THREE ESSAYS ON TEXTUAL ANALYSIS AND AUDIT QUALITY

by

YANGIN YOON

A Dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

written under the direction of

Dr. Kevin Moffitt

and approved by

Newark, New Jersey

October, 2020

ABSTRACT OF THE DISSERTATION

Three Essays on Textual Analysis and Audit Quality

By YANGIN YOON

Dissertation Director:

Dr. Kevin Moffitt

This dissertation comprises three essays. The first two apply textual analysis methodologies in the accounting and auditing domain and the last essay investigates materiality as a key audit component.

The first essay develops a novel measurement, management self-efficacy, through analysis of earnings conference call transcripts. This measurement captures managers' level of judgment about the ability of their company to achieve its goals. This essay provides empirical evidence of the negative association between the level of management self-efficacy and the company's future financial performance. This result supports the overconfidence and obfuscation arguments. This study also introduces a natural language processing methodology that uses artificial neural networks for dictionary building.

The second essay documents the usefulness of the textual analysis of earnings conference calls in identifying accounting misreporting. This essay demonstrates that differences in tone between CEOs and CFOs can predict misstatements in financial statements. When accounting misreporting exists, the negative tone of CFOs during Q&A session increases significantly more than that of the CEOs. This essay documents the empirical evidence regarding narrative tone sensitivities based on a person's job title and related personal characteristics.

The third essay concerns the materiality level, an important aspect of auditing. An external auditor sets a materiality level at an early stage of the auditing process and the subsequent audit procedures and results are directly affected by this level. However, due to the scarcity of archival data, few studies have considered the determinants of the materiality level. This essay provides empirical evidence that both new clients and long-tenured clients have a higher materiality level than other clients. Furthermore, the materiality level is negatively associated with the modified audit opinion when it is a new client.

Acknowledgments

First and the most, I express my appreciation to my dissertation committee members, Dr. Kevin Moffitt, Dr. Miklos Vasarhelyi, Dr. Helen Brown-Liburd, and Dr. Isabel Martínez Conesa. Thanks to your insightful guidance and advice, I could finish my dissertation. In particular, Kevin, I have always enjoyed discussions with you, and they were filled with novel curiosities and cheerful encouragements.

I also want to thank my Rutgers faculty members and colleagues, Dr. Alexander Kogan, Dr. Bharat Sarath, Dr. Divya Anantharaman, Dr. Won Gyun No, Ph.D. students, and Barbara Jensen. It has been my pleasure to work with you all. I appreciate the financial support from the Korea Financial Supervisory Service and Rutgers University.

To my parents, my sister, and parents-in-law, I sincerely thank you for your unconditional love and support. And to my beloved wife, I could achieve this great result with your dedicated supports during the whole program. Now, hopefully, it is my turn to return.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The advancement of technology has allowed individuals to utilize various sources of accounting information for their decision-making. Such information sources have expanded and continue to grow, as has been proposed and anticipated in the academic literature (Vasarhelyi and Halper 1991; Bovee et al. 2005), and recently, a wide variety of information is actively proposed and implemented in the Big Data era (Moffitt and Vasarhelyi 2013; Brown-Liburd and Vasarhelyi 2015; Vasarhelyi et al. 2015). For example, the use of drones (Appelbaum and Nehmer 2017), legal contract analytics (Zhaokai and Moffitt2019), and voice recognition and natural language processing (Li and Liu 2020) have been proposed to assist auditors.

One of the major accounting information sources is textual data; in addition to numerical financial information, languages can provide an abundance of information. Accounting information users and academic researchers are increasingly utilizing this textual information (Moffitt and Vasarhelyi 2013; Brown-Liburd and Vasarhelyi 2015; Vasarhelyi et al. 2015; Loughran and McDonald 2016). In the first two chapters of this dissertation, we analyze textual information, especially in the form of earnings conference call transcripts, and propose a new measurement: management self-efficacy. In Chapter 2 of this dissertation, we document the association between the level of management self-efficacy and company's future financial performance. In Chapter 3, we analyze the tone changes during conference calls over the years. Based on this, we propose a model for identifying accounting misreporting.

In Chapter 4, we investigate a traditional auditing concept: the level of materiality. The materiality level is negatively associated with the reporting reliability (PCAOB AS 2105, ISA 320, Choudhary et al. 2019). We empirically examine the level of materiality and find evidence that the materiality levels in the early audit years are higher than in others. This finding may have implications on the existing literature or reveal the mechanism that indicates the low accounting quality of the initial audit tenure (Johnson et al. 2002; Carcello and Nagy 2004).

This dissertation's findings can be utilized by various stakeholders. Investors can use our findings for their investment decisions when predicting a company's future performance or identifying accounting misreporting. Auditors can use these findings when evaluating the potential risk level of their new and existing clients. In particular, stakeholders can use these findings in the early years of new auditing processes based on the results of the third essay with regard to materiality.

This dissertation also proposes some technical aspects of analyzing textual information. Starting with the bag-of-words approach, we add a method that utilizes sentence structure to accurately capture contextual meaning. In addition to this, we introduce up-to-date artificial intelligence (AI) Natural Language Process (NLP) techniques—Word2Vec and Doc2Vec models—to the accounting and auditing domains. (Bengio et al. 2003; Mikolov et al. 2013b, 2013c; Le and Mikolov 2014). This research shows these up-to-date NLP models have significant potential.

Chapter 2: Management self-efficacy and future firm performance

2.1. Introduction

The availability of resources is a major determinant of outcomes for organizations and people (Barney 1991; Penrose 1959; Wernerfelt 1984). However, performance still varies between two people or two organizations when both have equal access to resources. In this situation, the internal motivation of the entity and its belief in itself are key differentiators. Bandura (1986) described individuals' belief in their ability to succeed at a specific task as "self-efficacy." Stronger belief is typically followed by improved performance. In this paper, we focus on managers' belief in their company's ability to succeed, which we term management self-efficacy.

At the individual level, the relationship between perceived self-efficacy and performance has been widely researched across many types of tasks (Gist et al. 1989; Gist and Mitchell 1992; Harrison et al. 1997; Hill et al. 1987; Taylor et al. 1984). Self-efficacy has also been investigated at the group and organizational levels (Gully et al. 2002; Lester et al. 2002; Lindsley et al. 1995). At these levels, self-efficacy is referred to as group efficacy, collective efficacy, or group potency. Nearly all self-efficacy research has measured this construct via survey responses from research participants (Gong et al. 2009; Lent et al. 1987; Taylor et al. 1984) and, to the best of our knowledge, this study is the first to measure self-efficacy through the application of textual analysis to transcribed interviews.

To measure management self-efficacy in this study, we propose and develop a measure that captures self-efficacy at the organizational level by analyzing manager transcripts from earnings conference calls. We then investigate the relationship between management self-efficacy and future firm performance. We perform additional analyses to ensure that the new measure does not capture the same information as through existing widely used methods. Furthermore, we apply an up-to-date textual analysis technique, Word2Vec, to enhance the initially proposed measurement.

First, we provide empirical evidence of the determinants of management self-efficacy. Companies with low ROA exhibit higher levels of management self-efficacy during earnings conference calls. Besides, larger total assets and lower debt ratios are related to higher levels of management self-efficacy during earnings conference calls. Our second finding is that companies with higher self-efficacy experience decreased return on assets (ROA) in the year subsequent to the earnings call and this trend continues for the following year. This suggests that managers' expressed belief in their companies' abilities may often be unwarranted and indicate overconfidence. In the additional analysis section, this study shows that Word2Vec method can be effectively applied to generate a word list of synonyms by using domain specific textual documents. Furthermore, if we adjust the level of management self-efficacy by using the difference between the Q&A session and the presentation session, a positive association is found between management self-efficacy and a company's future financial performance.

Our study contributes to the literature on self-efficacy and text analytics in accounting in several ways. First, our research proposes a new method for measuring self-efficacy using textual analysis of transcripts of managers' speech. This is a measure of self-efficacy at the organizational level, which we term management self-efficacy. Unlike previous studies that administered surveys to a small sample of start-up CEOs (Forbes 2005; Hmieleski and Corbett 2008), our method allows for a much more comprehensive sample of corporate CEOs, and the data are generated in a real business context. Second, we document empirical evidence that self-efficacy at the organizational level—or management self-efficacy—has a negative association with a firm's future financial performance, which is a valuable indicator for capital market participants. Third, we develop a new and contextually rich measure of self-efficacy that combines sentence structure, word lists, and verb tenses. The measurements in prior studies that relied solely on curated lists of words (i.e., bag-of-words approach) to identify and measure phenomena lacked substantial context. Fourth, our research contributes to the literature regarding the informativeness of earnings conference calls.

The remainder of this chapter is organized as follows: in Section 2, we discuss relevant prior literature and develop our hypotheses. Section 3 documents the construction of the management self-efficacy measure. Section 4 presents the research sample and empirical research design. We discuss the results of our hypothesis testing in Section 5, and we perform additional analyses by comparing the existing measurements and introducing an AI textual analysis technology, Word2Vec, in Section 6. We conclude this chapter with a discussion of its contributions in Section 7.

2.2. Literature review and hypotheses

2.2.1. Self-efficacy

Bandura (1986) defined self-efficacy as “people’s judgements of their capability to organize and execute courses of action required to attain designated types of performances.” And Wood and Bandura (1989) use the definition, “self-efficacy refers to beliefs in one’s capabilities to mobilize the motivations, cognitive resources, and courses of action needed to meet situational demands.”

Bandura (2000) argues that individuals who have high perceived self-efficacy have stronger internal motivation to achieve their goals and they better overcome difficulties and obstacles. Self-efficacy increases the expected probability of a positive outcome.

The relationship between self-efficacy and performance has been studied in a variety of settings including the productivity of academic faculty (Taylor et al. 1984), individual job interview tasks (Stumpf et al. 1987), innovative technology adoptability (Hill et al. 1987), and training effectiveness (Gist et al. 1989). Within the business environment, positive associations between self-efficacy and performance have been documented for employee’s sales activities (Gong et al. 2009), and managers’ effectiveness (Luthans 2002). On the other hand, negative outcomes have also been observed where increased self-efficacy has led to overconfidence and poorer performance in standardized tests (Moores and Chang 2009) and memory games (Vancouver et al. 2002).

Self-efficacy originates from the individual; however, it can be applied to a group of people or a team. Similar to individuals, if the collective membership of a team has high self-efficacy, it can strongly motivate team members and lead to positive outcomes (Feltz and Lirgg 1998; Gully et al. 2002; Peterson et al. 2000; Prussia and Kinicki 1996).

Relatively recent research in the self-efficacy domain has explored the relationship between the self-efficacy of entrepreneurs and firm performance. Forbes (2005) finds that entrepreneurial self-efficacy has a positive relation with new venture companies' performance as measured by sales growth and overall subjective performance evaluation. Hmieleski and Corbett (2008) study the relationships between entrepreneurial self-efficacy and improvisational behavior and new venture performance.

We make several contributions to the self-efficacy literature. We develop a method for measuring self-efficacy outside the survey methodology. Second, there may be a conceptual difference between the entrepreneur's/CEO's self-efficacy and the company's self-efficacy, even though the CEO is the most influential person in the company. Entrepreneurial self-efficacy measures the entrepreneur's individual belief in their ability to succeed. Management self-efficacy, while still measured from the CEO and CFO, reflects their belief in the company's ability to reach its goals. Lastly, research on CEOs and their self-efficacy is couched in the limited scope of startup companies while there is a paucity of research on self-efficacy for management from more seasoned companies.

2.2.2. Textual analysis

Textual analysis a way of extracting information from unstructured text (Loughran and McDonald 2016). It is a different approach from traditional analyses of companies using financial and accounting data in that it focuses on the textual disclosures rather than numerical information in corporate disclosures. Although the textual analysis existed in the 1300s (Loughran and McDonald 2016), automated approaches are recently enabled by improvements in information technology.

Many researchers have tried to capture various constructs of textual data using different techniques. One seminal work in the area comes from Loughran and McDonald (2011), wherein the authors measured the general tone of companies' 10-Ks. They created four categories of word lists that are specific to accounting and finance including positivity, negativity, litigiousness, and uncertainty. They document that the tone of the materials is related to stock market return, trading volume, volatility of the returns, and the unexpected earnings.

Researchers try to predict companies' future earnings by deploying textual analysis. Li (2008) analyzes an association between the earnings persistence and annual report readability. Li (2010b) uses a naïve Bayesian model, identifying the forward-looking sentences in the MD&A section of 10-K and 10-Q filings. Huang et al. (2014) investigate the abnormal positive tone in the earnings press release and companies' negative future earnings.

Earnings conference call transcripts are one of the major sources to researchers. Larcker and Zakolyukina (2012) and Burgoon et al. (2016) analyze conference call transcripts in order to identify the linguistic cues about accounting misreporting. Hollander et al. (2010) investigate the management behavior—refuse to answer certain questions from analysts during the conference call—and investors' reaction. Lee (2016) examines whether the management relies on the prepared scripts when the management answers during the Q&A session in the conference call.

In addition, there exists a research stream to measure personal characteristics by utilizing the textual analysis. Hermann (1999) identifies leadership style by analyzing a person's speech. Mairesse et al. (2007) develop a model to measure the Big Five personality traits based on a linguistic model. Green et al. (2019) try to measure the level of extraversion of the management from the conference calls. They document the relationship between the level of extraversion of the CEO and their salaries or company's sales growth.

In this study we merge a word list/bag-of-words approach with sentence structure and part-of-speech analysis to identify self-efficacy statements made by management in earnings conference call transcripts. Self-efficacy is a subjective and psychological concept by nature. Furthermore, in our study, we try to measure the self-efficacy at the organizational level. Thus, it is important to properly operationalize and develop a metric for measuring the organizational perceived self-efficacy.

We use earnings conference call transcripts because the statements from management can be construed as the company's official stance. The conference call participants are the top management and they have the most comprehensive understanding of a company's capabilities, circumstances, and the expected future outcomes. Moreover, they are the most influential people in the company and they develop and execute the business strategy. For these reasons, the earnings conference call is ideal for measure management self-efficacy.

2.2.3. Hypotheses

One of the features and limitations of entrepreneurial self-efficacy measures is that they measure the CEO's self-efficacy at the individual level. Consider the differences in the two sentences: (1) "I, as a CEO, can encourage our research team to produce innovative products." and (2) "Our company research team has full capacity to launch innovative products." Prior studies' main interests would be in the individual CEO's self-efficacy (sentence 1). Management self-efficacy corresponds to management belief in organizational abilities (sentence 2).

Thus, in this study we introduce a novel way to measure the self-efficacy of companies as stated by management. To do this, we apply textual analysis on earnings conference call transcripts in which management represents their respective organizations to outside stakeholders.

We evaluate the language of managers in the earnings conference call. Managers frequently comment on the ability of their companies to accomplish certain goals, making conference call transcripts a suitable data set for extracting management self-efficacy measures. From the company side, a typical earnings conference call includes the CEO, the CFO, and other high level managers. We treat all comments from employees as one, not taking into account job title or position in this Chapter.

In our first analysis, we investigate the determinant factors of management self-efficacy. There could be many related factors which contribute; however, during the earnings conference call, the main revolves around the current and future financial performance of the company. Thus, the magnitude of earnings likely has a significant impact on the content of calls self-efficacy language. If managers are buoyed up and optimistic due to higher earnings, then positive performance could lead to higher management self-efficacy. However, if managers want to project optimism and confidence in the face of disappointing financial performance, then it is possible that poor results could prompt higher management self-efficacy language. In order to account for either outcome, we hypothesize in the null:

H1: Firm financial performance has no effect on the level of management self-efficacy language in earnings conference calls.

Next we investigate the relationship between the management self-efficacy and future firm performance. Previous research documents a positive association between self-reported entrepreneurial self-efficacy and their startup companies' performance (Forbes 2005; Hmieleski and Corbett 2008). In our setting, management self-efficacy is unobtrusively measured, and it is an organizational-level measure. Management self-efficacy language is typically oriented toward what a company can do in the future. "We can meet our goals." "Our team is able to surpass the competition." If the language of earnings calls is high in management self-efficacy, then management is projecting confidence in the future which may result in or foreshadow strong firm performance. On the other hand, if strong results are not expected, management may try even harder to project confidence and ability to the analysts, investors, and media who will scrutinize the call. Some prior literature show that the management has overconfidence in their abilities (Malmendier and Tate 2005, 2008; Hirshleifer, Low, and Teoh 2012). And this overconfidence causes suboptimal decisions such as overinvestment. As a result, overconfidence could portend poorer future results. To account for both possibilities, we hypothesize in the null:

H2: There is no relationship between future firm financial performance and the current level of management self-efficacy

2.3. Measuring management self-efficacy

2.3.1. Management self-efficacy characteristics

Management self-efficacy is managements' confidence in their company's ability to succeed at its goals. The most fundamental goal for public corporations is maximizing shareholder value, but there are other types of goals such as developing an effective marketing strategy, incorporating cutting-edge technology to obtain a competitive advantage, and retaining valuable employees. In our measure we do not consider the specific type of goal.

