)	(25.87)	(26.15)	(26.92)	(25.80)	(26.34)	(26.92)	(27.33)
#Analysts	+	0.0010***	0.0010***	0.0010***	0.0010***	0.0010***	0.0010***	0.0010***	0.0010***
		(22.47)	(22.5)	(22.55)	(23.06)	(22.47)	(22.98)	(23.10)	(23.62)
Intercept		-0.0429***	-0.0346***	-0.0346***	0.0428***	-0.0429***	-0.0280***	-0.0282***	0.0404***
		(-4.39)	(-3.21)	(-3.30)	(5.37)	(-4.39)	(-2.85)	(-2.93)	(5.09)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n		19,003	19,003	19,003	19,003	19,003	19,003	19,003	19,003
Adjusted R-Squared		0.3302	0.3303	0.3302	0.3264	0.3302	0.3342	0.3333	0.3295

The *t*-statistics in parentheses. *, **, *** stands for $p < 0.1$, $p<0.05$, $p<0.01$, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 10

Correlation Matrix for the Association between Analyst Forecast Accuracy and the Use of Uncertain and Weak Modal Words in 10-K Filings

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Accuracy*		-0.0232*	0.0133	0.0738*	-0.0701*	0.0585*	-0.2256*	-0.1020*	-0.0084	-0.0982*	0.1238*	-0.0843*	-0.0417*	0.2844*	-0.1495*
2	*Uncertain*	-0.0475*		0.7313*	-0.3120*	0.2365*	0.2469*	-0.1099*	0.0404*	-0.1858*	-0.0135	0.0406*	0.0316*	0.2156*	0.0814*	-0.0158
3	*Weak_Modal*	-0.0259*	0.7343*		-0.0033	0.2171*	0.4227*	-0.2296*	0.1006*	-0.3275*	-0.0827*	0.1214*	0.0436*	0.3645*	0.2348*	-0.0512*
4	*File_Size*	0.0804*	-0.3496*	-0.0091		-0.1056*	0.2898*	0.2623*	-0.0132	0.1238*	0.1022*	-0.0117	0.0186*	0.0761*	-0.0320*	0.2493*
5	*Pos_Tone*	-0.0575*	0.2221*	0.2488*	-0.0974*		0.0992*	-0.0268*	-0.0003	-0.0874*	-0.0643*	-0.0224*	0.0324*	0.4192*	0.0426*	-0.0360*
6	*Neg_Tone*	0.0515*	0.2496*	0.4330*	0.2950*	0.0694*		-0.1137*	-0.0687*	-0.1177*	-0.0342*	0.0914*	0.0736*	0.3215*	0.1939*	0.0112
7	*Firm_Size*	-0.1589*	-0.1032*	-0.2157*	0.2588*	-0.0185*	-0.0873*		0.0332*	0.2721*	0.3869*	-0.2556*	0.0350*	-0.0784*	-0.5166*	0.7044*
8	*Growth*	-0.0056	0.0407*	0.1337*	0.0187*	0.0727*	0.0043	-0.0452*		-0.1222*	0.0227*	0.0278*	0.0124	0.0708*	0.0608*	-0.0131
9	*Segments*	0.0001	-0.1853*	-0.3272*	0.1274*	-0.0960*	-0.1109*	0.2818*	-0.1252*		0.0914*	-0.1172*	-0.0537*	-0.1952*	-0.2381*	0.0968*
10	*%_Inst*	-0.1285*	-0.0134	-0.0978*	0.1058*	-0.0657*	-0.0371*	0.3963*	-0.0712*	0.1103*		-0.1310*	0.0065	-0.1140*	-0.2573*	0.3539*
11	*10K_News*	0.0882*	0.0380*	0.1119*	-0.0149	0.0039	0.0854*	-0.2396*	0.0588*	-0.0999*	-0.1606*		-0.0218*	0.0494*	0.3724*	-0.1642*
12	*Adv*	-0.0129	0.0239*	0.0548*	0.0387*	0.0045	0.0626*	0.0630*	0.0208*	-0.0330*	-0.0205*	0.0069		-0.0005	-0.0395*	0.0640*
13	*R&D*	-0.0076	0.1748*	0.4135*	0.0946*	0.4515*	0.2317*	-0.1270*	0.2068*	-0.2900*	-0.1522*	0.0947*	-0.0753*		0.2063*	-0.0197*
14	*Std_Ret*	0.1989*	0.0445*	0.1866*	-0.0311*	0.0709*	0.1626*	-0.4507*	0.1489*	-0.2136*	-0.3212*	0.3846*	-0.0014	0.2662*		-0.3106*
15	*#Analysts*	-0.0686*	-0.0193*	-0.0459*	0.2291*	-0.0492*	0.0215*	0.7081*	-0.0612*	0.0967*	0.2961*	-0.1487*	0.0712*	-0.0477*	-0.2607*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 11

Multivariate Analysis for the Association between Analyst Forecast Accuracy and the Use of Uncertain and Weak Modal Words in 10-K Filings

DV: *Accuracy*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Uncertain	?		-1.2960**	-1.0987**	-3.6533***				
			(-2.22)	(-1.98)	(-7.49)				
Weak_Modal	?						-7.9428***	-6.9056***	-9.4055***
							(-8.62)	(-7.93)	(-10.37)
File_Size	+	0.0324***	0.0278***	0.0312***		0.0324***	0.0272***	0.0321***	
		(10.27)	(7.36)	(8.98)		(10.27)	(8.51)	(10.7)	
Pos_Tone	+	-1.9465**	-1.7223**		-2.1091**	-1.9465**	-1.2415		-2.1892**
		(-2.28)	(-2.00)		(-2.45)	(-2.28)	(-1.45)		(-2.57)
Neg_Tone	+	0.5120	0.7807*		1.9494***	0.5120	1.6509***		2.7824***
		(1.29)	(1.89)		(5.09)	(1.29)	(3.96)		(7.03)
Firm_Size	-	-0.0170***	-0.0169***	-0.0174***	-0.0155***	-0.0170***	-0.0183***	-0.0188***	-0.0168***
		(-13.26)	(-13.20)	(-13.71)	(-12.24)	(-13.26)	(-14.22)	(-14.72)	(-13.18)
Growth	-	-0.0056**	-0.0052**	-0.0054**	-0.0043*	-0.0056**	-0.0035	-0.0041	-0.0027
		(-2.19)	(-2.05)	(-2.12)	(-1.69)	(-2.19)	(-1.38)	(-1.60)	(-1.06)
Segments	+	0.0126***	0.0122***	0.0120***	0.0141***	0.0126***	0.0084***	0.0084***	0.0111***
		(3.98)	(3.86)	(3.78)	(4.46)	(3.98)	(2.63)	(2.61)	(3.47)
%_Inst	-	-0.0440***	-0.0432***	-0.0426***	-0.0414***	-0.0440***	-0.0417***	-0.0411***	-0.0407***
		(-7.26)	(-7.12)	(-7.02)	(-6.80)	(-7.26)	(-6.89)	(-6.78)	(-6.70)
10K_News	+	0.0186	0.0205	0.0231	0.0261	0.0186	0.0280	0.0313	0.0326
		(0.52)	(0.57)	(0.65)	(0.73)	(0.52)	(0.78)	(0.87)	(0.91)
Adv	+	-0.0305	-0.0258	-0.0217	-0.0123	-0.0305	-0.0056	0.0020	0.0057
		(-0.77)	(-0.65)	(-0.55)	(-0.31)	(-0.77)	(-0.14)	(0.05)	(0.14)
R&D	-	-0.0459***	-0.0438***	-0.0486***	-0.0354***	-0.0459***	-0.0258***	-0.0303***	-0.0158
		(-4.88)	(-4.64)	(-5.35)	(-3.77)	(-4.88)	(-2.67)	(-3.24)	(-1.64)