Management self-efficacy is forward-looking by nature. Thus, we seek to identify sentences that are future-oriented and mention an ability to accomplish something beneficial to the firm. We attempt to identify self-efficacy statements at the sentence level meaning. Each sentence uttered by management is analyzed to determine if it is a self-efficacy sentence.

2.3.2. Constructing the management self-efficacy measurement rules

We develop a rules-based measurement that combines sentence structure, words lists, and parts of speech in order to measure the management self-efficacy. In all there are six rules.

2.3.2.1. Bag-of-words

The bag-of-words method, using the pre-defined list of words, is widely deployed to measure the tone and other constructs in text (Loughran and McDonald 2016). We use a predefined "achievement" word list from LIWC (Linguistic Inquiry and Word Count) to help identify self-efficacy sentences. LIWC word lists were developed for general textual analysis purposes and have been widely used in accounting and finance research (Keila and Skillicorn 2005; Larcker and Zakolyukina 2012; Matsumoto et al. 2011).

The "achievement" category in LIWC word list contains 185 words including able, achieve, resourceful, opportunity, strength, and success—words related to the concept of self-efficacy. We modify the achievement word list and remove 23 words with negative connotations. Because many words are homographic (they have more than one meaning), the interpretability of the bag-of-words approach is inherently limited. To bolster our bag-of-words approach and the LIWC achievement dictionary we added 162 additional words. These words are synonyms of the efficacy words—ability, capability, able, and capable. One of our rules is that if a sentence contains an achievement word, and does not contain a past tense verb, then it is a self-efficacy sentence.

The words that our synonyms were based on (ability, capability, able, and capable) in addition to the plural forms capabilities and abilities are considered core self-efficacy terms. If a sentence contains any of those words, it is considered a self-efficacy sentence. We also identified 22 additional words that occur frequently in self-efficacy sentences (see Appendix A). We term this list of words and the *KevinNBen* word list, and if a sentence has one, it is classified as a self-efficacy sentence.

2.3.2.2. Sentence structure

We aim to identify sentences in which the subject is referring to the firm itself, and since management represents the firm, we search for first person pronouns in the subject of the sentence. One powerful indicator of a self-efficacy sentence is the presence of words such as "*I can*" or "*we are able to*". We call these "I can" sentences, but we exclude certain types of "I can" sentences such as "I can say", "I can tell you", and "I can answer". We also consider sentence negation. If the sentence contains the words 'not' or 'never' next to the verb, it does not count as a self-efficacy sentence.

We also take into account the forward-looking nature of a sentence and we look for words and phrases (together with first person pronouns) such as 'will', 'be going to', 'am going to', 'are going to', and 'is going to' and call them "*I will*" sentences. If, following one of these words, a sentence has one of our words from the modified LIWC "achievement" word list or "ability" synonym word list, it is considered a self-efficacy sentence. We also consider the present perfect tense (e.g. has been) which connects past events to the present (British Council 2020) and identify "*I have done*" sentences. In our reading of conference call transcripts, sentences with this tense insinuate preparation for continued success in the future, and we classify them as self-efficacy sentences when they appear with an achievement word.

Sentences can be very long, and "I" does not always occur next to "will". In order to connect the correct personal pronoun to the verb phrase that contains "will", "has been" etc., we use the Stanford CoreNLP sentence parser (Chen and Manning 2014). This tool decomposes a sentence into its constituent parts, and tags individual words with the appropriate part-of-speech. Using this tool, we first identify a verb phrase that contains relevant words (will, has done, etc.), and then search within the same independent clause (i.e., corresponding subject phrases) for a first-person pronoun from. Table 1 contains a summary of six rules.

Table 1 Summary of the rule for measuring the management self-efficacy

Rule No	Sentence Structure	Tense	Word list
1	'I can' Sentence	N/A	N/A
2	'I Will' Sentence	N/A	Modified LIWC or Modified synonym word lists
3	'I have done' sentence	N/A	Modified LIWC or Modified synonym word lists
4	N/A	N/A	One of 'ability, abilities, capability, capabilities, capable, and able' words
5	N/A	N/A	*KevinNBenWord* list
6	N/A	No past tense verb(s)	Modified LIWC or Modified synonym word lists

1) If a sentence meets any of the above six rules, the sentence is classified as a self-efficacy sentence.
2) If a sentence contains 'can say', 'can tell', or 'can answer', the sentence is excluded from the 'I can' sentence.
3) If a sentence contains a 'call' word and the total number of words is less than 15, the sentence is excluded from the 'I will' sentence.
4) All three conditions are applied simultaneously. For example, Rule 3 requires both having the 'I have done' sentence structure and containing more than one of necessary words in the sentence.
5) The personal pronouns of 'I, my, we, and our' are used for 'I can' sentences. 'I, my, we, our, and you' are used for 'I will' sentences or 'I have done' sentences.

2.3.3. Testing the initial rules and modifying the rules

In the previous section, we discussed the rules which are used in order to identify the firm's self-efficacy statement in the earnings conference call transcripts. We developed those rules based on our self-efficacy definition and extensive testing. In this section we describe the testing protocol used to refine and update the rules.

We first selected a set of 200 random sentences from the conference calls dataset. We manually coded each sentence as management self-efficacy or not. Our initial rules correctly identify the self-efficacy sentences with a rate of 75.5% among 200 sentences. Using these results, we heavily modified the LIWC achievement word list and our synonym word list. We also created and added the *KevinNBen* word list and garnered 84.0% overall accuracy with the same 200 sentences.

Using our updated word lists we randomly selected additional 200 sentences and manually coded them. We achieve an accuracy of 74.0%. Using these results, we added 8 words and removed 5 more from the existing word lists to achieve an overall accuracy of 78.8%. We applied this second modification over the first sample of 200 sentences and still achieved 84.0% accuracy.

We randomly selected 200 more sentences as a final test set and applied the most recent rules to achieve a classification accuracy of 76.5%. In our test sets, we found that the percentage of the self-efficacy sentences in the earnings conference call ranged between 10 and 20 percent. [1] Table 2 is the confusion matrix which contains the classification details of our final test. [2]

Table 2 Classification and validation results of the management self-efficacy

	Predicted as self-efficacy sentences	**Predicted as NOT self-efficacy sentences**	**Total**
Actual: Self-efficacy Sentences	27	8	35
Actual: Not Self-efficacy Sentences	39	126	165
Total	66	134	200

Overall accuracy: 76.5% (= [27 + 126] / 200)

2.3.4. Metric for the level of management self-efficacy

We use the rate of self-efficacy sentences in earnings calls to measure the degree of management self-efficacy.

$$\text{Management self-efficacy} = \frac{\text{the number of self–efficacy sentences}}{\text{the total number of sentences}}$$

[1] Although we have no prior empirical statistics about self-efficacy sentence distributions, it seems that this ratio is above the general percentage of the self-efficacy remarks made in daily, routine conversation. This may be because our sample comes from earnings conference call transcripts which explain company' outcome and its status from the management.

[2] Our measure's validity rate is similar or higher than the prior literature. Mairesse et al. (2007) documents 73% with their extraversion measure. Green et al. (2019) report 68% accuracy rate with their extraversion measure in the binary setting. Li (2010b) develops linguistic classification models. His tone measurement reports 67% accuracy under the three categories of tones. And for the three types of contents topic classification, he reports 82% accuracy.

Because there are two parts in the earnings call (the prepared presentation portion, followed by the question and answer session with analysts), we can devise three measures of management self-efficacy. SELF_EFFICACY_PT is the level of self-efficacy in the presentation portion of the call, SELF_EFFICACY_QNA is the level during the question and answer session, SELF_EFFICACY_ BOTH incorporates both parts.

2.4. Samples description and research design

2.4.1. Sample

Our sample consists of earnings conference call transcripts obtained from SeekingAlpha.com. In this study, we focus on the annual earnings (or 4th quarter) conference calls, which correspond with audited financial statements.[3] We exclude observations where conference call transcripts were uploaded an abnormally long time (over 30 days) after the call itself. We also eliminate calls from subsidiaries that were given the same ticker as the parent company. [4] We dropped the observations with missing relevant data in COMPUSTAT and CRSP databases and observations without the requisite amount of history to compute ROA volatility and future performance. If conference calls are held more than 3 months after the fiscal year end, those abnormal long audit delays may be related to unobservable specific circumstances, we removed

[3] Some conference calls occur with pre-audited earning numbers. However, when we measure the company's financial performance, we use the audited accounting numbers (from Compustat), which give us a higher assurance level than the unaudited quarterly numbers.
[4] If a subsidiary is also a listed company, it has its own ticker symbol. However, the transcript provider attached the parent company's ticker symbol in the transcripts in many cases.

those long delayed conference calls. Finally, we do not include conference call transcripts that only contain the presentation or Q&A portion of the call. A summary of the sample selection steps is given in Table 3, and Table 4 shows the yearly distribution of earnings conference calls in our sample.

Table 3 Sample Description

Source / Filter	Observations
All conference call transcripts from SeekingAlpha.com	149,164
Less transcripts that are not year end	(116,385)
Less transcripts uploaded more than 30 days after call	(335)
Less transcripts with a duplicate Ticker Symbol	(3,230)
Less transcripts not matched with Compustat (Matching the fiscal year end of the current year)	(2,682)
Less transcripts held more than 93 days after the fiscal year	(279)
Less transcripts not matched with Compustat (Two-year post for future predictions)	(7,973)
Less transcripts that do not have both prepared and Q&A sections	(784)
Final sample	17,496

Table 4 Distribution of year in the earnings conference call transcripts

Year	Number of earnings conference call transcripts
2005	130
2006	242
2007	1,171
2008	1,921
2009	1,151
2010	690
2011	1,042
2012	2,165
2013	2,662
2014	3,010
2015	2,985
2016	327
Total	17,496

The periods are classified by using the year of fiscal year-end of companies that are related to each earnings conference call.

2.4.2. Research Design

We first investigate the determinants of management self-efficacy level and we focus on characteristics of the firm that have been used in previous studies on the relationship between company-generated text and future performance (Li 2008, 2010b). Our main variables of main interest in this test are return on assets (ROA) and the change in ROA (D_ROA) from the previous year. Other variables are mainly financial variables or firm specific stock market variables. Appendix D provides the detailed definitions of the variables. SELF_EFFICACY in the model below is a placeholder for all three variants of our management self-efficacy measure (the prepared part, the Q&A part, and both parts).

$$\begin{aligned}
SELF_EFFICACY_{it} = \beta_0 &+ \beta_1 ROA_{it} + \beta_2 D_ROA_{it} + \beta_3 BIG4_{it} + \beta_4 LOG_TA_{it} \\
&+ \beta_5 DEBT_RATIO_{it} + \beta_6 CUR_RATIO_{it} \\
&+ \beta_7 RET_VOL_{it} + \beta_8 ROA_VOL_{it} \\
&+ \beta_9 ABS_TA_{it} + \beta_{10} BTM_{it} + \beta_{11} D_DIVIDEND_{it} \\
&+ \beta_{12} SEG_CNT_{it} + \beta_{13} FIRM_AGE_{it} \\
&+ \text{year fixed effects} + \text{industry fixed effects} + \varepsilon_{it}
\end{aligned} \tag{1}$$

We also test if management self-efficacy level is informative about future firm performance (H2). If the company has relatively high management self-efficacy, the company may plan more effectively and take the appropriate actions to achieve its goals, even in the face of obstacles and difficulties. On the other hand, high self-efficacy could indicate overconfidence or dissimulation and it may appear in an effort to assuage the concerns of analysts and investors. Financial performance can be measured in many ways. We use a standard measure of return on assets, which is defined as net income divided by total assets. We include control variables that are used in the prior literature of the earnings persistence research (Li 2008 and 2010b).

$$\begin{aligned} D\_ROA_{it+1} = {} & \beta_0 + \beta_1 \, SELF\_EFFICACY_{it} + \beta_2 \, ROA_{it} + \beta_3 \, D\_ROA_{it} \\ & + \beta_4 \, BIG4_{it} + \beta_5 \, LOG\_TA_{it} + \beta_6 \, DEBT\_RATIO_{it} \\ & + \beta_7 \, CUR\_RATIO_{it} + \beta_8 \, RET\_VOL_{it} + \beta_9 \, ROA\_VOL_{it} \\ & + \beta_{10} \, ABS\_TA_{it} + \beta_{11} \, BTM_{it} + \beta_{12} \, D\_DIVIDEND_{it} \qquad (2) \\ & + \beta_{13} \, SEG\_CNT_{it} + \beta_{14} \, FIRM\_AGE_{it} \\ & + \text{year fixed effects} + \text{industry fixed effects} + \varepsilon_{it} \end{aligned}$$

In addition to the next year’s financial performance, we test the persistence of the relationship between management self-efficacy future firm performances. The next model tests the relationship between management self-efficacy two consecutive years of improving ROA. The variable CONTI_D_ROA is a dummy variable that equals one if ROA has an increased trend for two consecutive years after the earnings call, otherwise the value is set to zero.

$$
\begin{aligned}
CONTI\_D\_ROA_{it+2} = {} & \beta_0 + \beta_1 SELF\_EFFICACY_{it} + \beta_2 ROA_{it} \\
& + \beta_3 D\_ROA_{it} + \beta_4 BIG4_{it} + \beta_5 LOG\_TA_{it} \\
& + \beta_6 DEBT\_RATIO_{it} + \beta_7 CUR\_RATIO_{it} \\
& + \beta_8 RET\_VOL_{it} + \beta_9 ROA\_VOL_{it} \\
& + \beta_{10} ABS\_TA_{it} + \beta_{11} BTM_{it} \qquad (3) \\
& + \beta_{12} D\_DIVIDEND_{it} \\
& + \beta_{13} SEG\_CNT_{it} + \beta_{14} FIRM\_AGE_{it} \\
& + \text{year fixed effects} + \text{industry fixed effects} + \varepsilon_{it}
\end{aligned}
$$

2.5. Empirical results

2.5.1. Descriptive statistics

Table 5 gives the descriptive statistics from our sample. The average rate of management self-efficacy sentences is 0.293. The measured management self-efficacy in the presentation part is 5.1 percent point (=31.7% - 26.6%) higher than that of the Q&A part. It could be that the presentation portion is more future oriented, and the Q&A portion, which depends on the questions asked, is more past-oriented. It could also be that the management shows less confidence in unprepared remarks.

There are two main dependent variables in our sample. One is the change of ROA in the next year, and the other is the increasing trend of ROA for the consecutive two years. In our sample, the mean of the change in ROA is close to zero, and the mean of increasing ROA trend over two consecutive years (CONTI_D_ROA) is 22%. Since there are four possibilities with CONTI_D_ROA (year 1 increase or decrease, year 2 increase or decrease), and three of those possibilities result in a zero value, and only increase both year 1 and year 2 gives a value of one, the final ratio is in line with what would happen in a random distribution. [5]

Table 5 Descriptive Statistics

Variables	N	Mean	St.Dev	p1	p25	Median	p75	p99
SELF EFFICACY PT	17,496	.317	.088	.115	.256	.316	.377	.527
SELF EFFICACY QNA	17,496	.266	.079	.094	.211	.262	.317	.481
SELF EFFICACY BOTH	17,496	.293	.068	.14	.245	.291	.339	.463
SELF EFFICACY DIFF	17,496	-.051	.097	-.28	-.116	-.051	.012	.195
D_ROA $_{t+1}$	17,421	-.009	.138	-.655	-.027	-.001	.018	.538
CONTI_D_ROA $_{t+2}$	17,429[6]	.221	.415	0	0	0	0	1
ROA	17,425	-.005	.178	-.925	-.004	.03	.071	.272
D_ROA $_t$	17,388	-.008	.131	-.586	-.027	-.001	.018	.557
BIG4	17,496	.842	.365	0	1	1	1	1
LOG TA	17,425	7.416	2.056	2.534	5.988	7.464	8.791	12.628
DEB_ RATIO	17,386	.552	.256	.07	.37	.55	.719	1.35
CURRENT_RATIO	14,481	.617	.434	.062	.323	.517	.788	2.508
RET_VOL	16,624	.111	.065	.029	.065	.095	.141	.373
ROA_VOL	17,388	.084	.151	.001	.015	.034	.084	1.057
ABS_T_ACCRUALS	17,417	.082	.097	.001	.026	.053	.099	.592
BTM	15,292	0.598	0.567	-0.457	0.255	0.466	0.782	3.443
D_DIVIDEND	17,487	.028	.164	0	0	0	0	1
SEG_CNT	17,496	1.744	1.397	1	1	1	2	7
FIRM AGE	17,496	22.613	16.119	3	11	18	29	65

All continuous variables are winsorized at the top and bottom 1 percent.
All variables are defined in Appendix D.

[5] The mean of D_ROA is near zero (-0.009 or -0.008). The chance of increase in the ROA or decrease in the ROA is roughly 50%, in our sample.

[6] There are companies that Compustat provides only the ROA of the next year without the ROA of the current year. If the ROA is decreased during the year of the next, the value of CONTI_D_ROA is assigned 0 even if there is no ROA information of the current year. Thus, the observation number of CONTI_D_ROA is larger than the observation number of D_ROA $_{t+1}$ in our sample.

Table 6 Correlations Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) SELF_EFFICACY_PT	1																		
(2) SELF_EFFICACY_QNA	0.335***	1																	
(3)SELF_EFFICACY_BOTH	0.836***	0.742***	1																
(4) SELF_EFFICACY_DIFF	-0.631***	0.514***	-0.150***	1															
(5) D_ROA t	-0.015**	-0.009	-0.015**	0.008	1														
(6) CONTI_D_ROA t+2	-0.01	0	-0.011	0.01	0.264***	1													
(7) ROA	-0.066***	-0.034***	-0.079***	0.032***	-0.263***	-0.149***	1												
(8) D_ROAt-1	0.019**	0.036***	0.028***	0.013*	-0.309***	-0.135***	0.359***	1											
(9) BIG4	0.011	0.011	-0.001	-0.002	0.031***	0.037***	0.172***	0.007	1										
(10) LOG_TA	-0.038***	-0.089***	-0.099***	-0.041***	0.031***	0.026***	0.375***	0.016**	0.400***	1									
(11) DEBT_RATIO	-0.023***	-0.095***	-0.070***	-0.058***	0.131***	0.076***	-0.074***	-0.029***	0.108***	0.385***	1								
(12) CURRENT_RATIO	-0.007	0.050***	0.029***	0.047***	-0.114***	-0.066***	-0.145***	0.055***	-0.116***	-0.360***	-0.522***	1							
(13) RET_VOL	-0.036***	-0.016**	-0.017**	0.019**	0.039***	0.047***	-0.401***	-0.041***	-0.139***	-0.391***	-0.054***	0.179***	1						
(14) ROA_VOL	0.064***	0.070***	0.098***	-0.001	0.019**	0.01	-0.559***	-0.025***	-0.207***	-0.447***	-0.093***	0.257***	0.387***	1					
(15) ABS_T_ACCRUALS	0.001	0.013*	0.021***	0.009	0.233***	0.137***	-0.521***	-0.285***	-0.104***	-0.312***	0.008	-0.008	0.335***	0.394***	1				
(16) BTM	-0.206***	-0.158***	-0.214***	0.055***	-0.055***	0.012	-0.015*	-0.086***	-0.040***	0.126***	-0.112***	-0.012	0.122***	-0.113***	0.004	1			
(17) D_DIVIDEND	-0.007	-0.067***	-0.032***	-0.049***	0.008	0.002	0.030***	0.005	0.057***	0.140***	0.074***	-0.122***	-0.132***	-0.078***	-0.058***	0.035***	1		
(18) SEG_CNT	0.023***	0.031***	0.033***	0.004	0.015**	-0.025***	0.114***	0.011	0.059***	0.181***	0.050***	-0.113***	-0.194***	-0.120***	-0.119***	-0.015*	0.084***	1	
(19) FIRM_AGE	-0.015*	-0.031***	-0.031***	-0.012*	0.028***	0.012*	0.165***	0.004	0.143***	0.377***	0.155***	-0.185***	-0.244***	-0.201***	-0.155***	-0.028***	0.267***	0.198***	1

*, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at the top and bottom 1 percent.
All variables are defined in Appendix D.

2.5.2. Regression results

We used equation (1) to investigate the major determinants of management self-efficacy. Table 7 shows the results of the OLS regression. There is a negative relationship between the current ROA and the self-efficacy measurements in both the presentation session and the Q&A session. And there exists a positive relationship between them and the change in ROA from the previous year.[7] This seems to indicate that concurrent low ROA is associated with higher self-efficacy, or high ROA is concurrent with low management self-efficacy. This could be due to several factors. First, high ROA firms may have less motivation to present an optimistic view about the future through words, when "the facts speak for themselves." More compelling explanation is the idea that low ROA firms shore up their investors' confidence with their own confidence in their ability to perform well in the future. This may be genuine, or it may be a face-saving effort to prevent negative reports from analysts.

Companies with more total assets and lower debt ratio express more self-efficacy than other companies. This may be explained by the fact that big companies have more resources to execute their future plans. The companies with lower debt ratio may be less restricted financially, and this may lead them to reveal the high self-efficacy. If a company has lower book-to-market ratio, it is positively associated with the high level of the management self-efficacy. This can be explained by the fact that if a company has a high value in the capital market, the company is currently evaluated as a more competent company than others. And there are some inconsistency between the management self-

---

[7] The negative relationship between management self-efficacy and current year's ROA is robust if we test it without the change of ROA from the prior year.

efficacy in the presentation session and the Q&A session (i.e., current ratio, total accruals, and firm age).

Table 7: Determinants of the management self-efficacy

	Management self-efficacy		
	(1)	(2)	(3)
Variables	Presentation Session	Q&A Session	Both Sessions
ROA	-0.0711***	-0.0374***	-0.0537***
	(-8.3993)	(-5.4035)	(-8.3486)
D_ROAt	0.0207***	0.0259***	0.0218***
	(3.6116)	(4.3628)	(4.8034)
BIG4	0.0007	0.0058**	0.0031
	(0.1879)	(1.9976)	(1.1131)
LOG_TA	0.0074***	0.0032***	0.0040***
	(6.8876)	(3.7158)	(4.7423)
DEBT_RATIO	-0.0324***	-0.0269***	-0.0270***
	(-4.9304)	(-5.3656)	(-5.4923)
CURRENT_RATIO	-0.0022***	-0.0005	-0.0014***
	(-3.9138)	(-1.0559)	(-3.2150)
RET_VOL	0.0022	-0.0229	-0.0086
	(0.1233)	(-1.4032)	(-0.6169)
ROA_VOL	-0.0039	-0.0009	0.0004
	(-0.5194)	(-0.1206)	(0.0603)
ABS_T_ACCRUALS	-0.0319***	-0.0072	-0.0181**
	(-2.8835)	(-0.7256)	(-2.0644)
BTM	-0.0287***	-0.0142***	-0.0197***
	(-11.9050)	(-6.8836)	(-10.4977)
D_DIVIDEND	-0.0077	-0.0258***	-0.0085
	(-0.6077)	(-2.8557)	(-0.9020)
SEG_CNT	-0.0006	0.0001	0.0001
	(-0.6458)	(0.1825)	(0.1197)
FIRM_AGE	-0.0002*	-0.0001	-0.0001
	(-1.8026)	(-0.9495)	(-1.3509)
Obs.	12,825	12,825	12,825
R-squared	0.1227	0.0888	0.1372
INDUSTRY FE	YES	YES	YES
YEAR FE	YES	YES	YES

Standard errors are clustered by client. t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

The results of our test of the relationship between management self-efficacy and future performance are documented in Table 8. SELF_EFFICACY_PT and SELF_EFFICACY_BOTH are each negatively related with change in ROA one year in the future. This means that higher levels of management self-efficacy in the current year during the conference calls are related to lower changes in ROA in the future, and vice versa. This goes against the notion that higher expressed self-efficacy is indicative of positive future outcomes. It also supports the idea higher self-efficacy is related to overconfidence or dissimulation; the former signifying ineptitude and the latter signifying duplicity.

Interestingly, management self-efficacy in the Q&A session has no significant relationship with future performance. The discussion topics are usually constrained by the analysts' questions during the Q&A session. This might lead to the reduced level of overconfidence or dissimulation of the managers. However, the coefficient itself of the relationship is still negative.

The current ROA level and the change of ROA in the current year have negative associations with the change of ROA in the next year. This is consistent with the descriptive statistics of our sample; the change of ROA in the long run is zero in the sample data (See Table 5 for the change of ROA). In addition, total assets, debt ratio, absolute total accruals, and the number of business segments have positive associations with the change in ROA. And the current ratio, volatility of market return, volatility of ROA, and book-to-market ratio have negative associations with the ROA changes.

Table 8 Management self-efficacy and future financial performance

Variables	ROA change in the next year (D_ROA_{t+1}) (1)	(2)	(3)	ROA improving trend in two years ($CONTI\_D\_ROA_{t+2}$) (4)	(5)	(6)
SELF_EFFICACY_PT	-0.0648***			-0.9654***		
	(-4.9658)			(-3.2181)		
SELF_EFFICACY_QNA		-0.0178			-0.0868	
		(-1.1498)			(-0.2832)	
SELF_EFFICACY_BOTH			-0.0666***			-0.9638**
			(-3.9268)			(-2.4954)
ROA	-0.1943***	-0.1903***	-0.1932***	-2.2818***	-2.2095***	-2.2612***
	(-9.5735)	(-9.3880)	(-9.5231)	(-10.5766)	(-10.2431)	(-10.4741)
D_ROAt	-0.2018***	-0.2027***	-0.2017***	-0.7797***	-0.7959***	-0.7793***
	(-10.1992)	(-10.2533)	(-10.1894)	(-4.1801)	(-4.2622)	(-4.1734)
BIG4	0.0029	0.0029	0.003	0.2068**	0.2070**	0.2086**
	(0.6222)	(0.6338)	(0.6562)	(2.4935)	(2.4928)	(2.5153)
LOG_TA	0.0058***	0.0054***	0.0056***	0.0701***	0.0631***	0.0669***
	(5.7706)	(5.4017)	(5.5895)	(3.2299)	(2.9138)	(3.0861)
DEBT_RATIO	0.0143*	0.0159*	0.0146*	0.2847*	0.3147**	0.2902*
	(1.6804)	(1.8614)	(1.7117)	(1.9213)	(2.1197)	(1.9538)
CURRENT_RATIO	-0.0036***	-0.0035***	-0.0036***	-0.0382**	-0.0355**	-0.0371**
	(-4.2763)	(-4.1129)	(-4.2198)	(-2.4466)	(-2.2880)	(-2.3818)
RET_VOL	-0.0931***	-0.0937***	-0.0938***	-1.5813***	-1.5751***	-1.5914***
	(-2.7636)	(-2.7777)	(-2.7840)	(-3.1727)	(-3.1643)	(-3.1946)
ROA_VOL	-0.0828***	-0.0825***	-0.0825***	-1.2480***	-1.2346***	-1.2385***
	(-4.0591)	(-4.0458)	(-4.0429)	(-4.8437)	(-4.8152)	(-4.8214)
ABS_T_ACCRUALS	0.2542***	0.2561***	0.2551***	2.0999***	2.1286***	2.1111***
	(8.6688)	(8.7257)	(8.6929)	(6.9535)	(7.0640)	(6.9962)
BTM	-0.0308***	-0.0292***	-0.0303***	-0.027	-0.0009	-0.0184
	(-8.0904)	(-7.6560)	(-7.9316)	(-0.4608)	(-0.0147)	(-0.3148)
D_DIVIDEND	-0.0027	-0.0027	-0.0028	0.2367	0.2383	0.2372
	(-0.2635)	(-0.2655)	(-0.2705)	(0.8405)	(0.8501)	(0.8403)
SEG_CNT	0.0029***	0.0029***	0.0030***	-0.006	-0.0053	-0.0054
	(3.9974)	(4.0602)	(4.0618)	(-0.2662)	(-0.2343)	(-0.2382)
FIRM_AGE	0.0001	0.0001	0.0001	-0.001	-0.0009	-0.001
	(1.1851)	(1.3430)	(1.2601)	(-0.5581)	(-0.4699)	(-0.5252)
Observations	12,821	12,821	12,821	12,815	12,815	12,815
R-squared (Pseudo R^2)	0.2155	0.2143	0.215	0.0753	0.0744	0.075
INDUSTRY FE	YES	YES	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES	YES	YES

Standard errors are clustered by client. t-statistics and z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

The findings from the logistic regression of Equation (3) are documented in Table 8 in the last three columns. Sustained increases in ROA are negatively associated with management self-efficacy language and it is negatively related with current ROA. In other words, if the management reveals high management self-efficacy, it has low probability to have the ROA improving trends in the next two years. Consistently with the results of Equation (2), the management self-efficacy in the Q&A session has no significant association with ROA improving trend. This result also supports the overconfidence argument rather than the positive self-efficacy arguments in the business environment.

The sign of the coefficients with most of the control variables have consistent results across two years. The level of ROA and the ROA change in the current year are negatively associated with the future financial performance. Higher total assets are positively related to the ROA improvements in both the following year and the two consecutive years. Current ratio and return volatilities (both in market returns and accounting returns, ROA) have negative associations with future ROA improvements. In our sample, the absolute value of total accruals in the current year is positively associated with the company's ROA changes in the future.

It is an interesting result that the level of management self-efficacy during the Q&A session has no statistical significance with the future financial performance (Column [2] and [5] in Table 8). Participants in the earnings conference call in both the prepared session and the Q&A session are the same management. However, the correlation of management self-efficacy between the prepared session and the Q&A session is relatively low, considering that speakers are the same (0.335 in Table 6). It is still an open question that the analysis results are different between the presentation session and the Q&A session.[8]

Based on the overall statistical analysis, our results support the overconfidence argument (Malmendier and Tate 2005, 2008; Hirshleifer, Low, and Teoh 2012) and dissimulation arguments. Especially, our results are significant with self-efficacy measurements of the prepared-presentation session. The management can carefully choose the words in order to carve the positive impression during the earnings conference call. This study suggests that investors and analysts should be careful to interpret the contents of the earnings conference calls.

2.6. Additional analysis

In this section, we added additional aspects to our management self-efficacy measurement. First, we adjust our measurement by using the two different sessions in a conference call in order to control the entity specific effects. Second, in the previous section, we do not distinguish "*I*" sentences from "*we*" sentences. In this section, we will test the effect of these two types of sentences. Third, we compare our proposed measurements with the existing methodologies. We will compare our measurement with a traditional textual analysis methodology (Loughran and McDonald's word list). Lastly, we will apply the most up-to-date textual technique (Word2Vec) in our setting.

---

[8] In the additional analysis section, we use the difference between the presentation session and Q&A session.

2.5.3.1. Adjusting the measurement of management self-efficacy

Larcker and Zakolyukina (2012) try to control individual-specific effects when they analyze the earnings conference calls. One of their procedures is to take the difference between the presentation session and the Q&A session.[9] By using this method, the entity specific narrative tone effects can be mitigated when we measure the management self-efficacy by analyzing the earnings conference call. Our sample shows that the management self-efficacy level between the presentation session and the Q&A session is significantly correlated (0.335 in Table 6). And the presentation session has a higher level of management self-efficacy, compared to the Q&A session (0.317 and 0.266, respectively, in Table 5). As Larcker and Zakolyukina (2012) expected, unobservable entity-specific characteristics may have an impact on the measured management self-efficacy. We use the same approach of prior literature in taking the difference between the two parts in the conference call. (i.e., the level of Q&A part minus the level of presentation part). Table 9 shows the results. If we remove the entity-specific effects, the management self-efficacy has positive associations with the future financial performance. As we mentioned before, the information contents themselves may be different between these two sessions. Or, the improvisation of the Q&A session may bring less obfuscated management self-efficacy, compared to the prepared remark session. Larcker and Zakolyukina (2012) also point out that how to measure or adjust the tone of the textual information is an interesting and arguable topic.