TABLE 11 (Continued)

DV: Accuracy		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Std_Ret	+	0.3427***	0.3459***	0.3493***	0.3602***	0.3427***	0.3539***	0.3615***	0.36756***
		(15.79)	(15.91)	(16.13)	(16.6)	(15.79)	(16.32)	(16.73)	(16.95)
#Analysts	+	0.0021***	0.0021***	0.0021***	0.0022***	0.0021***	0.0022***	0.0023***	0.0023***
		(7.71)	(7.76)	(7.98)	(8.06)	(7.71)	(8.13)	(8.41)	(8.49)
Intercept		-0.2776***	-0.2125***	-0.2571***	0.1362***	-0.2776***	-0.2003***	-0.2500***	0.1261**
		(-4.34)	(-3.02)	(-3.76)	(2.62)	(-4.34)	(-3.11)	(-3.97)	(2.43)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n		15,213	15,213	15,213	15,213	15,213	15,213	15,213	15,213
Adjusted R-Squared		0.0958	0.0961	0.0957	0.0929	0.0958	0.1002	0.0992	0.0959

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 12

Correlation Matrix for the Association between Uncertainty in Analysts' Overall Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Uncertainty_Overall*		0.0153	0.0950*	0.1251*	-0.0686*	0.1224*	-0.3314*	-0.1472*	-0.0583*	-0.1538*	0.1940*	-0.1216*	0.0075	0.4098*	-0.1563*
2	*Uncertain*	-0.0430*		0.7303*	-0.3266*	0.2354*	0.2462*	-0.1294*	0.0416*	-0.1905*	-0.0184*	0.0436*	0.0301*	0.2134*	0.0872*	-0.0363*
3	*Weak_Modal*	-0.0072	0.7331*		-0.0171*	0.2141*	0.4200*	-0.2570*	0.1042*	-0.3328*	-0.0862*	0.1248*	0.0436*	0.3616*	0.2431*	-0.0746*
4	*File_Size*	0.1048*	-0.3648*	-0.0223*		-0.1102*	0.2801*	0.2426*	-0.0136	0.1219*	0.0730*	-0.0047	0.0169*	0.0786*	-0.0217*	0.2243*
5	*Pos_Tone*	-0.0566*	0.2201*	0.2464*	-0.1020*		0.0987*	-0.0275*	0.0002	-0.0825*	-0.0664*	-0.0265*	0.0333*	0.4176*	0.0360*	-0.0455*
6	*Neg_Tone*	0.0655*	0.2494*	0.4303*	0.2850*	0.0686*		-0.1333*	-0.0673*	-0.1233*	-0.0440*	0.0965*	0.0754*	0.3241*	0.2023*	-0.0066
7	*Firm_Size*	-0.1827*	-0.1212*	-0.2411*	0.2384*	-0.0199*	-0.1042*		0.0257*	0.2774*	0.3286*	-0.2486*	0.0301*	-0.0806*	-0.5209*	0.6855*
8	*Growth*	-0.0008	0.0439*	0.1397*	0.0179*	0.0784*	0.0044	-0.0517*		-0.1258*	0.0237*	0.0295*	0.0151	0.0715*	0.0592*	-0.0218*
9	*Segments*	-0.0105	-0.1895*	-0.3326*	0.1244*	-0.0928*	-0.1171*	0.2860*	-0.1269*		0.0777*	-0.1154*	-0.0585*	-0.1910*	-0.2363*	0.0968*
10	*%_Inst*	-0.1378*	-0.0182*	-0.1045*	0.0739*	-0.0682*	-0.0472*	0.3390*	-0.0745*	0.0974*		-0.1113*	0.0032	-0.1144*	-0.2331*	0.2941*
11	*10K_News*	0.1078*	0.0421*	0.1149*	-0.0118	0.0015	0.0907*	-0.2349*	0.0618*	-0.0975*	-0.1462*		-0.0216*	0.0464*	0.3706*	-0.1489*
12	*Adv*	-0.0114	0.0231*	0.0538*	0.0378*	0.0048	0.0636*	0.0612*	0.0239*	-0.0339*	-0.0240*	0.0024		0.0042	-0.0371*	0.0553*
13	*R&D*	0.0125	0.1726*	0.4133*	0.0982*	0.4509*	0.2342*	-0.1321*	0.2083*	-0.2895*	-0.1504*	0.0943*	-0.0726*		0.2028*	-0.0272*
14	*Std_Ret*	0.2364*	0.0479*	0.1938*	-0.0190*	0.0651*	0.1717*	-0.4564*	0.1509*	-0.2120*	-0.3046*	0.3863*	-0.003	0.2634*		-0.3037*
15	*#Analysts*	-0.0620*	-0.0335*	-0.0613*	0.2087*	-0.0569*	0.0109	0.6940*	-0.0664*	0.0945*	0.2442*	-0.1380*	0.0690*	-0.0503*	-0.2519*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 13

Multivariate Analysis for the Association between Uncertainty in Analysts' Overall Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings

DV: *Uncertainty_Overall*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Uncertain	+		-0.9548*	-0.8239	-3.5870***				
			(-1.78)	(-1.61)	(-8.04)				
Weak_Modal	+						-7.1363***	-6.2906***	-8.7673***
							(-8.46)	(-7.86)	(-10.55)
File_Size	+	0.0340***	0.0306***	0.0332***		0.0340***	0.0292	0.0332***	
		(11.82)	(8.82)	(10.36)		(11.82)	(9.99)	(12.08)	
Pos_Tone	-	-1.6522**	-1.4938*		-1.9572**	-1.6522**	-1.0359		-2.0954***
		(-2.12)	(-1.90)		(-2.49)	(-2.12)	(-1.32)		(-2.69)
Neg_Tone	+	0.3846	0.5780		1.8250***	0.3846	1.3795***		2.5590***
		(1.06)	(1.52)		(5.17)	(1.06)	(3.62)		(7.04)
Firm_Size	-	-0.0197***	-0.0196***	-0.0200***	-0.0182***	-0.0197***	-0.0210***	-0.0214***	-0.0195***
		(-16.7)	(-16.67)	(-17.17)	(-15.54)	(-16.7)	(-17.71)	(-18.20)	(-16.51)
Growth	-	-0.0052**	-0.0050**	-0.0051**	-0.0038	-0.0052**	-0.0032	-0.0037	-0.0022
		(-2.17)	(-2.05)	(-2.12)	(-1.58)	(-2.17)	(-1.31)	(-1.53)	(-0.91)
Segments	+	0.0113***	0.0110***	0.0108***	0.0130***	0.0113***	0.0076***	0.0075**	0.0104***
		(3.91)	(3.81)	(3.74)	(4.51)	(3.91)	(2.59)	(2.56)	(3.59)
%_Inst	-	-0.0403***	-0.0398***	-0.0393***	-0.0384***	-0.0403***	-0.0388***	-0.0382***	-0.0382***
		(-7.10)	(-7.01)	(-6.92)	(-6.74)	(-7.10)	(-6.84)	(-6.75)	(-6.72)
10K_News	+	0.0335	0.0351	0.0375	0.0405	0.0335	0.0422	0.0455	0.0454
		(1.00)	(1.05)	(1.12)	(1.21)	(1.00)	(1.26)	(1.36)	(1.36)
Adv	+	0.0050	0.0085	0.0118	0.0241	0.0050	0.0266	0.0336	0.0394
		(0.14)	(0.23)	(0.33)	(0.66)	(0.14)	(0.73)	(0.93)	(1.08)
R&D	-	-0.0394***	-0.0379***	-0.0421***	-0.0286***	-0.0394***	-0.0213**	-0.0251***	-0.0102
		(-4.56)	(-4.38)	(-5.05)	(-3.32)	(-4.56)	(-2.41)	(-2.93)	(-1.15)

TABLE 13 (Continued)