---

[9] "*We also computed linguistic measures using the difference in the cues between the MD (Management discussion) and Q&A,*" (Larcker and Zakolyukina 2012)

Table 9 Results with the adjusted management self-efficacy

Variables	D_ROA_{t+1} (1)	$CONTI\_D\_ROA_{t+2}$ (2)
SELF_EFFICACY_DIFF	0.0379***	0.6673***
	(3.1321)	(2.7333)
ROA	-0.1909***	-2.2343***
	(-9.4283)	(-10.3799)
D_ROA_t	-0.2034***	-0.8020***
	(-10.2923)	(-4.2995)
BIG4	0.0026	0.2037**
	(0.5713)	(2.4549)
LOG_TA	0.0055***	0.0656***
	(5.5231)	(3.0438)
DEBT_RATIO	0.0162*	0.3134**
	(1.8988)	(2.1161)
CURRENT_RATIO	-0.0036***	-0.0370**
	(-4.1698)	(-2.3785)
RET_VOL	-0.0923***	-1.5563***
	(-2.7401)	(-3.1249)
ROA_VOL	-0.0826***	-1.2439***
	(-4.0491)	(-4.8236)
ABS_T_ACCRUALS	0.2554***	2.1137***
	(8.7107)	(6.9994)
BTM	-0.0295***	-0.0090
	(-7.7543)	(-0.1549)
D_DIVIDEND	-0.0016	0.2499
	(-0.1557)	(0.8949)
SEG_CNT	0.0029***	-0.0057
	(4.0196)	(-0.2540)
FIRM_AGE	0.0001	-0.0009
	(1.3029)	(-0.4932)
Obs.	12,821	12,815
R-squared	0.2148	0.0750
INDUSTRY FE	YES	YES
YEAR FE	YES	YES

Standard errors are clustered by client. t-statistics and z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

2.5.3.1. Distinguishing "*I*" sentences from "*our company*" sentences

When we develop the management self-efficacy measurement, we assume that the comments during the earnings conference call would be the official representation of his/her company, rather than his/her personal opinion.[10] In this additional test, we exclude sentences whose subject parts have a singular first person pronoun from the management self-efficacy sentences. Those sentences with the pronoun "*I*" or "*my*" may be related to the speaker's personal opinion, which does not necessarily reflect the assessments of their company's ability. Thus, we exclude sentences such as "*I can*" or "*my skill can*" sentences. We use this modification when we measure the level of management self-efficacy and compare the results.[11]

2.5.3.2. Comparing positive and negative word list

Word lists or bag-of-words approach is widely used for the accounting textual analysis. And specific word lists were proposed for the textual analysis, especially for the accounting or finance narratives. In this section, we applied these existing word lists for the prediction of firm future performance.

One of the most widely used word lists in the accounting domain is Loughran and McDonald's positive and negative word lists (Loughran and McDonald 2011). It has the

[10] As discussed in the hypothesis development section, we did not differentiate these two sentences: (1) "I, as a CEO, can encourage our research team to produce innovative products." and (2) "Our company research team has full capacity to launch innovative products."

[11] We found that sometimes, the speakers indicate their company by using the phrases "the company." So, we include those sentences as self-efficacy sentences (i.e., 'the company can', 'the company will', and 'the company has been' sentences.) For the detailed definition of the self-efficacy sentences, please see Table 1.

positive word list for measuring the positive tone of the accounting textual materials, and the negative word list for the negative tone. We use these word lists for the same analysis which we performed in the previous section.

We analyze our transcripts at the sentence level. If a sentence contains any of the positive words, it is classified as a positive sentence. In the same way, if a sentence has any of the negative words, it is categorized as a negative sentence. If a sentence has both types of words, or none of these words, the sentence is classified neither a negative sentence nor a positive sentence. And then we measured the ratio between the number of the positive sentences and the total number of the sentences in the earnings conference call. We did the same with the negative sentences in order to measure the negative tone.

Table 10, Table 11, and Table 12 show the results of the additional analysis. Separating the first person singular pronouns ("*I*" and "*my*") from the first person plural pronouns, "*we*" and "*our*", makes little difference in the results. The correlation is nearly 1 (0.999). And the results stay consistent (Column [1] and [2] in Table 11). Positive tone has significant associations with the future financial performances (Column [3] and [4] in Table 11). The more positive tone expresses the management during the earnings conference call, the higher ROA they will produce in the two consecutive years. For the Q&A session, the result is insignificant (Column [1] in Table 12). The results with negative tones are insignificant over the years (Column [4], [5] and [6] in Table 12).

We compared the positive tone and the management self-efficacy by putting together (Column [5] and [6] in Table 11). Interestingly, both relationships remain significant with the same directions. Moreover, the t-statistics become stronger on both measurements. This may indicate that our self-efficacy measurement captures the unmeasured area by the existing positive tone measurement. [12]

Table 10 Correlation Matrix with additional variables

Variables	(1)	(2)	(3)	(4)
(1) SELF_EFFICACY_PT	1			
(2) SELF_EFFICACY_PT_Excluding_I_my	0.999***	1		
(3) LM_POSITIVE_PT	0.563***	0.563***	1	
(4) LM_NEGATIVE_PT	-0.245***	-0.247***	-0.400***	1

*, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

Table 11 Analysis with the modified measurement and positive tone

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	D_ROA_{t+1}	$CONTI\_D\_ROA_{t+2}$	D_ROA_{t+1}	$CONTI\_D\_ROA_{t+2}$	D_ROA_{t+1}	$CONTI\_D\_ROA_{t+2}$
SELF_EFFICACY_PT					-0.1106***	-1.8942***
					(-6.6050)	(-5.4197)
SELF_EFFICACY_PT_ Excluding_I_my	-0.0652***	-0.9867***				
	(-4.9853)	(-3.2895)				
LM_POSITIVE_PT			0.0267*	0.7499**	0.0979***	1.9635***
			(1.7032)	(2.0962)	(4.8856)	(4.7544)
Obs.	12,821	12,815	12,821	12,815	12,821	12,815
R-squared	0.2155	0.0754	0.2144	0.0748	0.2169	0.0773
Controls	YES	YES	YES	YES	YES	YES
INDUSTRY FE	YES	YES	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES	YES	YES

Standard errors are clustered by client. t-statistics and z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

[12] Among our 28 words in our *KevenNBen* word list and 354 words in the positive word list of Loughran and McDonald's, only one word, "*able*" is commonly shared by both word lists.

Table 12 Regression results with the future financial performance and the existing proposed tones

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	D_ROA $_{t+1}$	CONTI_ D_ROA $_{t+2}$	D_ROA $_{t+1}$	CONTI_ D_ROA $_{t+2}$	D_ROA $_{t+1}$	CONTI_ D_ROA $_{t+2}$
LM_POSITIVE_QNA	0.0183	1.3628***				
	(0.9540)	(3.7524)				
LM_NEGATIVE_PT			-0.0946***	-0.0994		
			(-3.8996)	(-0.1985)		
LM_ NEGATIVE _QNA					-0.0459	-0.5682
					(-1.3202)	(-1.0073)
Obs.	12,821	12,815	12,821	12,815	12,821	12,815
R-squared	0.2143	0.0755	0.2152	0.0744	0.2144	0.0745
Controls	YES	YES	YES	YES	YES	YES
INDUSTRY FE	YES	YES	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES	YES	YES

Standard errors are clustered by client. t-statistics and z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

2.5.3.3. Developing a word list by using the Word2Vec method

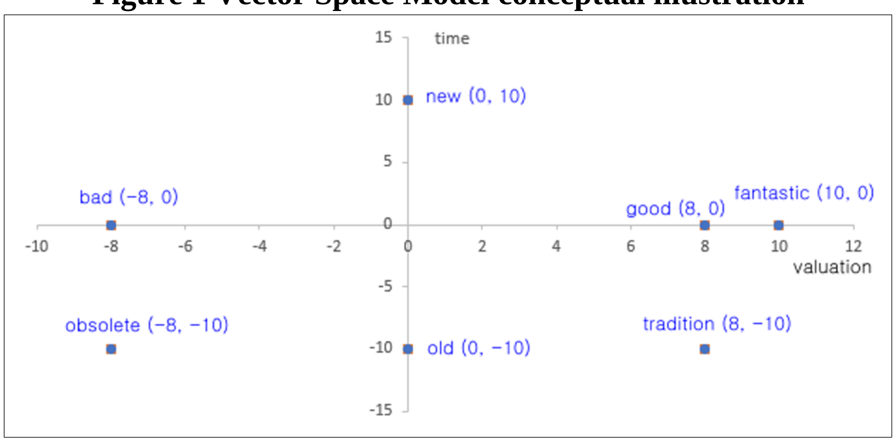
In this section, we will introduce an up-to-date statistical language methodology to improve our current management self-efficacy measurement.[13] By using the Word2Vec methodology, we can build a word list that bears efficacy meanings from words of our target documents, the earnings conference call transcripts. We will apply this new word list in order to measure the management self-efficacy. Before discussing detailed results, we will very briefly introduce the vector space model and Artificial Neural Network on which the Word2vec is based.

Statistical language researchers or computational language researchers develop ways to convert language into numerical information. Once it is converted into numerical

[13] The new methodology, Word2Vec, was actually proposed more than a decade ago (Bengio et al. 2003). And it gains popularity by reducing the amount of the computational requirements for the implementation (Mikolov et al.2013b). However, to our best knowledge, in the accounting domain, Jung et al. (2020) is the only literature that uses the Word2Vec methodology.

information, it can be further analyzed or processed by various numerical models. However, the meaning of language is much more complex by nature. Therefore, it is always challenging to convert the language into numerical information while preserving the full semantic meaning. The Vector Space Model converts words into numerical values. The simplest one would be representing each word by using a single numerical value. For example, we can map the word *good* to a value of 8 and *bad* to -8. The word *fantastic* or *excellent* would be 10. In this case, a higher numerical value corresponds to more positive valuation. However, the single numerical value can convey only limited information. In order to solve this limitation, more dimensions can be added. For example, we can add a time-related dimension in our word representation model. For example, the words *new*, *old*, *tradition*, and *obsolete* are all somehow different from each other. The word *new* or *old* can be a value-neutral word. The words *tradition* and *obsolete* contain positive and negative evaluation, respectively. Figure 1 illustrates one conceptual example of two-dimension word representations. In this case, the first coordinate value is related to the valuation; the second coordinate value is corresponding to the time aspects.

Figure 1 Vector Space Model conceptual illustration



Depending on the situation, a researcher can add as many as dimensions into his or her representation system as he or she might need. The important principle, in any multi-dimensional model, is to preserve the relative semantic relations between given words.

Recently, the Word2Vec method is widely used in order to convert language into the VSM (Young et al. 2017). This method deploys an Artificial Neural Network (ANN) in order to convert a word into a numerical vector. An ANN is a general-purpose model for the prediction or classification tasks. In a shallow artificial neural network, there are three sequential layers: input nodes, hidden nodes, and output nodes. Panel A Figure 2 shows a basic structure of the three-layer ANN model.[14] [15]

For the Word2Vec methodology, the number of input nodes becomes the number of unique words in the target documents, and each word corresponds to each input node. The ANN model will have the same number of output nodes (i.e., the number of unique words). The number of hidden nodes is set by the number of the desired dimensions in the VSM.

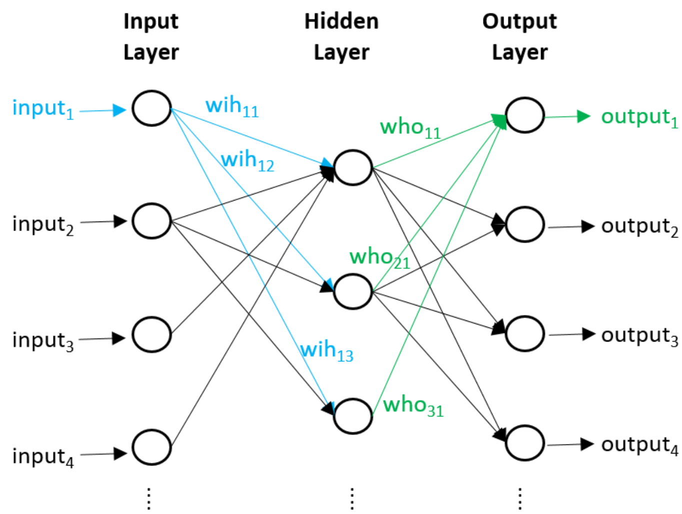
---

[14] Variables shown in Figure 2 are all scalar values (i.e., $input_1$, $input_2$, wih_{11}, wih_{12}, who_{11}, $output_1$, etc.). An ANN model receives an input and converts it into output results like the following. The two values connected to each other in Figure 2 are multiplied. And then, those values are added up into one singular scalar value. That singular value (one scalar value) is passed over to the next layer. The hidden layer converts values by using non-linear function (activation function). The same procedures are performed between the hidden layer and output layer. By comparing the output values, the final classification result or the output result is decided (i.e., selecting the largest value among multiple output values, or choosing the maximum value based on the Softmax function). During the training process, all the weights are updated in order to produce the correct output values (Back-propagation). The prepared input and output data are used for the training process. After the training process, we can use the trained ANN for the prediction with a new dataset. The structure of ANN can be modified depending on researchers' purposes. For more details of the ANN and its variations, please see Alpaydin (2014) and Rashid (2016).

[15] Our Word2Vec method uses a simple 3-layer ANN model. It does not use a non-linear activation function in the hidden layer (Mikolov et al.2013c). However, there are also other proposed ANN VSMs with the non-linear activation function in the hidden layer (i.e., Bengio et al. 2003).

Figure 2 Word2Vec flow and structure illustration

Panel A: An Example of 3-layer Artificial Neural Network structure



Panel B: Sample text for the illustration

Sample text:

A cat sits on the yellow mat. The cat is enjoying a sunny day…

Training data point:

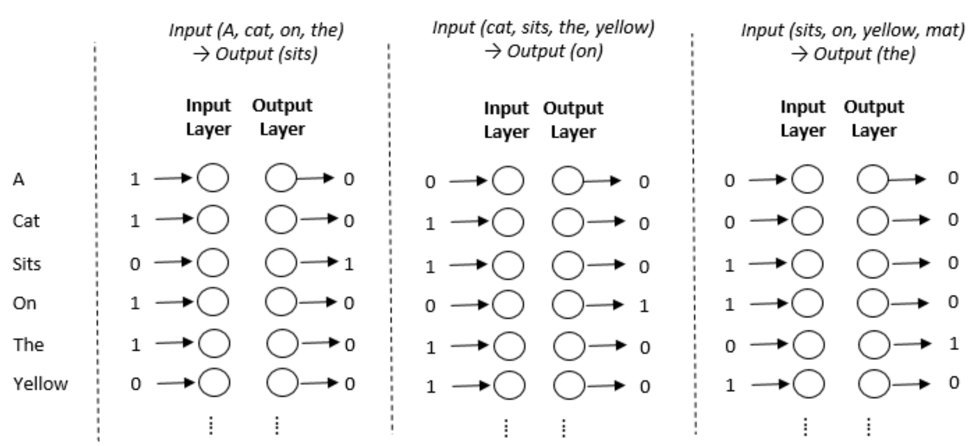
Input (A, cat, on, the) → Output (sits)

Input (cat, sits, the, yellow) → Output (on)

Input (sits, on, yellow, mat) → Output (the)

⋮

Panel C: Data points and training the Word2Vec model illustration



In a general ANN model, input values can be any scalar values. In the Word2Vec implementation, input values are assigned with the binary value (0 or 1).[16] After setting this structure, the ANN model (Word2Vec) is trained by using sentences from target documents.[17] Panel C Figure 2 shows the training examples. Based on the input values, the training process updates the weights between nodes in order to produce the desired output value. [18] After the training process, the n weights of the *i*-th input node become the corresponding vector of the *i*-th word.[19] [20] [21] In other words, we can convert each word into a vector with n-dimension by using those n weights of each node.

These word representations are based on the distributional hypothesis. It assumes that words of similar meaning are plotted near to each other (Turney and Pantel 2010). So, if we select words that are closely located to one specific word, it is expected that

[16] Usually, one-hot vector is used as an input in the Word2Vec model. In the Word2Vec, input and output are one word, generally. So, if a word is assigned as *i*-th word in the entire word list, the corresponding one-hot vector as an input is a vector that has 1 at the i-th element and 0 with other elements.

[17] There are two proposed methods in order to train the Word2Vec model. One is predicting a target word (center word) by using context words (surrounding words) from a sentence. The other method is predicting a context word by deploying a target word. The former is called Continuous Bag-of-Words Model, and the latter is named Continuous Skip-gram Model (Mikolov et al. 2013c).

[18] Panel C Figure 1 shows an example that has 4 surrounding words (i.e., the word window is 2). The weights and the hidden layer are omitted for simplicity. In the original paper (Mikolov et al. 2013c), input values are concatenated one-hot vectors. And the converted values are averaged in the hidden layer. The final results would be the same between the original paper's ANN structure and the illustration of Panel C Figure 2, because there is no non-linear activation function in the hidden later and each word shares the same weights. Please see the illustration and the figure in the original paper for more details (Mikolov et al. 2013c; Le and Mikolov 2014).

[19] Mikolov et al. (2013a) propose to use the weights with the input-layer. Meanwhile, each output node has n weights, too. The VSM value of each word can be obtained by using the weights of the output nodes as well. One of those weights, or both weights can be used as a word vector of each word.

[20] This method itself is not complex, but it needs huge computational power. In order to solve this problem, Mikolov et al. (2013b) introduce a new strategy, the negative sample, which reduces necessary computational power while producing reasonably good results.

[21] For the actual implementation, we used the Gensim Python package (https://radimrehurek.com/gensim/).

those words share the similar meaning. Based on the distributional hypothesis, we can derive a list of words from a given word.

Now, we will find the synonyms based on our six predefined efficacy words—'*ability, abilities, able, capability, capabilities, and capable*' by deploying Word2Vec model. As we discussed, we trained our Word2Vec model by using the earnings conference call transcripts. And then, we can map every word into a vector space. Based on this, we simply select words which are closely located to our six core efficacy words in our vector space.

Interestingly, some words, such as '*mastery, recourses, cohesiveness, dependability, and unrivalled*' are correctly identified although each word is used less than 22 times in our entire earnings conference call transcript sample (In our sample, the total number of words is about 106 million). It is necessary to include words that are more frequently used in order to compare the results with existing word list (i.e., *KevinNBen* word list). So, we choose words that are used more than 50 times in our sample. (Again, the total number of words in our sample is about 106 million.) Table 13 shows the word list obtained by Word2Vec as well as our *KevinNBen* word list. There are multiple ways of implementing (or training) the Word2Vec method. We applied eight different settings with Word2Vec in order to get a word list. [22]

Some words generated by Word2Vec have opposite meanings to *efficacy*, such as *vulnerabilities* and *inability*. This indicates one of the limitations of the Word2Vec method. We can, in some cases, solve this problem by increasing the number of

[22] The detailed settings of each method and full word list are documented in Appendix E.

dimensions of the vector space. In this example, we use 100 dimensions for each word. Except these two words, all other words are relatively related to the concept of efficacy.

Table 13 Identified word list by using the Word2Vec methodology

Word Category	Words	Total frequency in the transcripts
KevinNBen (22 words)	approach, approaches, approached, approaching, capacity, <u>capacities</u>, complete, completed, grow, grows, grew, growing, grown, growth, growths, milestone, milestones, ready, patent, patents, quality, and qualities	580,079
Word2Vec (Continuous Bag-of-Words method, 50 words)	expertise, skills, agility, credentials, flexibility, desire, resources, knowhow, competencies, strengths, processes, tools, talent, versatility, algorithms, specialization, scalability, solutions, talents, functionality, credibility, knowledge, commitment, vulnerabilities, infrastructures, <u>capacities</u>, willingness, technology, freedom, ecosystems, platform, mission, teams, backbone, wherewithal, workflow, autonomy, creativity, continuity, relevance, efforts, relevancy, needs, roadmaps, relationships, infrastructure, designed, prowess, profitably, and skill	270,734
Word2Vec (Skip-gram method, 50 words)	expertise, knowhow, prowess, flexibly, skills, competencies, scalability, agility, platform, wherewithal, adaptability, flexibility, knowledge, scalable, versatility, unparalleled, enables, tools, autonomy, scale, unmatched, enable, technology, ecosystems, intimacy, decisioning, processes, unrivaled, enabling, strengths, unparallel, competence, solutions, talent, enabled, sophistication, differentiated, talents, breadth, proficiency, competency, efficiently, resources, firepower, <u>capacities</u>, bioinformatics, optimized, adaptable, advantages, and nurture	205,868

The common word that is included in KevinNBen word list is underlined.

There are two different approaches in training the Word2Vec method. [23] We test both methods. And we need to decide how many words we will select from the Word2Vec model. There is no pre-fixed number in this. We try multiple combinations. We choose the key words of 25, 50, 100, and 200 words (Table 13 shows the lists of 50 words).

[23] One is Continuous Bag-of-Words (CBOW), and the other is Continuous Skip-Gram (SG) Model (Mikolov et al. 2013c).

The regression results with new word lists are documented in Table 14 and Table 15. We have all six rules for the identification of the self-efficacy sentences (See Table 1). One of those six rules is to use KevinNBen word list. In this section, we replaced this rule with the Word2Vec word list. The rest of the rules remain the same. Table 14 shows the correlation matrix among the new variables. Correlation reduced more than 20% although only one rule is replaced from the original six rules. The correlations between the original self-efficacy and the modified self-efficacy with Word2Vec word list are around between 0.8 and 0.75. Table 15 shows the additional analysis results with these new word lists. Interestingly, the new measurement with Word2Vec word list has more statistically significant associations with the company's future performance (Column (3) to (6) in Table 14).

The results show that the Word2Vec method effectively suggests a list of words that are related to the *efficacy*. The total frequency of these word usage is half of the usage of the *KevinNBen* word list (see Table 13). However, it effectively identifies the relevant words. t-statistics from the regression analysis are increased from 4.45 into 8.28 (Column (1) and (3) in Table 15). If the new word list is combined with the positive tone measurement, the significance is more increased (Column (9) in Table 14). Among our 8 word lists which are generated by the Word2Vec methodology, the word list with 200 words from Skip-Gram method performs best in terms of the t-statistics and z-statistics. The full list of 200 words are documented in Appendix E. The word list with 25 words from CBOW method performed less significant results. However, the results are still much more significant than the original KevinNBen word list (t-statistics: 7.70 and z-statistics: 3.97; this is untabulated).

Table 14 Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) SELF_EFFICACY_PT	1									
(2) Word2Vec_CBOW_25_PT	0.807***	1								
(3) Word2Vec_CBOW_50_PT	0.799***	0.976***	1							
(4) Word2Vec_CBOW_100_PT	0.795***	0.960***	0.988***	1						
(5) Word2Vec_CBOW_200_PT	0.756***	0.888***	0.921***	0.936***	1					
(6) Word2Vec_SG_25_PT	0.804***	0.975***	0.983***	0.970***	0.904***	1				
(7) Word2Vec_SG_50_PT	0.805***	0.979***	0.987***	0.977***	0.914***	0.993***	1			
(8) Word2Vec_SG_100_PT	0.797***	0.968***	0.980***	0.975***	0.921***	0.982***	0.990***	1		
(9) Word2Vec_SG_200_PT	0.788***	0.942***	0.967***	0.971***	0.935***	0.955***	0.963***	0.973***	1	
(10) LM_POSITIVE_PT	0.559***	0.468***	0.474***	0.488***	0.490***	0.469***	0.478***	0.486***	0.486***	1

*, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

Table 15 Relationship between the future performance and Word2Vec word list

Variables	(1) D_ROA $_{t+1}$	(2) CONTI_D_ROA $_{t+2}$	(3) D_ROA $_{t+1}$	(4) CONTI_D_ROA $_{t+2}$	(5) D_ROA $_{t+1}$	(6) CONTI_D_ROA $_{t+2}$	(7) D_ROA $_{t+1}$	(8) CONTI_D_ROA $_{t+2}$	(9) D_ROA $_{t+1}$	(10) CONTI_D_ROA $_{t+2}$
SELF_EFFICACY_PT	-0.0648***	-0.9654***								
	(-4.9658)	(-3.2181)								
Word2Vec_CBOW_50_PT			-0.1330***	-1.2761***						
			(-8.9263)	(-3.9720)						
Word2Vec_SG_50_PT					-0.1354***	-1.3763***				
					(-8.8418)	(-4.1956)				
Word2Vec_SG_200_PT							-0.1268***	-1.2896***	-0.1780***	-2.2721***
							(-9.2561)	(-4.3336)	(-10.9333)	(-6.6034)
LM_POSITIVE_PT									0.1440***	2.1563***
									(7.5346)	(5.2592)
Obs.	12,821	12,815	12,821	12,815	12,821	12,815	12,821	12,815	12,821	12,815
R-squared	0. 2155	0.0744	0.2188	0.0758	0.2188	0.0760	0.2190	0.0760	0.2219	0.0784
CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
INDUSTRY FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Standard errors are clustered by client. t-statistics and z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix D.

Based on the results, the *KevinNBen* word list contains more general terms which are more frequently used compared to the Word2Vec word list. Due to its general usage, we can argue that our *KevinNBen* word list can be applicable to more general situations. For the conference call situations, the Word2Vec model can be a good alternative for developing a word list based on a few core concept words.[24]

The usefulness of the domain specific word list is underscored in the prior literature (Loughran and McDonald 2011). Loughran and McDonald (2011) manually identified the word list based on the word usage frequency. Their word list is widely used in the accounting and finance domain.

By using the Word2Vec, this research shows that we can build a list of words effectively and efficiently by feeding the relevant domain documents to the Word2Vec model. We apply the Word2Vec methodology in order to identify the positive word list on our own. Table 16 shows our result. We build the word list by selecting the 50 closest words to the core positive word list. [25]

This method shows additional words such as "*nice, healthy, clear, solid, robust, and meaningful.*" These words are not included in the existing word list (Loughran and McDonald 2011). The existing words are constructed based on the 10-K filings, and our word list is built from the earnings conference call transcripts. We may argue that our

---

[24] In this study, we used six core words in order to develop a word list, 'ability, abilities, able, capability, capabilities, and capable.'

[25] We choose 16 core positive words based on Loughran and McDonald's positive word list (2018 version). The words are "best, enable, enjoy, good, great, happy, improve, optimistic, perfect, pleasant, positive positively, strong, succeed, success, and win."

word list can capture the more comprehensive positive tone.[26] [27] This would be another future research.

Similarly, Liu and Moffitt (2016) use the strong modal in identifying the restatement from the SEC comment letters. Table 17 shows the original word list and the newly identified word list based on the original word list. In this analysis, we just use all the original 19 words as the core words. We can find additional words bearing the '*strong'* meaning by deploying the Word2Vec method (i.e., absolutely, certainly, incredible). We may extend the existing word list by using the Word2Vec method.

Table 16 Positive word list by using the Word2Vec methodology

50 Positive Words from the conference call transcripts
compelling, nice, beneficial, terrific, satisfactory, fruitful, healthy, wonderful, remarkable, powerful, clear, fantastic, solid, stunning, pleasing, robust, spectacular, formidable, convincing, phenomenal, attentive, stellar, satisfying, encouraging, respectable, brisk, resilient, decent, credible, rewarding, superb, gratifying, receptive, doable, reassuring, demonstrable, confident, attractive, generous, vibrant, promising, impressed, excellent, conducive, superior, meaningful, buoyant, viable, lucrative, and visible (*And 16 core positive words: best, enable, enjoy, good, great, happy, improve, optimistic, perfect, pleasant, positive positively, strong, succeed, success, and win*)

The proposed words by Loughran and McDonald are underlined.

[26] There is one additional consideration when we measure the tone of the conference calls. Some of the phrases may not reflect the actual positive tone (i.e., "have a nice day; good morning; good afternoon; that's a good question, etc."). We may need to eliminate these utterances from the positive tone when we measure the positive tone. We may not find the complete utterances of these cases. However, we can identify the reasonable cases of phrases if we use the N-gram frequency with those words.

[27] For the information, the word '*successful*' appeared on the 56th. '*Profitable*' is located on the 76th on our word list. However, there are some limitations as well. The words 'lackluster, challenging, and unattractive' are in the list on the 53rd, 60th, and 61th.

Table 17 Strong Modal word list by using the Word2Vec methodology

Word Category	Words
Original Strong Modal Word list (from Loughran and McDonald 2011)	always, best, clearly, definitely, definitively, highest, lowest, must, never, strongly, unambiguously, uncompromising, undisputed, undoubtedly, unequivocal, unequivocally, unparalleled, unsurpassed, will
Word2Vec Methods #1 (the words used more than 50 times in the entire sample)	unmatched, unparallel, invaluable, unrivaled, enduring, credible, exemplary, elusive, absolutely, unwavering, fierce, certainly, incredible, formidable, unrelenting, uniquely, unacceptable, objectively, inspiring
Word2Vec Methods #2 (the words used more than 1,000 times in the entire sample)	absolutely, certainly, incredible, uniquely, consistently, truly, firmly, compelling, greatest, acceptable, clear, essential, indeed, enormous, optimal, dominant, exceptional, also, adequate

1) Each word list has 19 words in the list. The entire sample transcripts have more than 106 million words.
2) Original words are used as core words. Thus, they are not included in the Word2Vec word list.

2.7. Conclusion and contributions

In this chapter, we propose a new metric for measuring business communications, management self-efficacy. We base our measure on modified versions of existing word lists, new word lists, parts of speech and sentence structure. This paper argues that earnings conference call transcripts are an ideal dataset for the research questions explored here. Earnings conference calls typically contain narratives and answers to questions by the CEO, CFO, and other C-level officers, and their language reveals their companies' attitudes about the future. Our paper is unique in self-efficacy research in that it has a large sample size of over 12,000 observations spanning 12 years. The survey-based nature of self-efficacy research limits samples to under 1,000 observations. (Forbes 2005; Hmieleski and Corbett 2008).

We find that management self-efficacy has prediction power for two-year ROA trends. Management self-efficacy has a negative relationship with future financial performance.

Our research contributes to the existing literature in many ways. First, we augment self-efficacy literature by proposing the new measurement of management self-efficacy. Our method is the first to measure self-efficacy using textual analysis.

Second, we provide empirical evidence that high management self-efficacy is related to negative future performance at the organizational level. There are abundant studies on individual self-efficacy (Harrison et al. 1997) as well as group and organizational efficacy (Prussia and Kinicki 1996; Peterson et al. 2000; Gully et al. 2002; Forbes 2005; Hmieleski and Corbett 2008). Unlike the existing literature, we apply a novel approach for measuring the management self-efficacy: textual analysis. This approach allows us to use a larger sample size than in previous studies. Furthermore, the prior literature has been mainly limited to new venture companies. However, in this study, we include all of the major listed and long-standing companies in our sample, which makes our findings more generalizable in the business context.

Third, we propose a new textual measurement that incorporates sentence structure to include more context than a standard bag-of-words approach. We also incorporate the co-occurrence of words and phrases. Thus, it provides a more comprehensive understanding of the texts.

Fourth, the application of the Word2Vec methodology for developing a word list suggests new possibilities in accounting and finance research. By analyzing documents themselves without any human judgment, the Word2Vec method effectively identifies the words that have a similar meaning to the target words. Future textual analysis may benefit from developing a word list using this method.

Lastly, our new measure of management self-efficacy may be applicable to many contexts and open a new stream of research. Unobtrusive personality measures such as this, which can be automated, can have many applications in other financial disclosures or related domains including finance, marketing, management, and psychology.

Chapter 3: Accounting misreporting and tone difference between CEOs and CFOs during earnings conference calls

3.1. Introduction

Auditors and investors can utilize multiple channels to make their decisions regarding companies and their clients. Besides the numerical financial information, textual disclosures—both mandatory and voluntary—have increasingly become major sources of information to guide decision-making. Earnings conference calls are one of the voluntary disclosures channels that stakeholders can use.[1] Earnings conference call are actively studied to find its unique roles as corporate disclosure channels (Matsumoto et al. 2011; Larcker and Zakolyukina 2012; Blau et al. 2015; Burgoon et al. 2016; Huang et al. 2018). Unlike other corporate communication channels, earnings conference calls usually feature extemporaneous discussions during the Q&A session. By nature, this reflects the individual personal characteristics of the participants as well as the corporate characteristics (Green et al. 2019). These conference calls' unique setting (i.e., extemporaneous Q&A session and reflection of personal characteristics) has produced useful resources for researchers (Mayew 2008; Hollander et al. 2010; Lee 2016; Merkley et al. 2017). In this study, we use earnings conference call transcripts to identify accounting misreporting.

[1] It is possible to argue that once a company starts to have an earnings conference call, it becomes an annual (quarterly) mandatory event. However, companies still choose not to open earnings conference calls. In addition, Hollander et al. (2010) document that management sometimes decides not to answer a specific question during the Q&A session in the conference call. Based on these facts, we can classify the earnings conference call as a voluntary disclosure, or at least, a pseudo-voluntary disclosure.

Information asymmetry between stakeholders provides an incentive to use deceptions (Akerlof 1978). Despite audit reports and other regulations on disclosure, accounting misreporting continues, with huge social cost. As a result, research that aims to uncover and predict accounting misreporting has been widely documented (Dechow et al. 1995; Sloan 1996; Beneish 1999; Dechow and Dichev 2002; Hribar et al. 2014; Dechow et al. 2011; Price et al. 2011; Purda and Skillicorn 2015; Perols et al. 2017).

This study examines whether this unique setting of earnings conference calls can be utilized to identify and predict companies' financial misreporting. We find that CFOs respond more sensitively compared to CEOs when accounting misreporting exists. In other words, a CFO's negative tone during a conference call increases more rapidly than the negative tone of a CEO. Accounting information users can utilize our findings for their decision-making, especially in risk assessments.

Our research contributes to the literature on accounting and textual analysis. First, investors, auditors, and regulators can employ our measurement to assess the level of risk of companies with respect to accounting misreporting. Second, our study provides empirical evidence about personal characteristics depending on job titles. While prior research mainly addressed the personal characteristics of the CEO (Kaplan and Sorensen 2017; Mayew et al. 2013; Falato et al. 2015; Green et al. 2019), our study provides findings that focus specifically on responses to accounting misreporting, employing multiple job titles in the same company, including the CFO. Third, this research provides evidence about the difference between presentation sessions and question-and-answer sessions. Some research has not differentiated between the two sessions (Larcker and Zakolyukina 2012), while other research takes advantage of the differences between two parts (Lee 2016). Our study extends the conference call research by contributing our findings about the differences between presentation sessions and the question-and-answer sessions. Fourth, we also compare the models based on the misclassification cost. The results show that our model performs better than—or at a similar level to—the model previously applied in the literature (i.e., the abnormal audit fee model). Lastly, we introduce an up-to-date NLP model in the form of the Doc2Vec methodology. We demonstrate the new possibility of using the Dov2Vec model in the accounting disclosure domain.