DV: *Uncertainty_Overall*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Std_Ret	+	0.3636***	0.3661***	0.3687***	0.3832***	0.3636***	0.3739***	0.3808***	0.3900***
		(17.58)	(17.66)	(17.87)	(18.52)	(17.58)	(18.09)	(18.50)	(18.86)
#Analysts	+	0.0026***	0.0026***	0.0026***	0.0027***	0.0026***	0.0027***	0.0028***	0.0028***
		(10.59)	(10.62)	(10.84)	(10.90)	(10.59)	(11.02)	(11.3)	(11.37)
Intercept		-0.2703***	-0.2212***	-0.2565***	0.1666***	-0.2703***	-0.1972***	-0.2375***	0.1565***
		(-4.54)	(-3.37)	(-4.01)	(3.41)	(-4.54)	(-3.28)	(-4.04)	(3.22)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n		14,308	14,308	14,308	14,308	14,308	14,308	14,308	14,308
Adjusted R-Squared		0.1312	0.1313	0.1311	0.1266	0.1312	0.1355	0.1347	0.1295

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 14

Correlation Matrix for the Association between Uncertainty in Analysts' Common Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Uncertainty_Common*		-0.0726*	-0.0959*	0.0574*	-0.0675*	-0.0253*	0.0943*	-0.0287*	0.0818*	0.0950*	-0.0207*	0.004	-0.0991*	0.0239*	0.1107*
2	*Uncertain*	-0.0704*		0.7302*	-0.3267*	0.2354*	0.2459*	-0.1292*	0.0415*	-0.1904*	-0.0184*	0.0438*	0.0299*	0.2135*	0.0870*	-0.0360*
3	*Weak_Modal*	-0.0947*	0.7330*		-0.0172*	0.2142*	0.4198*	-0.2568*	0.1041*	-0.3327*	-0.0862*	0.1250*	0.0435*	0.3617*	0.2429*	-0.0743*
4	*File_Size*	0.0600*	-0.3649*	-0.0223*		-0.1100*	0.2801*	0.2427*	-0.0135	0.1219*	0.0728*	-0.0048	0.0169*	0.0787*	-0.0217*	0.2244*
5	*Pos_Tone*	-0.0710*	0.2202*	0.2465*	-0.1019*		0.0988*	-0.0275*	0.0002	-0.0826*	-0.0663*	-0.0264*	0.0334*	0.4176*	0.0361*	-0.0456*
6	*Neg_Tone*	-0.0181*	0.2491*	0.4301*	0.2850*	0.0687*		-0.1330*	-0.0674*	-0.1232*	-0.0440*	0.0967*	0.0753*	0.3242*	0.2020*	-0.0062
7	*Firm_Size*	0.0942*	-0.1210*	-0.2409*	0.2385*	-0.0200*	-0.1039*		0.0258*	0.2773*	0.3287*	-0.2488*	0.0303*	-0.0806*	-0.5208*	0.6854*
8	*Growth*	-0.0081	0.0439*	0.1396*	0.0179*	0.0784*	0.0043	-0.0517*		-0.1257*	0.0238*	0.0297*	0.0151	0.0715*	0.0591*	-0.0217*
9	*Segments*	0.0810*	-0.1894*	-0.3325*	0.1244*	-0.0928*	-0.1169*	0.2859*	-0.1269*		0.0776*	-0.1155*	-0.0584*	-0.1910*	-0.2362*	0.0967*
10	*%_Inst*	0.1093*	-0.0182*	-0.1045*	0.0738*	-0.0681*	-0.0472*	0.3391*	-0.0745*	0.0974*		-0.1114*	0.0032	-0.1144*	-0.2331*	0.2942*
11	*10K_News*	-0.0184*	0.0422*	0.1150*	-0.0119	0.0015	0.0908*	-0.2349*	0.0618*	-0.0975*	-0.1463*		-0.0216*	0.0464*	0.3708*	-0.1491*
12	*Adv*	0.0068	0.0231*	0.0538*	0.0378*	0.0049	0.0636*	0.0612*	0.0239*	-0.0339*	-0.0241*	0.0023		0.0042	-0.0372*	0.0555*
13	*R&D*	-0.1151*	0.1726*	0.4134*	0.0982*	0.4509*	0.2342*	-0.1321*	0.2083*	-0.2895*	-0.1504*	0.0942*	-0.0726*		0.2029*	-0.0273*
14	*Std_Ret*	0.0192*	0.0478*	0.1936*	-0.0190*	0.0652*	0.1715*	-0.4563*	0.1509*	-0.2119*	-0.3046*	0.3863*	-0.003	0.2634*		-0.3035*
15	*#Analysts*	0.1070*	-0.0332*	-0.0610*	0.2089*	-0.0570*	0.0113	0.6939*	-0.0664*	0.0944*	0.2444*	-0.1381*	0.0690*	-0.0503*	-0.2518*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 15

Multivariate Analysis for the Association between Uncertainty in Analysts' Common Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings

DV: *Uncertainty_Common*		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Uncertain	+		-1.3166	-1.1622	-5.4532***				
			(-0.78)	(-0.72)	(-3.87)				
Weak_Modal	+						-6.3455**	-5.5240**	-9.0535***
							(-2.37)	(-2.18)	(-3.44)
File_Size	+	0.0528***	0.0480***	0.0531***		0.0528***	0.0485***	0.0546***	
		(5.79)	(4.38)	(5.24)		(5.79)	(5.23)	(6.27)	
Pos_Tone	-	-4.0846*	-3.8662		-4.5927*	-4.0846*	-3.5367		-5.2947**
		(-1.65)	(-1.55)		(-1.85)	(-1.65)	(-1.43)		(-2.15)
Neg_Tone	+	0.7664	1.0327		2.9909***	0.7664	1.6502		3.6080***
		(0.67)	(0.86)		(2.68)	(0.67)	(1.37)		(3.14)
Firm_Size	-	0.0040	0.0040	0.0032	0.0063*	0.0040	0.0028	0.0019	0.0053
		(1.07)	(1.08)	(0.86)	(1.71)	(1.07)	(0.75)	(0.52)	(1.43)
Growth	-	0.0120	0.0124	0.0122	0.0142*	0.0120	0.0139*	0.0133*	0.0154**
		(1.57)	(1.62)	(1.59)	(1.85)	(1.57)	(1.8)	(1.73)	(2.01)
Segments	+	0.0217**	0.0213**	0.0209**	0.0245***	0.0217**	0.0184**	0.0182**	0.0232**
		(2.38)	(2.33)	(2.29)	(2.69)	(2.38)	(1.99)	(1.97)	(2.52)
%_Inst	+	0.1767***	0.1774***	0.1786***	0.1796***	0.1767***	0.1781***	0.1794***	0.1790***
		(9.83)	(9.86)	(9.93)	(9.98)	(9.83)	(9.90)	(9.98)	(9.94)
10K_News	+	-0.2754***	-0.2732***	-0.2678**	-0.2647**	-0.2754***	-0.2677**	-0.2617**	-0.2622**
		(-2.60)	(-2.58)	(-2.53)	(-2.5)	(-2.60)	(-2.53)	(-2.47)	(-2.47)
Adv	+	0.0253	0.0301	0.0360	0.0545	0.0253	0.0445	0.0528	0.0657
		(0.22)	(0.26)	(0.31)	(0.47)	(0.22)	(0.39)	(0.46)	(0.57)
R&D	-	-0.2476***	-0.2456***	-0.2566***	-0.2310***	-0.2476***	-0.2316***	-0.2427***	-0.2131***
		(-9.07)	(-8.95)	(-9.73)	(-8.48)	(-9.07)	(-8.23)	(-8.93)	(-7.63)

TABLE 15 (Continued)