This rest of this chapter is organized as follows: in Section 2, we review prior literature, and in Section 3, we develop the hypotheses. In Section 4, we document the research sample and design. In Section 5, we discuss the results of the main hypothesis testing. We extend our discussions with the additional analysis, including analysis of the presentation sessions, implementation of the Dov2Vec model, and comparison of various existing models in Section 6. We conclude this chapter in Section 7.

3.2. Literature review

Researchers try to uncover deceptions by analyzing the linguistic cues. However, the findings are mixed, and it is a still challenging topic. It is pointed out that the deception cues are difficult to find in the experimental settings (Vrij 2004; Lancaster et al. 2013). The reason is that emulating high-stakes settings is difficult under experimental research methods that are restricted by ethical guidelines preventing punishments to participants. In addition, if speakers can have an opportunity to rehearse speech before the official saying, the speakers intentionally avoid some of the pre-identified or well-known cues of the deceptions. For example, it is conceived that the liars would have a lack of details. However, if a speaker prepares an untruthful speech in advance, the deceiver can provide more information than a truth teller (Zhou et al. 2004; Braun et al. 2015).

Responding to this, some researchers try to identify both cues (Burgoon et al. 2016). In other words, the *unskillful* liars may reveal straight-forward cues; the *skillful* tellers use sophisticated cues to conceal the truth. Burgoon et al. (2016) define the former as non-strategic spoken language and the latter as strategic spoken language. These double twisted dynamics make it tricky to distinguish deceptions from true statements.

Unlike other speeches, earnings conference calls provide various unique settings. These characteristics bring attention from researchers. Burgoon et al. (2016) focus on high-stake settings of the earnings conference call. When the management provides false information during the earnings conference calls, the consequences vary greatly depending on whether the false information is identified or not by listeners. (Burgoon et al. 2016; Lee 2016). This high-stake setting cannot be reproducible by using the experimental setting (Burgoon et al. 2016).

Besides this high-stake setting of earnings conference calls, the earnings conference calls have unique structure. It has two distinctive parts. The first part is the management presentation about the company's operation outcomes. This part is usually prepared in advance by the management (Lee 2016). The other part is an extemporaneous question-and-answer part between the outside participants and the management.

This unique setting is explored by the researchers. Lee (2016) documents that not all managers respond to questions spontaneously. He argues that some managers prepare the anticipated questions and they read the prepared scripts as a response when they receive the anticipated questions from the analysts.[2] Lee (2016) also documents that if the manager remains on the scripted answers during the Q&A session, investors consider this behavior as 1) a negative signal about the inability of the managers to run the company (Kaplan et al. 2012) , or 2) cue of existing negative private information that the manager tries to conceal. This leads to negative reactions in the capital market.

There are also other studies which focus on the difference between the prepared part and the Q&A part during the earnings conference call. Matsumoto et al. (2011) document that the Q&A provide additional information to investors over the prepared presentation part. In addition, some managers discriminate unfavorable analysts from favorable or prestigious analysts by limiting question opportunities during the conference calls (Mayew 2008). Or, managers explicitly deny answering certain questions (Hollander et al. 2010). All these show that the manager tries to prepare the Q&A session in advance. However, these studies also show that the manager cannot perfectly deal with the question-and-answer session. Researchers and investors try to uncover the information cues from the earnings conference calls, especially during the Q&A session.

There is not much previous research using the earnings conference calls to detect accounting misreporting. Burgoon et al. (2016)'s analysis is one of the extensive studies related to our research. They analyze six quarterly earnings conference calls by using

[2] Lee (2016) assumes that the style of written language is different from the style of speech based on the prior research (Rowley-Jolivet and Carter-Thomas 2005). He shows the validity of his construct by using the conference call transcript samples.

written transcripts as well as the audio vocal information. They show potentials about how we can use the earnings conference call transcripts in order to identify the accounting fraud. Although they conduct extensive analysis, they briefly analyze the calls with respect to the difference between the CEO and the CFO. They did not use the differences between the CEO and the CFO for the identification of the misreporting. In addition, their sample is limited to a single company.

Larcker and Zakolyukina (2012) deploy a large sample of the earnings conference call transcripts. They used 29,663 quarterly earnings transcripts. They measure the word frequencies such as personal pronouns, positive and negative emotions words, shareholder value related words, etc. However, they do not focus on the difference between the CEO and the CFO, either.

For more general studies in predicting material accounting misstatements, it has a long history and has been widely examined. Beneish (1999) develops a model with the financial variables in order to identify the misstatements (so called M-score). There are multiple versions of the accrual models for this purpose, as well. Dechow et al. (1995) modifies the original Jones model by Jones (1991) (so called the Modified Jones Model). Sloan (1996) uses working capital accruals. Dechow and Dichev (2002) introduce the accruals quality in their model. Dechow et al. (2011) also develop a similar model with M-score with mainly financial variables (so called F-score). Hribar et al. (2014) deploy the unexplained audit fee for identifying the misstatements. Price et al. (2011) compare these academic models with the commercial model (Accounting and Governance Risk by Audit Integrity, LLP). They provide evidence that the commercial model outperforms the academic models in identifying accounting misreporting.

3.3. Hypothesis development

As discussed in the previous section, the earnings conference call has two distinct sessions. They are the prepared presentation part and the instant question-and-answer part between the management and the analysts. In addition, we can identify exact individuals who spoke those specific comments during the conference call. Furthermore, we can obtain the information about the participants' names and their job titles (e.g., CEO, CFO, COO, etc.).

This unique setting of the earnings conference call may reveal cues with respect to accounting misreporting of companies. One of the findings in the prior literature is that negative emotion is related to deception (Vrij et al. 2000; Bond and Lee 2005; Larcker and Zakolyukina 2012; Braun et al. 2015; Burgoon et al. 2016). This is consistent with the hypothesis that negative emotions can be expressed because of being fearful of the detection or feeling of guilt. At the same time, positive emotions can also be used, strategically, in order to disguise these negative emotions by the speakers (Buller and Burgoon, 1996; Burgoon et al. 2016). Control perspective theory argues that deceivers, intentionally and strategically, can remove any cues that might cause problems to the deceivers. A meta-analysis shows that the positive and negative emotions are related to the deception (Hauch et al. 2014).

As a result, negative emotions can be either high or low during the earnings conference call with accounting misreporting. Under the business setting, we may consider one more aspect that is future litigation risk. If the false information is uncovered later, the management would face huge negative consequences. It can be monetary losses or even criminal punishment (i.e., imprisonment). In order to avoid the worst-case scenario, the CEO or CFO may want to plant a last resort during a conference call. In other words, they may put the negative clauses around accounting misreporting issues. Thus, even if they get caught later, they can use these negative comments in order to defend themselves in the court.[3] [4]

Larcker and Zakolyukina (2012) also approach from this litigation risk perspective. They focus on the shareholder value, and they measure the frequency of terms which are related to the shareholder values. In other words, they limit the litigation risk to the shareholder value. In this study, we remove this limitation. The general negative tone can be related to the litigation risk (Francis et al. 1994; Rogers et al. 2011). The management can use this strategically.

The person's speaking style can vary depending on the person's personality. There has been research about the CEO characteristics (Mayew et al. 2013; Kaplan and Sorensen 2017; Green et al. 2019). Green et al. (2019) document that an extraverted personality (i.e., outgoing and energetic personality) has an association with longer CEO tenures, and less job turnover from the CEO position. Moreover, they find that among

---

[3] The accounting number itself does not change due to these negative comments from the CEO or the CFO. So, the fact whether a company meets or beats the analysis expectation, or the amount of their compensation bonus does not change, due to these negative comments. The negative tone from the management may increase the suspicion from the investors. However, it can also work as a sign of honesty of the management. Thus, the direct and immediate impact of the increased negative tone would be limited. And we argue that the management strategically can use this for their benefits.

[4] The conference call transcripts remain almost permanently. And we can identify every single comment and the individual who spoke that specific comment. Technically, the conference call transcripts can be used either favorably or unfavorably in the court. The CEO or CFO may want to leave favorable comments to themselves in order to prepare the worst case scenario.

CFOs, extraverted personality CFOs have a higher chance to be promoted to the CEO position. As a result, we assume that, in general, the personality between the CEO and the CFO is different. And because of the difference of the personality, the response during the conference call will be different when there exists accounting misreporting.

The extraverted personality is associated with being more talkative, assertive, and energetic. (John and Srivastava 1999). It is an outgoing personality. And associations between the types of personalities and the usage of linguistics are documented (Argamon et al. 2005; Oberlander and Nowson 2006; Mairesse et al. 2007). So, we assume that the unique characteristics of the CEOs is associated with the linguistic difference during the conference calls between the CEOs and the CFOs when there exist accounting misreporting. On top of this, the deception is associated with the negative tone as we discussed (Buller and Burgoon, 1996; Burgoon et al. 2016). In sum, we assume that the negative tone of the CEO is different from the negative tone of the other person (i.e., CFO) because of the different personality type, when accounting misreporting exists.

The other factors can also cause differences between the CEO and the CFO. We assume that the degree of knowledge about the company’s misreporting can vary between the CEO and the CFO when accounting misreporting exists. Thus, the personal characteristics and the degree of knowledge about the misreporting, collectively, produce the difference in the negative tone between the CEOs and the CFOs.

In addition, the CFO is more directly responsible for accounting misreporting even though the ultimate responsibility lies on the CEO. We assume that the CFO is more fearful of termination of his or her employment, or potential litigation risk against himself or herself. Empirically, Hennes et al. (2008) document that the CFO is replaced more likely than the CEO if there is a restatement. So, I assume that the CFO would try to be more conservative during the earnings conference call when negative private information, accounting misreporting, exists. Especially, this trait will be more observable during the Q&A session, which are carried out with improvisation.[5] All the elements (e.g., personal characteristics, degree of knowledge about the accounting, potential losses when it is revealed, etc.) indicate that the response of the CFOs would be more sensitive than the one of the CEOs.

H1: The negative tone change differences between the CEO and the CFO during the conference call are positively associated with accounting misreporting.

As we discussed in the second chapter, the raw level of the management self-efficacy is negatively associated with the company's future performance. And if we rely on the control theory of the deceptions, speakers make more aggressive speech to hide the deceptions. Similar mechanisms can be applicable to the accounting deceptive cases. In this case, the CEO could be more sensitive in the self-efficacy tone. This is because CEOs generally have more high levels of self-efficacy or overconfident personalities.

[5] Larcker and Zakolyukina (2012) find that "there is also a positive association between deception and negation words for CFOs, but not for CEOs" (pp. 522). However, they did neither focus on the difference between the job titles. They build two separate models by using the CEO sample and the CFO sample, separately. Furthermore, they obtain mixed results, when adding an individual fixed effect. Thus, this topic still remains as an empirical question.

Thus, we can assume that the CEOs can intentionally cover accounting misreporting by increasing the tone of management self-efficacy.[6]

H2: The CEO will increase the level of the management self-efficacy tone more than the CFO when accounting misreporting exists.

3.4. Sample description and research design

4.1. Sample data

We used the same dataset that are used in the second Chapter. We obtain the earnings call transcripts from SeekingAlpha.com. We collect accounting misreporting from AuditAnalytics. AuditAnalytics provides the restatement data with the reason of its restatement (i.e., fraud, accounting rule application failures, accounting and clerical applications errors, and other reasons). We use the three categories of the restatement cases for accounting misreporting cases (i.e., fraud, accounting rule application failures, accounting and clerical applications errors). [7] We use the financial data from Compustat.

---

[6] The prior research shows that lying politicians intentionally produce longer words in order to hide their deceptions and increase the specificity (Braun et al. 2015). Usually, the lack of specificity is regarded as untruthful. However, some politicians can utilize this common belief in order to conceal their fabrication.

[7] In AuditAnalytics dataset, other reasons include material weakness internal control over financial reporting, audit or auditor related restatement, etc. AuditAnalytics responded to us that they classified the restatements as the error category when a company explicitly announced that the restatement was due to errors when they announced the restatement. We include this error type restatement as our restatement dataset because 1) this classification judgment is solely based on the company's own disclosure and it is not decided by the independent entity, 2) AuditAnalytics database indicates that many of these error cases are also associated with the SEC investigations or securities class action litigation legal cases, and 3) anyway, it ultimately caused the explicit restatement of the financial statements on which the information users rely.

Some restatements span more than two consecutive years. In this case, we only include the first-year observation in our dataset and excludes the rest of the restatement periods. We did this because we focus on the narrative tone changes from the none-misreporting year to the first misreporting year. Sometimes, one fiscal year of a company has multiple restatements afterwards. Although the fiscal year is the same, the reasons and the disclosure dates for the restatements are different. In this case, we include this as two separate observations of restatement. This is because the restatement observation is still rare and this makes our dataset unbalanced between two events. As a result, we include as many restatement observations as possible. [8] And other years, which are not covered by AuditAnalytics restatements dataset, are considered years without accounting misreporting. Table 18 and Table 19 show the sample composition and distribution.

Table 18 Sample Description

Source / Filter	**Observations**	**Remaining Observations**
All conference call transcripts from SeekingAlpha.com	149,164	149,164
Less transcripts that are not year end	(116,385)	32,779
Less transcripts uploaded more than 30 days after call	(335)	32,444
Less transcripts with a duplicate Ticker Symbol	(3,230)	29,214
Less transcripts not matched with Compustat	(2,682)	26,532
Less transcripts that the CEO or the CFO is changed , or either the CEO or the CFO did not participate in the Q&A session	(16,537)	9,995
Added restatement occurs multiple time in one year	34	10,029
Less restatements span more than one year (Removing the restatement observations except the first year)	(441)	9,558
Less transcripts not matched audit fees and capital market variables	(2,592)	6,966
Final sample	6,966	

[8] We winsorized the dataset before every analysis at 1% and 99% level. The main results of our hypotheses remain robust when we include only one data point although there were multiple restatements in the same year (i.e., the main results unchanged when we exclude 34 additional observations.)

Table 19 Distribution of year in the earnings conference call transcripts

Year	Total Observations	Restatement Cases	Ratio
2007	76	2	2.6%
2008	100	1	1.0%
2009	487	22	4.5%
2010	379	15	4.0%
2011	283	20	7.1%
2012	375	36	9.6%
2013	450	39	8.7%
2014	978	73	7.5%
2015	1,062	55	5.2%
2016	1,112	45	4.0%
2017	881	31	3.5%
2018	783	22	2.8%
Total	6,966	361	5.2%

One of the key challenges for preparing our sample is matching personal names. There are many variations in the English names (i.e., “Robert” with Robert, Rob, and Bobby, or “Anthony” with Anthony or Tony). And there are frequent typos in the complex names. For the Asian names, family name and given name of the same person are sometimes switched over the years. All of these create hurdles for name matching. And we manually matched these variations and standardized the name variations.[9] [10]

4.2. Research design

General narrative tone of each person varies depending on a number of factors. Each person's personality can have an effect on its tone (Argamon et al. 2005; Oberlander and Nowson 2006; Mairesse et al. 2007; Burgoon et al. 2016). At the same time, financial outcomes of the company would also cause these tone variations because the main topic of the earnings conference calls would be financial results of their companies. The importance of controlling for individual personality factors is also empathized in the prior literature (Larcker and Zakolyukina 2012; Burgoon et al. 2016). [11] In order to control for unobservable factors of each individual, we employ a tone-change variable across years within the same person. This will enable us to eliminate those unobservable individual factors (Burgoon et al. 2016).

As discussed, financial outcomes of the company can also cause narrative tone variations (Loughran and McDonald 2011). We can assume that the direction of the tone changes is related to the company performance (Loughran and McDonald 2011). And the

---

[9] By comparing the names and the job titles, we can identify the same person across years. We replace various names with one standardized name. By doing this, we can match the CEOs and CFOs across multiple years in the same company.

[10] Larcker and Zakolyukina (2012) also encounter this issue when they analyze the conference call transcripts. They also document that "*we manually correct these inconsistencies*" with their more than 17,000 earnings conference call transcripts.