DV: ***Uncertainty_Common***		***Uncertain***				***Weak_Modal***			
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Std_Ret	+	0.5739***	0.5773***	0.5817***	0.6042***	0.5739***	0.5831***	0.5905***	0.6098***
		(8.77)	(8.80)	(8.91)	(9.24)	(8.77)	(8.89)	(9.05)	(9.32)
#Analysts	+	0.0054***	0.0054***	0.0055***	0.0055***	0.0054***	0.0055***	0.0056***	0.0056***
		(6.99)	(7.00)	(7.15)	(7.15)	(6.99)	(7.10)	(7.25)	(7.29)
Intercept		-0.0658	0.0018	-0.0743	0.6113***	-0.0658	-0.0008	-0.0800	0.5864***
		(-0.35)	(0.01)	(-0.37)	(3.96)	(-0.35)	(0.00)	(-0.43)	(3.81)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
n		14,306	14,306	14,306	14,306	14,306	14,306	14,306	14,306
Adjusted R-Squared		0.0604	0.0603	0.0603	0.0591	0.0604	0.0607	0.0605	0.0589

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

APPENDIX 1

Variable Definitions for Determinants of the Use of Uncertain and Weak Modal Words in 10-K Filings

Variable	Definition
Uncertain	The proportion of uncertain words to the total words in 10-K filings as defined in Loughran and McDonald (2011).
Weak_Modal	The proportion of weak modal words to the total words in 10-K filings as defined in Loughran and McDonald (2011).
Firm_Size	The logarithm of the market value of equity at the fiscal year-end.
MTB	The ratio of the market value of equity plus book value of liability to the book value of total assets at the fiscal year-end.
Age	The number of years of firms that appear in the CRSP monthly stock return database.
Special_Items	The amount of special items divided by the book value of assets.
Std_Ret	The standard deviation of the monthly stock returns in the prior year.
Std_Earn	The standard deviation of the operating earnings during the past five fiscal years.
Segments	The logarithm of the number of business segments.
M&A	1 for firms that engage in M&A as an acquirer in a specific firm-year and 0 otherwise.
SEO	1 for firms that have the seasoned equity offering in a specific firm-year and 0 otherwise.
Delaware	1 for firms that are incorporated in Delaware state and 0 otherwise.

APPENDIX 2

Variable Definitions for the Association between the Use of Uncertain and Weak Modal Words in 10-K Filings and Analyst Forecast Attributes

Variable	Definition
#Analysts	The number of analysts included in the first I/B/E/S consensus earnings forecast after 10-K filings.
Dispersion	The standard deviation of the individual analyst forecasts in the first analyst consensus earnings forecast after the 10-K reporting, scaled by share price 90 days before the consensus earnings forecast.
Accuracy	The squared value of the difference between the reported earnings in I/B/E/S and the most recent analyst consensus forecast, scaled by stock price 90 days before the consensus earnings forecast.
Uncertainty_Overall	The sum of uncertainty related to analysts' private information and uncertainty related to common (public) information to all analysts.
Uncertainty_Common	The ratio of common uncertainty to overall uncertainty, and it measures the average analyst's belief reflects between common and private information.
Uncertain	The proportion of uncertain words to the total words in 10-K filings as defined in Loughran and McDonald (2011).
Weak_Modal	The proportion of weak modal words to the total words in 10-K filings as defined in Loughran and McDonald (2011).
File_Size	The natural logarithm of the net file size of 10-K filings.
Fog	The Fog Index of the 10-K filing calculated as (average words per sentence + percent of complex words) × 0.4.
Pos_Tone	The proportion of positive words to the total number of words in 10-K filings as defined in Loughran and McDonald (2011).
Neg_Tone	The proportion of negative words to the total number of words in 10-K filings as defined in Loughran and McDonald (2011).
Std_Ret	The standard deviation of the monthly stock returns in the prior year.
Firm_Size	The logarithm of the market value of equity at the fiscal year-end.
Growth	The difference of sales volume between the current year and prior year divided by prior year sales volume.

Segments	The logarithm of the number of business segments.
Adv	Advertising expense divided by operating expense.
R&D	Research and development expense divided by operating expense.
%_Inst	The percentage of a firm's shares that are held by institutional investors.
10K_News	Two-day event window for market-adjusted return to control the informativeness of the 10-K filing.

APPENDIX 3

Additional Tests

TABLE 16

Correlation Matrix for the Association between Analyst Following and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog Replacement

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	*#Analysts*		0.0021	-0.0223*	0.0291*	0.0114	0.0239*	0.7247*	0.0827*	0.0978*	0.4899*	-0.1653*	0.0576*	-0.0037	-0.2800*
2	*Uncertain*	-0.0182*		0.6796*	-0.0349*	0.2530*	0.1874*	-0.0795*	0.0314*	-0.1592*	0.0347*	0.0442*	0.0417*	0.2493*	0.0998*
3	*Weak_Modal*	-0.0413*	0.6849*		0.1251*	0.2267*	0.3975*	-0.1855*	0.0683*	-0.2843*	-0.0525*	0.1263*	0.0502*	0.4173*	0.2621*
4	*Fog*	0.0442*	-0.0805*	0.0863*		0.0091	0.1489*	-0.0026	0.0065	-0.0131	0.0373*	0.0002	-0.0049	0.0783*	0.008
5	*Pos_Tone*	-0.0092	0.2359*	0.2506*	-0.0091		0.0584*	-0.0051	0.0046	-0.0883*	-0.0349*	0.0039	0.0131	0.3777*	0.0553*
6	*Neg_Tone*	0.0290*	0.1942*	0.4039*	0.1349*	0.0346*		-0.1133*	-0.0953*	-0.0808*	-0.0069	0.1129*	0.0701*	0.2964*	0.2556*
7	*Firm_Size*	0.7217*	-0.0766*	-0.1759*	0.0230*	0.0069	-0.0962*		0.1591*	0.2615*	0.5553*	-0.2604*	0.0163*	-0.0983*	-0.4915*
8	*Growth*	0.0017	0.0426*	0.1179*	0.0160*	0.0513*	-0.0312*	0.0523*		-0.0599*	0.0819*	-0.0358*	-0.0256*	0.0533*	-0.0552*
9	*Segments*	0.0992*	-0.1612*	-0.2863*	0.0014	-0.0951*	-0.0799*	0.2753*	-0.0803*		0.1266*	-0.1153*	-0.0501*	-0.1868*	-0.2166*
10	*%_Inst*	0.3896*	0.0296*	-0.0644*	0.0419*	-0.0352*	-0.0089	0.5464*	-0.0116	0.1405*		-0.2010*	0.0129	-0.1359*	-0.3609*
11	*10K_News*	-0.1429*	0.0422*	0.1098*	-0.0047	0.0164*	0.0947*	-0.2551*	0.0178*	-0.1000*	-0.2207*		-0.0054	0.1001*	0.3781*
12	*Adv*	0.0907*	0.0118	0.0315*	-0.0384*	0.0066	0.0307*	0.0627*	0.0199*	-0.0181*	-0.0285*	0.0038		-0.001	-0.0144
13	*R&D*	-0.0138	0.1942*	0.4347*	0.0784*	0.4110*	0.2075*	-0.1073*	0.1826*	-0.2681*	-0.1538*	0.1055*	-0.0750*		0.2399*
14	*Std_Ret*	-0.2337*	0.0621*	0.2111*	-0.0170*	0.0665*	0.2096*	-0.4322*	0.0624*	-0.1904*	-0.3873*	0.4090*	0.0024	0.2421*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 17

Multivariate Analysis for the Association between Analyst Following and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