[11] "…they (auditors and analysts) would look for variation in vocalics and linguistics for the same individual over time for the purpose of identifying fraudulent topics. Moreover, a within-firm analysis removes the potential for unobserved heterogeneity across executives with respect to linguistic speaking style…" (Burgoon et al. 2016 pp.132)

changes would move in the same direction across job titles (i.e., if the financial result of the current year is positively outstanding, this rosy financial result leads to higher positive tone of both the CEO and the CFO in general. If the financial result is not desirable, vice versa.). In order to control for a year-company specific factors, we employ the difference between the CEO and the CFO. By taking the difference, we can eliminate the company-specific effects over the tone. As a result, this approach enables us to mitigate observable and unobservable company-specific effects. We construct a tone change variable between the CEO and the CEO as following:

$$TONE\_DIFF\_QnA\_CEO\_CFO_{i,t} = (TONE\_QnA\_CEO_{i,t} - TONE\_QnA\_CEO_{i,\, t-1}) - (TONE\_QnA\_CFO_{i,t} - TONE\_QnA\_CFO_{i,t-1})$$

where $TONE\_QnA\_CEO_{i,t}$ = the measured tone of CEO during the Q&A session of company i, at year t

$TONE\_QnA\_CEO_{i,t-1}$ = the measured tone of CEO during the Q&A session of company i, at year t-1

$TONE\_QnA\_CFO_{i,t}$ = the measured tone of CFO during the Q&A session of company i, at year t

$TONE\_QnA\_CFO_{i,t-1}$ = the measured tone of CFO during the Q&A session of company i, at year t-1

We assume that the CFO is more sensitive than the CEO with respect to accounting misreporting. Thus, the CFO will significantly increase the negative tone during the current year compared to the CEO. And if so, the value of our construct, $TONE\_DIFF\_QnA\_CEO\_CFO_{i,t}$, will be significantly negative when accounting misreporting occurs.

If there is no accounting misreporting, the value of $TONE\_DIFF\_QnA\_CEO\_CFO_{i,t}$ would be zero or around zero. The reason is that all the personality related tones would be canceled out by using the difference between the same person. In addition, all the company related tone changes would be also removed by deploying the two persons' tones from the same company. Thus, our measure of interest would be zero when there is no accounting misreporting. We investigate the relationship between the management's tone change and accounting misreporting by using the following equation.

$$
\begin{aligned}
RESTATEMENT_{i,t} = \beta_0 &+ \beta_1\, TONE\_DIFF\_QnA\_CEO\_CFO_{i,t} \\
&+ \beta_2\, Big4_{i,t} + \beta_3\, LOG\_TOTAL\_ASSETS_{i,t} \\
&+ \beta_4\, ROA_{i,t} + \beta_5\, DEBT\_RATIO_{i,t} \\
&+ \beta_6\, CURRENT\_RATIO_{i,t} \\
&+ \beta_7\, T\_ACCRUALS_{i,t} + \beta_8\, BTM_{i,t} \\
&+ \beta_9\, ABNORMAL\_AUDIT\_FEE_{i,t} \\
&+ \text{year fixed effects} + \text{industry fixed effects} + \varepsilon_{i,t}
\end{aligned}
$$

If there is a financial restatement, RESTATEMENT would be 1, otherwise 0. We assume that if there is accounting misreporting, the value of TONE_DIFF_QnA_CEO_CFO would be negative. So, the main coefficient of interest, β_1, is expected to have a negative value.

Prior textual analysis literature does not use financial control variables. They directly analyze the relationship between the linguistics cues and accounting misreporting (Larcker and Zakolyukina 2012; Burgoon et al. 2016). Price et al. (2011) evaluate predictive power of the financial models in order to identify accounting misreporting. They compare the six financial models (Dechow et al. 1995; Sloan, 1996; Beneish, 1999; Dechow and Dichev 2002; Dechow et al. 2011; Hribar et al. 2014). There is no financial or market variable that is used in all these six models commonly. In other words, there has been no financial variable that has a consistent relationship with accounting misreporting.[12]

Price et al. (2011) document that the abnormal audit fee model (Hribar et al. 2014) produces the high prediction power among academic models. So, in this study, we include the abnormal audit fee variable in order to identify the incremental ability of our model. Besides, we include the basic financial and market variables which are used in evaluating the company's financial statement as control variables.[13] All variables are defined in Appendix F.

3.5. Empirical results

Table 20 shows the summary of descriptive statistics. In our dataset, the restatement cases are about 5% among the entire firm-year dataset.[14] The average change of the tone

[12] Price et al. (2011) did not use other control variables except for the misreporting indicator variable.
[13] We also conduct the analysis without control variables like the prior literature (Larcker and Zakolyukina 2012; Burgoon et al, 2016).
[14] Prior research has 2-14% of restatement cases among the total firm-quarters dataset (Larcker and Zakolyukina, 2012).

during the conference calls converged into zero. We construct the tone changes with respect to the job titles (i.e., CEO and CFO) and the conference call session (i.e., the prepared part and the Q&A part), respectively. All the medians of tone changes are zero or near zero.

Table 20 Summary statistics

Variables	N	Mean	St.Dev	p1	p25	Median	p75	p99
RESTATEMENT	15,673	0.048	0.214	0	0	0	0	1
NEGATIVE_CHANGE_CEO _PT	13,533	0.000	0.068	-0.185	-0.040	0.000	0.039	0.195
NEGATIVE_CHANGE_CFO _PT	12,566	0.001	0.080	-0.220	-0.046	0.000	0.046	0.223
NEGATIVE _CHANGE_CEO_QNA	13,112	-0.001	0.067	-0.188	-0.041	0.000	0.038	0.185
NEGATIVE _CHANGE_CFO_QNA	11,323	-0.002	0.105	-0.323	-0.059	0.000	0.056	0.328
NEGATIVE_CHANGE_ CEO_CFO_PT	10,605	0.000	0.093	-0.252	-0.056	0.001	0.057	0.244
NEGATIVE_CHANGE_ CEO_CFO_ QNA	9,588	0.001	0.120	-0.341	-0.068	0.000	0.070	0.348
SELF_EFFICACY_ CHANGE_CEO_CFO_PT	10,605	0.000	0.134	-0.338	-0.088	0.000	0.088	0.333
SELF_EFFICACY_ CHANGE _CEO_CFO_QNA	9,588	-0.002	0.183	-0.495	-0.113	-0.001	0.109	0.480
BIG4	15,673	0.835	0.371	0	1	1	1	1
LOG_TA	15,641	7.46	2.032	2.636	6.062	7.517	8.816	12.554
ROA	15,641	-0.011	0.187	-1.012	-0.008	0.028	0.068	0.272
DEBT RATIO	15,610	0.566	0.256	0.073	0.387	0.561	0.731	1.381
CURRENT RATIO	12,802	2.613	2.325	0.396	1.251	1.908	3.025	14.52
T_ACCRUALS	15,641	-0.065	0.099	-0.538	-0.091	-0.048	-0.016	0.203
ABNORMAL AUDIT FEE	14,589	0.08	0.525	-1.232	-0.269	0.079	0.431	1.39
BTM	14,045	0.579	0.562	-0.617	0.244	0.455	0.761	3.316

All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix F.

Table 21 illustrates the correlation matrix. Table 21 shows a moderate level correlation between the CEOs and the CFOs during the presentation part (correlation 0.220). Interestingly, there are reduced correlations between the presentation part and the

Q&A part (correlation 0.121 with the CEOs, 0.115 with the CFOs, respectively). Based on these around 0.1 correlations, it may suggest that contents or tones during the earnings conference calls are different between the presentation part and the Q&A part.

Table 22 illustrates the results of the main hypothesis. As we expected, the change of the negative tone during the Q&A session has a statistically significant association with the incident of the restatement of their financial statements. The results are robust with various model specifications (Column [1], [2], [3], and [4]). The sign of the coefficient is also negative as we expected. This can be explained by our assumption that the CFOs are more sensitive to accounting misreporting than the CEOs. The CFOs are more concerned of future litigations—revealing more negative emotions (i.e., emotion theory from Larcker and Zakolyukina [2012]) —or less skillful at hiding the negative emotion (i.e., control theory from Larcker and Zakolyukina [2012]).

In order to investigate the main result further, we test the individual negative tone changes. Column (6) and (7) in Table 22 show the individual tone analysis based on each job title. Interestingly, the negative tone of the CEOs is not positively but negatively associated with accounting misreporting. The CEOs may intentionally reduce the negative tone during the conference calls (control theory). On the contrary, the negative tone of the CFOs is positively associated with accounting misreporting. The CFOs may implicitly express guilt or negative emotions during the conference calls (emotion theory). Or, the CFOs are more concerns of future litigation risk. So, the CFOs may increase the negative tone in order to defend themselves. This finding is consistent with the prior literature. Larcker and Zakolyukina (2012) document that they find the positive relation between the accounting irregularities and the negative tone for CFOs, but there is not for CEOs. However, our findings are not statistically significant.[15] If we only focus on the individual tone changes, we cannot find statistically significant associations. However, if we investigate the tone differences between the CEO and the CFO, we can identify accounting misreporting. Column (1)-(7) of Table 22 supports this argument.

[15] In our analysis, we use only the Q&A part. Larcker and Zakolyukina (2012) use the pooled dataset with both the presentation part and the Q&A part. This may cause the difference in the statistical significant level.

Table 21 Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) RESTATEMENT	1																
(2) NEGATIVE _ CHANGE_CEO_PT	0.02**	1															
(3) NEGATIVE _ CHANGE_CFO_PT	0.01	0.22***	1														
(4) NEGATIVE _ CHANGE_CEO_QNA	-0.01	0.12***	0.09***	1													
(5) NEGATIVE _ CHANGE_CFO_QNA	0.01	0.08***	0.11***	0.07***	1												
(6) NEGATIVE _ CHANGE_CEO_CFO_PT	0.00	0.54***	-0.69***	0.01	-0.05***	1											
(7) NEGATIVE _ CHANGE_CEO_CFO_QNA	-0.02*	0.00	-0.05***	0.49***	-0.83***	0.05***	1										
(8) SELF_EFFICACY _ CHANGE_CEO_CFO_PT	0.01	-0.07***	0.06***	0.00	0.00	-0.10***	0.00	1									
(9) SELF_EFFICACY _ CHANGE_CEO_CFO_QNA	0.00	0.00	0.02*	0.00	0.02**	-0.02	-0.02**	0.03***	1								
(10) ABNORMAL_AUDIT_FEE	0.04***	0.00	-0.01	0.00	-0.01	0.01	0.01	0.00	-0.01	1							
(11) BIG4	0.01	-0.01	0.01	-0.01	0.01	-0.02**	-0.01	0.01	-0.02*	-0.09***	1						
(12) LOG_TA	0.02*	0.00	0.01	-0.01	-0.02**	-0.01	0.01	0.01	-0.02*	0.11***	0.41***	1					
(13) ROA	0.00	0.00	-0.03***	0.00	0.00	0.03***	0.01	0.01	-0.03***	-0.05***	0.18***	0.40***	1				
(14) DEBT_RATIO	0.03***	0.00	-0.01	-0.01	-0.01	0.00	0.01	0.01	0.00	0.17***	0.09***	0.35***	-0.09***	1			
(15) CURRENT_RATIO	-0.03***	-0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.02*	-0.17***	-0.10***	-0.34***	-0.14***	-0.52***	1		
(16) T_ACCRUALS	0.00	-0.02**	-0.07***	0.01	-0.02*	0.05***	0.02	0.02	-0.03**	-0.03***	0.05***	0.24***	0.54***	-0.04***	0.06***	1	
(17) BTM	0.02*	0.04***	0.03***	0.01	0.02**	-0.01	-0.00	-0.00	-0.00	-0.04***	-0.04***	0.12***	0.01	-0.15***	-0.01	0.03***	1

*, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix F.

Next, we test the second hypothesis. Column (8) in Table 22 shows the result. There is no relationship between the management self-efficacy and the financial misreporting. This result shows that the level of management self-efficacy is not an effective indicator in order to identify accounting misreporting. The reason for this would be that the management self-efficacy is more effective in measuring the positive aspects of the company. Or the sensitivity about management self-efficacy can be similar between the CEOs and the CFOs.

In sum, the negative tone would be more effective in order to identify accounting misreporting than the level of management self-efficacy. In addition, only when we use the combined measure with the CEO and the CFO together, we can uncover accounting misreporting.

Table 22 Logistic Regression Analysis

	RESTATEMENT							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NEGATIVE_CHANGE_QNA_CEO_CFO	-0.6419*	-0.6639*	-0.9524**	-1.0214**	-1.0542**			
	(-1.6932)	(-1.6817)	(-2.0976)	(-2.2653)	(-2.2638)			
NEGATIVE_QNA_CHANGE_CEO						-0.7343		
						(-0.9738)		
NEGATIVE_QNA_CHANGE_CFO							0.4404	
							(0.8318)	
SELF_EFFICACY_CHANGE_QNA_CEO_CFO								-0.1898
								(-0.5931)
ABNORMAL AUDIT FEE				0.2880**	0.3552**	0.3410***	0.3981***	0.3515**
				(2.213)	(2.5753)	(2.7586)	(3.0655)	(2.5438)
BIG4			0.0688	0.2193	0.1817	-0.0681	0.2437	0.1838
			(0.3487)	(1.1477)	(0.903)	(-0.4063)	(1.2771)	(0.9157)
LOG_TA			-0.0079	-0.0101	-0.0321	-0.0016	-0.0353	-0.0332
			(-0.1682)	(-0.2395)	(-0.6574)	(-0.0386)	(-0.7995)	(-0.6792)
ROA			0.2302	0.0423	0.1257	-0.0831	0.083	0.1301
			(0.4419)	(0.0862)	(0.2376)	(-0.2303)	(0.1734)	(0.2455)
DEBT_RATIO			0.1445	0.3698	0.1318	0.0291	0.243	0.1232
			(0.3887)	(1.0358)	(0.3466)	(0.095)	(0.6925)	(0.325)
CURRENT_RATIO			-0.1154**	-0.0616	-0.0959*	-0.0980**	-0.0832*	-0.0963*
			(-2.4120)	(-1.3937)	(-1.9278)	(-2.3592)	(-1.8227)	(-1.9297)
TOTAL_ACCRUAL			0.1449	1.1482	0.5845	0.5847	0.0825	0.522
			(0.1627)	(1.3415)	(0.6385)	(0.8335)	(0.0967)	(0.5721)
BTM			0.4045***	0.4345***	0.4565***	0.2485**	0.4725***	0.4561***
			(3.1334)	(3.9005)	(3.3066)	(2.156)	(3.8799)	(3.297)
Obs.	9,588	9,283	7,060	6,966	6,651	9,261	7,836	6,651
Pseudo R^2	0.0007	0.0537	0.0664	0.0138	0.0709	0. 0512	0.0693	0.0693
INDUSTRY FE	NO	YES	YES	No	YES	YES	YES	YES
YEAR FE	NO	YES	YES	No	YES	YES	YES	YES

Standard errors are clustered by clien. z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix F.

3.6. Additional analysis

In this section, we add some additional tones in our model. In addition, we will introduce an up-to-date textual analysis methodology, the Doc2Vec, into our model. And we will evaluate our models with the existing model.

The main hypothesis in this chapter was analyzing the tones in the Q&A sessions during the earnings conference calls. And we focused on the differences between the CEOs and the CFOs. However, for some companies, the CFO does not appear in the earnings conference calls. Furthermore, either the CEO or the CFO can be changed during a year. If this is the case, we cannot use our previous model, which requires the participations of the unchanged CEO and the unchanged CFO simultaneously during the Q&A session. In this session, we extend our analysis, focusing on the prepared presentation session of the earnings conference call.

Lee (2016) shows that the style of the language of the presentation session is different from the style of the Q&A session, even if they are from the same company. He shows it by building his own construct to identify the language style.[15] Lee (2016)'s finding suggests that the presentation session bears the different aspects from the one of the Q&A session. The presentation session can also be used in order to identify accounting misreporting.

Table 23 shows the results with respect to the negative tone changes during the presentation session. The negative tone change of the CEO during the presentation sessions over the years has a positive association with accounting misreporting (Column

[15] His construct is based on the prior literature findings (Rowley-Jolivet and Carter-Thomas 2005)

[2] Table 23).[16] This finding is consistent with the prior literature. Larcker and Zakolyukina (2012) do not find the association between the deception and the negative tone from the CEO's comments with their initial plain analysis. However, when they adjust the level of the negative tone by using individual person's own average level of tones which are obtained from the previous quarters, they find a positive association between the deception and the negative tone during the conference call.[17] This implies that the CEOs increase the negative tone during the prepared presentation remarks, when they have accounting misreporting.

In the hypothesis development discussion, we assume that the CFO would be more sensitive to accounting misreporting. However, during the presentation session, CFO's remarks have a marginal significant relationship with accounting misreporting (Column [3] Table 23). [18] This can be explained by the assumption that the CFO may try to avoid the controversial topic itself during the presentation session. [19] Or, the general discussion may be carried out by the CEO, including the overall financial outcomes. If this is the case, the CFO can stand aside from the problematic topic during the presentation session.