DV: *#Analysts*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Uncertain	+		8.2556***	10.3344***			
			(4.35)	(5.53)			
Weak_Modal	+					28.8694***	39.2189***
						(7.99)	(11.37)
Fog	+	0.0024	0.0048	0.0090***	0.0024	0.0028	0.0059*
		(0.76)	(1.51)	(2.84)	(0.76)	(0.91)	(1.89)
Pos_Tone	+	4.4712	1.8548		4.4712	1.0572	
		(1.39)	(0.57)		(1.39)	(0.33)	
Neg_Tone	+	18.1086***	17.5328***		18.1086***	14.7657***	
		(13.04)	(12.57)		(13.04)	(10.20)	
Firm_Size	+	0.3681***	0.3693***	0.3684***	0.3681***	0.3712***	0.3713***
		(101.45)	(101.56)	(101.17)	(101.45)	(101.93)	(101.96)
Growth	+	-0.0838***	-0.0865***	-0.0952***	-0.0838***	-0.0923***	-0.1015***
		(-7.26)	(-7.49)	(-8.22)	(-7.26)	(-7.98)	(-8.78)
Segments	-	-0.1369***	-0.1319***	-0.1310***	-0.1369***	-0.1214***	-0.1161***
		(-11.19)	(-10.73)	(-10.61)	(-11.19)	(-9.82)	(-9.36)
%_Inst	+	0.6982***	0.6942***	0.7129***	0.6982***	0.6927***	0.7064***
		(29.92)	(29.74)	(30.47)	(29.92)	(29.73)	(30.28)
10-K_News	+	0.1013	0.0901	0.1217	0.1013	0.0869	0.1087
		(0.79)	(0.70)	(0.94)	(0.79)	(0.68)	(0.85)
Adv	+	0.8616***	0.8393***	0.9808***	0.8616***	0.7809***	0.8663***
		(5.12)	(4.99)	(5.81)	(5.12)	(4.64)	(5.14)
R&D	+	0.5042***	0.4924***	0.5282***	0.5042***	0.4338***	0.4357***
		(14.16)	(13.80)	(15.21)	(14.16)	(11.85)	(12.15)

TABLE 17 (Continued)

DV: *#Analysts*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Std_Ret	+	0.4887***	0.4820***	0.6396***	0.4887***	0.4621***	0.5766***
		(6.87)	(6.78)	(9.10)	(6.87)	(6.50)	(8.19)
Intercept		-2.1011***	-2.2130***	-2.1036	-2.1011***	-2.1539***	-2.0736***
		(-12.85)	(-13.38)	(-12.76)	(-12.85)	(-13.19)	(-12.8)
Year/Ind. Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
n		15,898	15,898	15,898	15,898	15,898	15,898
Adjusted R-Squared		0.6010	0.6015	0.5975	0.6010	0.6026	0.6000

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 18

Correlation Matrix for the Association between Analyst Forecast Dispersion and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Dispersion*		0.0795*	0.1851*	0.0272*	-0.0211*	0.2068*	-0.4284*	-0.1949*	-0.1236*	-0.2746*	0.2284*	-0.0740*	0.1067*	0.4645*	-0.1795*
2	*Uncertain*	0.0163		0.6690*	-0.0514*	0.2416*	0.1722*	-0.1300*	0.0378*	-0.1780*	0.0135	0.0474*	0.0317*	0.2466*	0.1235*	-0.0443*
3	*Weak_Modal*	0.1040*	0.6758*		0.1113*	0.2172*	0.3808*	-0.2623*	0.0843*	-0.3094*	-0.0778*	0.1406*	0.0484*	0.4230*	0.2965*	-0.0870*
4	*Fog*	0.0085	-0.0955*	0.0747*		0.0169	0.1449*	0.0009	0.0082	-0.0026	0.0473*	-0.0135	-0.005	0.0763*	0.0026	0.0393*
5	*Pos_Tone*	0.0139	0.2221*	0.2399*	-0.0036		0.0674*	-0.0360*	0.0015	-0.0920*	-0.0673*	-0.0027	0.0237*	0.3861*	0.0650*	-0.0316*
6	*Neg_Tone*	0.1434*	0.1812*	0.3863*	0.1289*	0.0396*		-0.1618*	-0.0935*	-0.0929*	-0.0398*	0.1218*	0.0821*	0.3175*	0.2750*	-0.0165
7	*Firm_Size*	-0.3675*	-0.1226*	-0.2469*	0.0244*	-0.0230*	-0.1357*		0.1107*	0.2719*	0.4550*	-0.2584*	0.0151	-0.1250*	-0.5124*	0.6845*
8	*Growth*	-0.0186*	0.0489*	0.1321*	0.0124	0.0558*	-0.0291*	0.0089		-0.0774*	0.0437*	-0.0171	-0.016	0.0617*	-0.0416*	0.0383*
9	*Segments*	-0.1138*	-0.1796*	-0.3097*	0.0124	-0.1002*	-0.0919*	0.2793*	-0.0925*		0.1060*	-0.1243*	-0.0538*	-0.1898*	-0.2168*	0.0921*
10	*%_Inst*	-0.3239*	0.0079	-0.0962*	0.0516*	-0.0685*	-0.0448*	0.4582*	-0.0475*	0.1203*		-0.1768*	0.0129	-0.1584*	-0.3432*	0.3776*
11	*10K_News*	0.2349*	0.0476*	0.1303*	-0.0108	0.0175	0.1115*	-0.2549*	0.0373*	-0.1076*	-0.2093*		-0.0126	0.1011*	0.3759*	-0.1515*
12	*Adv*	0.0281*	0.0173	0.0449*	-0.0363*	0.0045	0.0497*	0.0452*	0.0305*	-0.0332*	-0.0518*	0.0068		0.0114	-0.0051	0.0480*
13	*R&D*	0.1529*	0.1928*	0.4466*	0.0753*	0.4215*	0.2167*	-0.1388*	0.1865*	-0.2743*	-0.1728*	0.1120*	-0.0740*		0.2479*	-0.0327*
14	*Std_Ret*	0.4207*	0.0827*	0.2428*	-0.0201*	0.0752*	0.2324*	-0.4560*	0.0819*	-0.1937*	-0.3947*	0.4201*	0.0148	0.2456*		-0.2807*
15	*#Analysts*	-0.1264*	-0.0554*	-0.0866*	0.0486*	-0.0424*	0.0073	0.6966*	-0.0247*	0.0892*	0.3028*	-0.1336*	0.0778*	-0.0300*	-0.2270*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 19

Multivariate Analysis for the Association between Analyst Forecast Dispersion and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog Replacement

DV: *Dispersion*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Uncertain	?		-0.5170***	-0.4895***			
			(-5.91)	(-5.69)			
Weak_Modal	?					-1.5384***	-1.2926***
						(-9.23)	(-8.03)
Fog	+	0.0001	-0.0000	0.0000	0.0001	0.0001	0.0002
		(0.97)	(-0.14)	(0.25)	(0.97)	(0.71)	(1.22)
Pos_Tone	+	-0.1134	0.0362		-0.1134	0.0367	
		(-0.77)	(0.24)		(-0.77)	(0.25)	
Neg_Tone	+	0.2514***	0.2815***		0.2514***	0.4049***	
		(3.87)	(4.32)		(3.87)	(6.06)	
Firm_Size	-	-0.0061***	-0.0062***	-0.0062***	-0.0061***	-0.0063***	-0.0064***
		(-26.60)	(-26.97)	(-27.49)	(-26.60)	(-27.59)	(-28.04)
Growth	-	-0.0019***	-0.0017***	-0.0018***	-0.0019***	-0.0014***	-0.0016***
		(-3.64)	(-3.30)	(-3.54)	(-3.64)	(-2.75)	(-3.18)
Segments	+	0.0023***	0.0020***	0.0020***	0.0023***	0.0015***	0.0017***
		(4.10)	(3.50)	(3.56)	(4.10)	(2.68)	(2.96)
%_Inst	-	-0.0176***	-0.0174***	-0.0172***	-0.0176***	-0.0173***	-0.0171***
		(-15.89)	(-15.73)	(-15.61)	(-15.89)	(-15.73)	(-15.55)
10-K_News	+	0.0379***	0.0384***	0.0390***	0.0379***	0.0387***	0.0395***
		(6.13)	(6.22)	(6.32)	(6.13)	(6.30)	(6.41)
Adv	+	0.0337***	0.0351***	0.0377***	0.0337***	0.0385***	0.0413***
		(4.51)	(4.70)	(5.06)	(4.51)	(5.16)	(5.55)
R&D	+	0.0055***	0.0062***	0.0066***	0.0055***	0.0093***	0.0093***
		(3.37)	(3.82)	(4.23)	(3.37)	(5.58)	(5.71)

TABLE 19 (Continued)

DV: *Dispersion*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Std_Ret	+	0.0841***	0.0847***	0.0873***	0.0841***	0.0858***	0.0891***
		(24.01)	(24.20)	(25.31)	(24.01)	(24.53)	(25.76)
#Analysts	+	0.0009***	0.0009***	0.0010***	0.0009***	0.0009***	0.0010***
		(17.76)	(17.79)	(18.33)	(17.76)	(17.98)	(18.63)
Intercept		0.0319***	0.0393***	0.0419***	0.0319***	0.0365***	0.0397***
		(3.51)	(4.30)	(4.62)	(3.51)	(4.03)	(4.42)
Year/Ind. Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
n		11,820	11,820	11,820	11,820	11,820	11,820
Adjusted R-Squared		0.3151	0.317	0.3161	0.3151	0.3199	0.3179

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 20

Correlation Matrix for the Association between Analyst Forecast Accuracy and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Accuracy*		-0.0173	0.0118	-0.0065	-0.0540*	0.0637*	-0.1967*	-0.0964*	-0.0063	-0.0843*	0.1071*	-0.0655*	-0.0465*	0.2702*	-0.1164*
2	*Uncertain*	-0.0390*		0.6639*	-0.0606*	0.2203*	0.1698*	-0.1270*	0.0488*	-0.1739*	0.0267*	0.0492*	0.0399*	0.2334*	0.1471*	-0.0409*
3	*Weak_Modal*	-0.0275*	0.6742*		0.1090*	0.1842*	0.3743*	-0.2503*	0.1073*	-0.3072*	-0.0494*	0.1344*	0.0523*	0.4042*	0.3126*	-0.0793*
4	*Fog*	0.0108	-0.1027*	0.0739*		0.0121	0.1479*	0.011	0.0013	-0.002	0.0552*	-0.0184	0.0003	0.0667*	0.0033	0.0420*
5	*Pos_Tone*	-0.0567*	0.2034*	0.2112*	-0.0098		0.0733*	-0.0285*	-0.0006	-0.0686*	-0.0541*	-0.0119	0.0330*	0.3755*	0.0686*	-0.0336*
6	*Neg_Tone*	0.0647*	0.1778*	0.3797*	0.1310*	0.0413*		-0.1266*	-0.0805*	-0.0800*	-0.0129	0.1020*	0.0956*	0.3320*	0.2562*	0.0071
7	*Firm_Size*	-0.1509*	-0.1223*	-0.2389*	0.0317*	-0.0133	-0.1015*		0.0524*	0.2652*	0.3949*	-0.2389*	0.0320*	-0.0938*	-0.5011*	0.7030*
8	*Growth*	-0.0247*	0.0487*	0.1398*	0.0002	0.0503*	-0.0155	-0.0235*		-0.1072*	0.0133	0.0099	-0.0139	0.0717*	0.0145	-0.0111
9	*Segments*	-0.0023	-0.1764*	-0.3075*	0.0126	-0.0785*	-0.0798*	0.2724*	-0.1105*		0.0859*	-0.1206*	-0.0506*	-0.1633*	-0.2254*	0.0901*
10	*%_Inst*	-0.1049*	0.0129	-0.0743*	0.0579*	-0.0518*	-0.0183	0.4034*	-0.0735*	0.1048*		-0.1469*	0.0167	-0.1295*	-0.3043*	0.3591*
11	*10K_News*	0.0705*	0.0489*	0.1245*	-0.0163	0.012	0.0897*	-0.2235*	0.0619*	-0.1048*	-0.1679*		-0.0135	0.0848*	0.3750*	-0.1488*
12	*Adv*	0.0002	0.0287*	0.0531*	-0.0282*	0.0143	0.0615*	0.0805*	0.0386*	-0.0259*	-0.0374*	0.0032		0.0248*	-0.003	0.0738*
13	*R&D*	-0.0221*	0.1823*	0.4374*	0.0669*	0.4006*	0.2256*	-0.1188*	0.1897*	-0.2665*	-0.1474*	0.1072*	-0.0647*		0.2412*	-0.0122
14	*Std_Ret*	0.1989*	0.1067*	0.2646*	-0.0211*	0.0841*	0.2148*	-0.4303*	0.1314*	-0.2040*	-0.3573*	0.4074*	0.0139	0.2553*		-0.2918*
15	*#Analysts*	-0.0411*	-0.0509*	-0.0792*	0.0490*	-0.0399*	0.0269*	0.7113*	-0.0507*	0.0869*	0.2956*	-0.1294*	0.1000*	-0.0217*	-0.2359*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for p<0.05 two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 21

Multivariate Analysis for the Association between Analyst Forecast Accuracy and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

DV: *Accuracy*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Uncertain	?		-3.3171***	-3.3191***			
			(-5.59)	(-5.66)			
Weak_Modal	?					-9.9156***	-8.5436***
						(-8.74)	(-7.75)
Fog	+	0.0011	0.0000	0.0004	0.0011	0.0009	0.0014
		(1.11)	(0.04)	(0.45)	(1.11)	(0.90)	(1.47)
Pos_Tone	+	-2.6441***	-1.7887***		-2.6441***	-1.8753	
		(-2.65)	(-1.77)		(-2.65)	(-1.88)	
Neg_Tone	+	1.6933***	1.8746***		1.6933***	2.6260***	
		(3.84)	(4.24)		(3.84)	(5.81)	
Firm_Size	-	-0.0180***	-0.0185***	-0.0192***	-0.0180***	-0.0196***	-0.0204***
		(-11.52)	(-11.84)	(-12.38)	(-11.52)	(-12.53)	(-13.03)
Growth	-	-0.0084**	-0.0074**	-0.0078**	-0.0084**	-0.0056	-0.0066*
		(-2.44)	(-2.15)	(-2.29)	(-2.44)	(-1.63)	(-1.93)
Segments	+	0.0149***	0.0127***	0.0132***	0.0149***	0.0098**	0.0110***
		(3.96)	(3.37)	(3.48)	(3.96)	(2.57)	(2.91)
%_Inst	-	-0.0326***	-0.0313***	-0.0296***	-0.0326***	-0.0307***	-0.0288***
		(-4.34)	(-4.17)	(-3.94)	(-4.34)	(-4.09)	(-3.84)
10K_News	+	-0.0495	-0.0464	-0.0412	-0.0495	-0.0449	-0.0390***
		(-1.11)	(-1.04)	(-0.93)	(-1.11)	(-1.01)	(-0.88)
Adv	+	0.0279	0.0397	0.0604	0.0279	0.0637	0.0863*
		(0.53)	(0.76)	(1.16)	(0.53)	(1.22)	(1.65)
R&D	-	-0.0396***	-0.0348***	-0.0365***	-0.0396***	-0.0144	-0.0191*
		(-3.59)	(-3.15)	(-3.40)	(-3.59)	(-1.27)	(-1.72)

TABLE 21 (Continued)

DV: *Accuracy*		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Std_Ret	+	0.3413***	0.3481***	0.3642***	0.3413***	0.3573***	0.3768***
		(13.89)	(14.17)	(15.01)	(13.89)	(14.55)	(15.49)
#Analysts	+	0.0031***	0.0031***	0.0032***	0.0031***	0.0031***	0.0033***
		(9.08)	(9.10)	(9.67)	(9.08)	(9.30)	(9.99)
Intercept		0.0912	0.1348	0.1378***	0.0912	0.1202*	0.1230*
		(1.44)	(2.11)	(2.17)	(1.44)	(1.90)	(1.95)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
n		9,640	9,640	9,640	9,640	9,640	9,640
Adjusted R-Squared		0.1092	0.1120	0.1101	0.1092	0.1162	0.1127

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 22

Correlation Matrix for the Association between Uncertainty in Analysts' Overall Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Uncertainty_Overall*		0.0218*	0.0925*	0.0071	-0.0535*	0.1386*	-0.2834*	-0.1479*	-0.0442*	-0.1329*	0.1749*	-0.0809*	0.0091	0.3902*	-0.1014*
2	*Uncertain*	-0.0372*		0.6601*	-0.0721*	0.2174*	0.1684*	-0.1474*	0.0476*	-0.1810*	0.0285*	0.0500*	0.0407*	0.2315*	0.1503*	-0.0604*
3	*Weak_Modal*	-0.0139	0.6702*		0.1044*	0.1792*	0.3716*	-0.2812*	0.1073*	-0.3144*	-0.0476*	0.1373*	0.0543*	0.4019*	0.3179*	-0.1059*
4	*Fog*	0.0104	-0.1155*	0.0695*		0.0147	0.1505*	0.0146	-0.0004	0.0016	0.0577*	-0.0225*	0.0042	0.0716*	0.0014	0.0485*
5	*Pos_Tone*	-0.0610*	0.1994*	0.2072*	-0.0068		0.0736*	-0.0325*	-0.0035	-0.0638*	-0.0578*	-0.0166	0.0358*	0.3750*	0.0641*	-0.0441*
6	*Neg_Tone*	0.0787*	0.1774*	0.3772*	0.1344*	0.0408*		-0.1460*	-0.0840*	-0.0879*	-0.0216*	0.1104*	0.0990*	0.3368*	0.2645*	-0.0072
7	*Firm_Size*	-0.1650*	-0.1411*	-0.2667*	0.0334*	-0.0194	-0.1183*		0.0465*	0.2694*	0.3314*	-0.2313*	0.0267*	-0.1016*	-0.5028*	0.6825*
8	*Growth*	-0.0165	0.0510*	0.1422*	0	0.0511*	-0.0188	-0.0286*		-0.1085*	0.0146	0.0083	-0.0108	0.0676*	0.0057	-0.0202
9	*Segments*	-0.0095	-0.1829*	-0.3147*	0.0158	-0.0751*	-0.0880*	0.2760*	-0.1103*		0.0670*	-0.1207*	-0.0527*	-0.1611*	-0.2226*	0.0845*
10	*%_Inst*	-0.1131*	0.0148	-0.0754*	0.0600*	-0.0595*	-0.0279*	0.3409*	-0.0762*	0.0860*		-0.1310*	0.0158	-0.1338*	-0.2811*	0.2939*
11	*10K_News*	0.0889*	0.0503*	0.1289*	-0.0191	0.0102	0.0995*	-0.2216*	0.0663*	-0.1037*	-0.1658*		-0.013	0.0862*	0.3743*	-0.1327*
12	*Adv*	0.0101	0.0295*	0.0543*	-0.0258*	0.0145	0.0666*	0.0804*	0.0439*	-0.0240*	-0.0441*	0.0056		0.0283*	0.0015	0.0643*
13	*R&D*	-0.0077	0.1797*	0.4374*	0.0703*	0.4004*	0.2289*	-0.1275*	0.1823*	-0.2668*	-0.1488*	0.1100*	-0.0618*		0.2409*	-0.0218*
14	*Std_Ret*	0.2381*	0.1059*	0.2675*	-0.0197	0.0766*	0.2241*	-0.4350*	0.1281*	-0.2005*	-0.3471*	0.4136*	0.0139	0.2481*		-0.2810*
15	*#Analysts*	-0.0251*	-0.0645*	-0.0954*	0.0509*	-0.0500*	0.0199	0.6959*	-0.0559*	0.0807*	0.2384*	-0.1206*	0.0988*	-0.0246*	-0.2253*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 23

Multivariate Analysis for the Association between Uncertainty in Analysts' Overall Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

DV: Uncertainty_Overall		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Uncertain	+		-3.0118***	-3.0567***			
			(-5.61)	(-5.76)			
Weak_Modal	+					-8.9429***	-7.7987***
						(-8.69)	(-7.78)
Fog	+	0.0008	-0.0002	0.0001	0.0008	0.0006	0.0011
		(0.89)	(-0.25)	(0.17)	(0.89)	(0.65)	(1.24)
Pos_Tone	-	-2.8336***	-2.0775**		-2.8336***	-2.1854**	
		(-3.16)	(-2.29)		(-3.16)	(-2.44)	
Neg_Tone	+	1.5763***	1.7332***		1.5763***	2.3727***	
		(3.94)	(4.32)		(3.94)	(5.80)	
Firm_Size	-	-0.0190***	-0.0195***	-0.0202***	-0.0190***	-0.0206***	-0.0213***
		(-13.43)	(-13.79)	(-14.39)	(-13.43)	(-14.53)	(-15.07)
Growth	-	-0.0051	-0.0041	-0.0045	-0.0051	-0.0023	-0.0033
		(-1.58)	(-1.27)	(-1.41)	(-1.58)	(-0.71)	(-1.02)
Segments	+	0.0139***	0.0119***	0.0122***	0.0139***	0.0093***	0.0103***
		(4.10)	(3.49)	(3.58)	(4.10)	(2.71)	(3.01)
%_Inst	-	-0.0331***	-0.0320***	-0.0306***	-0.0331***	-0.0320***	-0.0305***
		(-4.71)	(-4.57)	(-4.36)	(-4.71)	(-4.58)	(-4.35)
10K_News	+	-0.0478	-0.0444	-0.0379***	-0.0478	-0.0438	-0.0363
		(-1.16)	(-1.08)	(-0.92)	(-1.16)	(-1.07)	(-0.88)
Adv	+	0.0749	0.08560*	0.1071**	0.0749	0.1072**	0.1304***
		(1.58)	(1.82)	(2.27)	(1.58)	(2.27)	(2.76)
R&D	-	-0.0318***	-0.0279***	-0.0307***	-0.0318***	-0.0095	-0.0149
		(-3.18)	(-2.79)	(-3.16)	(-3.18)	(-0.92)	(-1.49)

TABLE 23 (Continued)

DV: Uncertainty_Overall		***Uncertain***			***Weak_Modal***		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Std_Ret	+	0.3590***	0.3652***	0.3807***	0.3590***	0.3732***	0.3919***
		(15.53)	(15.8)	(16.69)	(15.53)	(16.17)	(17.14)
#Analysts	+	0.0035***	0.0035***	0.0036***	0.0035***	0.0035***	0.0037***
		(11.43)	(11.45)	(12.08)	(11.43)	(11.67)	(12.41)
Intercept		0.1307**	0.1717***	0.1729***	0.1307**	0.1600***	0.1601***
		(2.17)	(2.84)	(2.87)	(2.17)	(2.67)	(2.68)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
n		8,989	8,989	8,989	8,989	8,989	8,989
Adjusted R-Squared		0.1439	0.1545	0.1445	0.1439	0.151	0.1471

The *t*-statistics in parentheses. *, **, *** stands for $p < 0.1$, $p<0.05$, $p<0.01$, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

TABLE 24

Correlation Matrix for the Association between Uncertainty in Analysts' Common Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings: For Replacement

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Uncertainty_Common*		-0.0740*	-0.0976*	-0.0293*	-0.0530*	-0.0223*	0.0875*	-0.0117	0.0756*	0.0820*	-0.0212*	-0.0017	-0.1040*	0.0292*	0.1092*
2	*Uncertain*	-0.0722*		0.6599*	-0.0719*	0.2175*	0.1680*	-0.1471*	0.0474*	-0.1808*	0.0286*	0.0503*	0.0406*	0.2315*	0.1500*	-0.0599*
3	*Weak_Modal*	-0.0987*	0.6700*		0.1046*	0.1793*	0.3712*	-0.2808*	0.1071*	-0.3143*	-0.0475*	0.1377*	0.0541*	0.4020*	0.3176*	-0.1054*
4	*Fog*	-0.0305*	-0.1154*	0.0697*		0.0146	0.1508*	0.0144	-0.0004	0.0015	0.0578*	-0.0226*	0.0043	0.0716*	0.0016	0.0482*
5	*Pos_Tone*	-0.0564*	0.1995*	0.2073*	-0.0068		0.0738*	-0.0326*	-0.0036	-0.0638*	-0.0575*	-0.0165	0.0359*	0.3750*	0.0642*	-0.0442*
6	*Neg_Tone*	-0.0143	0.1769*	0.3768*	0.1346*	0.0409*		-0.1455*	-0.0842*	-0.0877*	-0.0216*	0.1107*	0.0988*	0.3369*	0.2642*	-0.0066
7	*Firm_Size*	0.0885*	-0.1407*	-0.2663*	0.0333*	-0.0196	-0.1178*		0.0467*	0.2692*	0.3315*	-0.2316*	0.0269*	-0.1016*	-0.5025*	0.6823*
8	*Growth*	-0.0018	0.0509*	0.1422*	0	0.0511*	-0.0189	-0.0285*		-0.1084*	0.0148	0.0085	-0.0109	0.0676*	0.0055	-0.02
9	*Segments*	0.0751*	-0.1827*	-0.3146*	0.0158	-0.0752*	-0.0877*	0.2758*	-0.1103*		0.0670*	-0.1209*	-0.0526*	-0.1611*	-0.2224*	0.0842*
10	*%_Inst*	0.0993*	0.0149	-0.0753*	0.0600*	-0.0593*	-0.0279*	0.3411*	-0.0762*	0.0860*		-0.1312*	0.0158	-0.1337*	-0.2812*	0.2940*
11	*10K_News*	-0.0241*	0.0505*	0.1292*	-0.0191	0.0103	0.0996*	-0.2217*	0.0663*	-0.1037*	-0.1659*		-0.0129	0.0863*	0.3746*	-0.1331*
12	*Adv*	0.0058	0.0296*	0.0544*	-0.0258*	0.0145	0.0666*	0.0805*	0.0439*	-0.0240*	-0.0442*	0.0056		0.0283*	0.0013	0.0646*
13	*R&D*	-0.1139*	0.1798*	0.4377*	0.0703*	0.4005*	0.2290*	-0.1275*	0.1823*	-0.2668*	-0.1489*	0.1100*	-0.0618*		0.2410*	-0.0218*
14	*Std_Ret*	0.0294*	0.1057*	0.2673*	-0.0196	0.0767*	0.2238*	-0.4348*	0.1280*	-0.2004*	-0.3472*	0.4137*	0.0139	0.2482*		-0.2806*
15	*#Analysts*	0.1094*	-0.0641*	-0.0949*	0.0508*	-0.0502*	0.0205	0.6958*	-0.0559*	0.0804*	0.2386*	-0.1207*	0.0989*	-0.0246*	-0.2251*	

This table reports the Pearson (below the diagonal) and Spearman (above the diagonal) correlation coefficients for the variables used in the regression analysis. * stands for $p<0.05$ two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. All variables are defined in Appendix 2.

TABLE 25

Multivariate Analysis for the Association between Uncertainty in Analysts' Common Information Environment and the Use of Uncertain and Weak Modal Words in 10-K Filings: Fog replacement

DV: Uncertainty_Common		Uncertainty			Weak Modal		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Uncertain	+		-8.5482***	-8.5315***			
			(-4.61)	(-4.66)			
Weak_Modal	+					-16.4710***	-14.9049***
						(-4.63)	(-4.30)
File_Size	+	-0.0056*	-0.0084***	-0.0079***	-0.0056*	-0.0060**	-0.0053*
		(-1.88)	(-2.77)	(-2.63)	(-1.88)	(-2.02)	(-1.79)
Pos_Tone	-	-3.8991	-1.7528		-3.8991	-2.7059	
		(-1.26)	(-0.56)		(-1.26)	(-0.87)	
Neg_Tone	+	1.7555	2.1980		1.7555	3.2190**	
		(1.27)	(1.59)		(1.27)	(2.27)	
Firm_Size	-	0.0000	-0.0014	-0.0023	0.0000	-0.0030	-0.0040
		(0.01)	(-0.29)	(-0.47)	(0.01)	(-0.61)	(-0.81)
Growth	-	0.0177	0.0206*	0.0199*	0.0177	0.0228**	0.0214*
		(1.59)	(1.85)	(1.79)	(1.59)	(2.05)	(1.93)
Segments	+	0.0360***	0.0302**	0.0306***	0.0360***	0.0275**	0.0288**
		(3.08)	(2.57)	(2.60)	(3.08)	(2.32)	(2.44)
%_Inst	+	0.1878***	0.1908***	0.1922***	0.1878***	0.1898***	0.1918***
		(7.75)	(7.88)	(7.95)	(7.75)	(7.84)	(7.93)
10K_News	+	-0.5008***	-0.4909***	-0.4835***	-0.5008***	-0.4931***	-0.4831***
		(-3.53)	(-3.46)	(-3.41)	(-3.53)	(-3.48)	(-3.41)
Adv	+	-0.0764	-0.0449	-0.0186	-0.0764	-0.0167	0.0146
		(-0.47)	(-0.27)	(-0.11)	(-0.47)	(-0.10)	(0.09)
R&D	-	-0.2269***	-0.2158***	-0.2172***	-0.2269***	-0.1857***	-0.1925***
		(-6.58)	(-6.25)	(-6.47)	(-6.58)	(-5.22)	(-5.54)

TABLE 25 (Continued)

DV: Uncertainty_Common		Uncertainty			Weak Modal		
IV	**Prediction**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Std_Ret	+	0.5806***	0.5980***	0.6181***	0.5806***	0.6066***	0.6321***
		(7.28)	(7.50)	(7.86)	(7.28)	(7.59)	(8.00)
#Analysts	+	0.0069***	0.0069***	0.0071***	0.0069***	0.0070***	0.0073***
		(6.64)	(6.65)	(6.88)	(6.64)	(6.75)	(7.06)
Intercept		0.6485***	0.7651***	0.7719***	0.6485***	0.7025***	0.7046***
		(3.13)	(3.67)	(3.72)	(3.13)	(3.38)	(3.42)
Year/Ind Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes
n		8,987	8,987	8,987	8,987	8,987	8,987
Adjusted R-Squared		0.0653	0.0674	0.0673	0.0653	0.0674	0.0670

The *t*-statistics in parentheses. *, **, *** stands for p < 0.1, p<0.05, p<0.01, respectively, two-tailed t-tests. All continuous variables are winsorized at 1% and 99%. Variables are defined in the Appendix 2.

VITA

MYUNG SUB KIM

Born, Seoul, Korea

1996-2002	Bachelor of Business Administration Dankook University Seoul, Korea
2002-2003	Junior Accountant Stryker Seoul, Korea
2004-2007	Senior Accountant Emerson Process Management Seoul, Korea
2008-2009	Assistant Manager in Finance and Accounting bioMérieux Seoul, Korea
2009-2012	Master of Professional Accountancy Georgia State University Atlanta, Georgia
2014-2018	Ph.D. Business Administration Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

Kim, M., (2018, January). "*Market Reaction to the OTCQX International Market Listing"* Paper presented at American Accounting Association International Accounting Section Midyear Meeting, Long Beach, California.