In short, the negative tone changes have an association with accounting misreporting in both CEOs and CFOs, although CFO's statistical significance level is marginal (i.e., around 12.8% p-value). These findings can be supported by our discussion

---

[16] Z-statistics is 2.0291. Corresponding p-value is 4.2%.
[17] Larcker and Zakolyukina (2012) did not distinguish the presentation part and the Q&A part. In our research, we find this positive relationship in the presentation part.
[18] Z-statistics is 1.5231. Corresponding p-value is 12.8%.
[19] The company's financial outcome can be explained by using various accounts (i.e., various line items in the financial statements) or events. Thus, the CFOs may try to explain the company's financial results without mentioning the problematic account(s) during the presentation session.

in the hypothesis development section, especially the emotion and the litigation risk argument.

Table 23 Logistic regression results in predicting the restatement

	RESTATEMENT			
Variables	(1)	(2)	(3)	(4)
NEGATIVE_CHANGE _CEO_CFO_PT	-0.4284			
	(-0.6873)			
NEGATIVE_CHANGE _CEO_PT		1.4362**		
		(2.0075)		
NEGATIVE_CHANGE _CFO_PT			1.0611	0.9998*
			(1.5231)	(1.6817)
ABNORMAL AUDIT FEE	0.3692***	0.2922**	0.3996***	
	(2.5841)	(2.3293)	(3.0043)	
BIG4	0.1055	-0.0604	0.1113	
	(0.5724)	(-0.3646)	(0.6309)	
LOG_TA	-0.0176	-0.004	-0.0174	
	(-0.3776)	(-0.0965)	(-0.4105)	
ROA	0.5313	-0.1323	0.4961	
	(1.2119)	(-0.3812)	(1.2347)	
DEBT_RATIO	0.0943	0.0726	0.2448	
	(0.2582)	(0.2370)	(0.7385)	
CURRENT_RATIO	-0.0734*	-0.0845**	-0.0508	
	(-1.6462)	(-2.0550)	(-1.2784)	
T_ACCRUALS	0.413	0.6812	0.1214	
	(0.5092)	(1.0105)	(0.1598)	
BTM	0.4690***	0.2753**	0.4792***	
	(3.9470)	(2.5246)	(4.3967)	
Obs.	7,453	9,566	8,850	12,171
Pseudo R^2	0.0686	0.0486	0.0685	0.0511
INDUSTRY FE	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES

Standard errors are clustered by client. z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix F.

3.6.1. Doc2Vec Methodology

In Chapter 2, we applied the Word2Vec methodology in order to improve the management self-efficacy measurement. In this section, we will test another up-to-date natural language processing model. We will adapt the Doc2Vec methodology and compare the prediction performances with respect to identifying accounting misreporting. Before we jump into the result, we will briefly discuss about the concept of the Doc2Vec methodology.

As we discussed in Chapter 2, the Word2Vec methodology converts a word into a vector, preserving the meaning of the word. As a result, by using these vectors, we were able to select a group of words that share the similar meanings. It is based on the assumption of the distributional hypothesis. Words with similar meaning are located in the vector space near to each other (Turney and Pantel 2010). And as we tested this in Chapter 2, the Word2Vec method worked effectively. Le and Mikolov (2014) extend this Word2Vec methodology from a word level into a document level. They propose a model that converts a document into a vector (so called Doc2Vec methodology).

The structure of Doc2Vec methodology is quite similar to the Word2Vec methodology. The Word2Vec model predicts the target word by using surrounding words (context words) with an Artificial Neural Network (ANN). On top of the Word2Vec model, the Doc2Vec model has additional input nodes that represents each document.[20]

[20] In other words, we have N input nodes when we have N unique words in the Word2Vec model. Similarly, in the Doc2Vec model, we have N input nodes and M additional input nodes, when we have N unique words and M unique documents. The number of output nodes is the same to the Word2Vec

By using this ANN structure, we can train our ANN parameters (weights) in order to produce the right output results. Input data is the binary value of document index and surrounding words (i.e., context words). And the output data is the target word.[21] After training process, each document node will have multiple connected weights in the ANN

model. In the Word2Vec model, we have N output nodes when we have N unique words in our dataset. In the Doc2Vec model, we also have N output nodes when we have N unique words.

[21] In the Word2Vec model, surrounding words and a target word slide word by word in a sentence. In the Doc2Vec model, surrounding words and a target words move in the same way. In addition, we use each document index as an additional input. As long as the words come from the same document, the same document binary index is repetitively used as an additional input during the training process.

model. Those multiple weights constitute a vector, corresponding that specific document. The brief structure of the Doc2Vec model is illustrated in Figure 3.[22] [23] [24] [25]

By using the Doc2Vec model, we can convert a document into a vector. We can choose the scope of the document, depending on our purposes. If we choose a single conference call as one document, we can convert a single conference call transcript into a vector. Or, if we define CEO's comments from the Q&A session of a single earnings conference call transcript as one document, we can convert this collection of the CEO's

---

[22] For more detail information about the Doc2Vec model, please see Le and Mikolov (2014).
[23] For the implementation, the multiple number of surrounding words are used for training process. So, word vectors can be averaged or concatenated. If they are averaged, the number of input nodes would be the number of unique words. If the input vectors are concatenated, the number of input nodes are the number of unique words multiplied by the number of surrounding words (i.e., the number of surrounding words is called window size).
[24] For more formal definition, it can be express like these: the Doc2Vec model maximizes below average log probability.

$$\sum_{d=1}^{ND} \quad \frac{1}{T}\sum_{t=k+1}^{T-k} \log p\,(w_t|w_{t-k},\dots,w_{t-1},w_{t+1},\dots,w_{t+k},d_d)$$

where

$$\log p\,(w_t|w_{t-k},\dots,w_{t-1},w_{t+1},\dots,w_{t+k},d_d) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}}$$

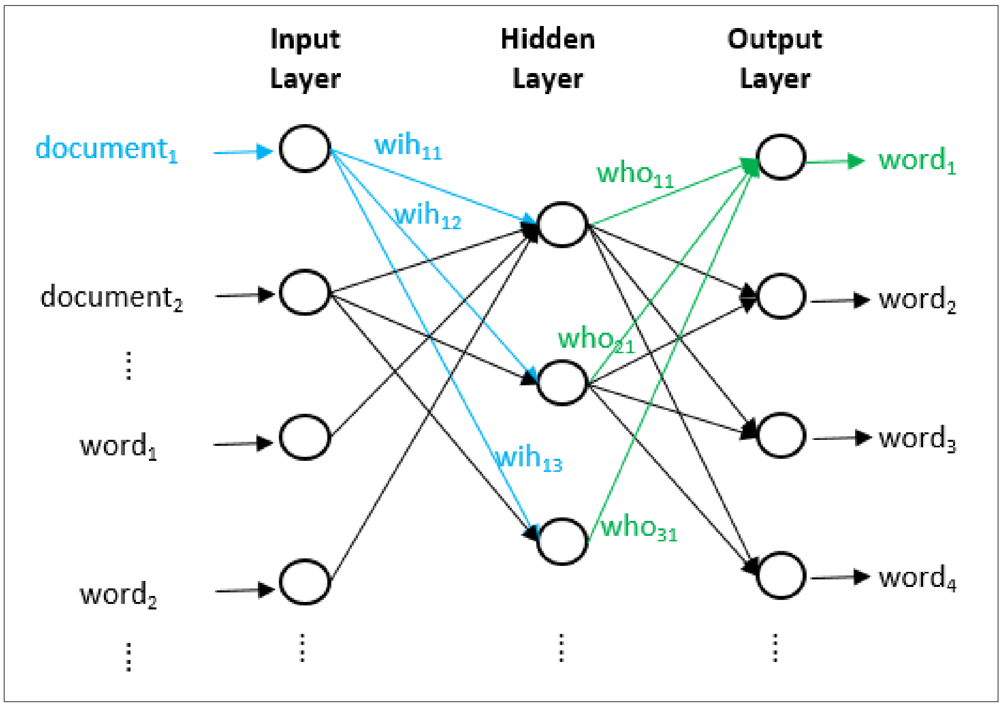
$$y = b + Uh\,(w_{t-k},\dots,w_{t-1},w_{t+1},\dots,w_{t+k},d_d;W,D)$$

ND is the number of documents. T is the number of words in the document. *k* is the number of surrounding words (i.e., window size). *p* is the output probability of the word given the surrounding words. y_i is the log probability for each output word i, given the 2k surrounding words. *h* is produced by concatenation or average of surrounding words with the look-up word matrices W and D. U and b are the softmax parameters. If we use Figure 3 for the explanation, W and D would be the weights between the input layer and the hidden layer (e.g., wih_{11}, wih_{12}, etc.). U and b would be the weights between the hidden layer and the output layer (e.g., who_{11}, who_{12}, etc.). In the hidden layer, there is no non-linear activation function. The above equations are slightly modified from the prior literature from Le and Mikolov (2014). There are two types of Doc2Vec models. The above equations are specifically related to the distributed memory model of the Dov2Vec model. The other model, the Distributed Bag of Words version of Paragraph Vector (PV-DBOW) of the Doc2Vec model, is maximizing the probability of surrounding words with given one document vector.
[25] The document input nodes and the words nodes can be treated separately unlike Figure 3. In other words, one input node for a document can have M weights. At the same time, one input node for a word can have N weights. M is not necessarily the same value of N. (In Figure 3, all input nodes are connected to all hidden nodes. So, the number of weights on each input node are the same.)

comments into a vector. Like a word vector in the Word2Vec model, each dimension in a document vector in the Doc2Vec model is assumed that it has a certain semantic measurement (i.e., positive tone, negative tone, closeness to a certain topic, etc.).

Figure 3

Illustration of the structure of the Doc2Vec model based on the artificial neural networks



The Doc2Vec is used for classifying document by sentiments, or by contents topics (Rehurek and Sojka 2010; Le and Mikolov 2014; Kim et al. 2019; Gensim 2020). A converted document vector has usually more than hundred dimensions (e.g., 400 dimensions). The classification task or prediction task is performed by using a logistic regression or artificial neural network, feeding these multi-dimensional vectors into the model (Le and Mikolov 2014).[26] [27]

In our task, we assume that one of the document dimensions can be related to the negative tone. If so, that element in the document vector will work as an effective discriminator for accounting misreporting. Furthermore, if the vector operation is still effective to the document vector, we can obtain the negative tone changes between the CEO and the CFOs across two consecutive years.[28] [29] As long as we know, the vector operation was only used at a word level (i.e., the Word2Vec model). In this research, we will extend this at a document level. We will investigate this new possibility of the Doc2Vec application.

3.6.2. Five models

In addition to the Doc2Vec model, we will add more models. There have been multiple models proposed in order to predict accounting misreporting (Dechow et al. 1995; Sloan 1996; Beneish 1999; Dechow and Dichev 2002; Dechow et al. 2011; Price et

---

[26] Building a classification model using numerical variables can be achieved by multiple ways (e.g., random forest, K-mean clustering, support vector machine, etc.).

[27] Recently, any objectives can be converted into a vector. For example, we can convert a word into a vector by using the Word2Vec. In this section, we convert a document into a vector with the Doc2Vec approach. Shin et al. (2019) convert one image into a vector for their prediction model. By using visual classification model with ANN, they obtain 4,096 dimensions from each image. Then, they use this vector in predicting the type of user's response in the social media.

[28] In the Word2Vec model, the vector operation preserves the semantic meaning. For example, all the following three vector operations (i.e., minus operation) have similar vector values: 1) vector of the word *man* – vector of the *woman*, 2) vector of the word *uncle* – vector of the *aunt*, and 3) vector of the word *king* – vector of the *queen*. (Mikolov et al. 2013a)

[29] Regarding the previous footnote, there are other examples: German with Berlin, Spain with Madrid, and Italy with Rome. All the pairs have similar relative locations to each other when we convert each word into a vector (Mikolov et al. 2013b) In other words, the differences between those two vectors produce one similar vector.

al. 2011; Hribar et al. 2014). Price et al. (2011) compare existing proposed models. Price et al. (2011) use the seven models and show that the commercial model outperforms the six academic models. Price et al. (2011) also documented that the abnormal audit fee model (Hribar et al. 2014) is relatively effective among the six academic models.[30] In this study, we include the abnormal audit fee model as a baseline model for the comparison purpose.

Hribar et al. (2014) propose the abnormal audit fee model for detecting accounting misreporting. Price et al. (2011) slightly modify the model by omitting four variables in order to increase the available sample size. [31] In this research, we use the same abnormal audit fee model of Price et al. (2011)'s. The abnormal audit fees are calculated by using the residual of the following regression. [32] The regression result of this audit fee model is documented in Appendix G.

[30] Price et al. (2011) document that the commercial model outperforms the five academic model with less than 1% p-value in their comparison. Only the abnormal model has more than 5% of p-value (i.e., 8%). In other words, only the abnormal audit fee model has a similar level of the estimated coefficient (i.e., 0.130) to the coefficient of the commercial model (i.e., 0.200). Price et al. (2011) do not perform the direct comparison among the academic models. They only compared the commercial model with each one of the academic models. However, based on their results, we can infer that the result of the abnormal audit fee model is the most effective among academic models, because the abnormal audit fee model's performance is closest to the commercial model. So, we use the abnormal model as a base line model in this research. Price et al. (2011) used Hribar et al.'s model of their working paper in 2010.

[31] Price et al. (2011) document that omitting four variables does not reduce the explanatory power significantly.

[32] LogFees is a natural logarithm of total assets at the fiscal year end. IndustryIndicators are dummy variables of the first two-digit SIC code. BigN is a dummy variable of Big four auditors. Assets is the amounts of total assets at the fiscal year end. Inventory is the amounts of the inventory at the end of the fiscal year. AvgAssets is the average of the total assets at the fiscal year start and end. Receivables is the amounts of the receivables. LTD is the amounts of the long term debt. Earn is the amounts of the operating income. Loss is an indicator variable of loss in the operating income. Qualified is an indicator variable of the modified opinion, including the going concern opinion. AuditorTenure is the number of years with the same auditor based on Audit Analytics dataset.

$$\begin{aligned}\text{LogAuditFees} = {} & \text{IndustryIndicators} + \beta_1 \text{ BigN} + \beta_2 \log(\text{Assets}) \\ & + \beta_3 \text{ Inventory / AvgAssets} + \beta_4 \text{ Receivables / AvgAssets} \\ & + \beta_5 \text{ LTD / AvgAssets} + \beta_6 \text{ Earn / AvgAssets} \\ & + \beta_7 \text{ Loss} + \beta_8 \text{ Qualified} + \beta_9 \text{ AuditorTenure} + \varepsilon\end{aligned}$$

In the previous section, we show that the tone changes between the CEOs and CFOs can be used effectively in order to identify the restatement. In addition to this, we will apply the Doc2Vec method for additional models. So, we will have six models.

Our first model will use the tone changes between the CEOs and CFOs. For the control variables, we use the same control variables which are used in the previous section except the abnormal audit fee variable. [33] The specific model like following:

$$\text{RESTATEMENT} = \beta_0 + \beta_1 \text{ TONE\_DIFF\_CEO\_CFO\_QnA} + \text{controls} + \varepsilon_{i,t}$$

The second model will use the Doc2Vec model at the document level. We convert each document into a vector with 100 dimensions. So, we have 100 variables in each document, which is the conference call transcript. The detailed model are like following:

$$\text{RESTATEMENT} = \beta_0 + \beta_1 \text{ DOC2VEC\_DIM1} + \beta_2 \text{ DOC2VEC \_DIM2} + \ldots$$

[33] There is no consensus result about a variable that has a consistent association with respect to the restatement. Price et al. (2011) use six academic prediction models. There is no single common variable that exists across all six models. In this analysis, we include the variables that are fundamental and frequently used across the financial empirical research. We include control variables with Big4 auditor, logarithms of total assets, return on assets, debt ratio, current ratio, total accruals, and book-to-market ratio.

$$+ \beta_{99}\ DOC2VEC\_DIM99 + \beta_{100}\ DOC2VEC\_DIM100 + controls + \varepsilon_{i,t}$$

The third model also uses the Doc2Vec model. However, we calculate the tone difference by applying the vector operation. We convert each comment into a vector at a job title level. In other words, we can get a vector from the CEO's comments during the Q&A session. And we can obtain an additional vector from the CFO's comments during the same Q&A session. And then, we apply the vector operation like following:

$$TONE\_DIFF\_DOC2VEC_{i,t} = (DOC2VEC\_CEO\_QNA_{i,t} - DOC2VEC\_CEO\_QNA_{i,t-1}) - (DOC2VEC\_CFO\_QNA_{i,t} - DOC2VEC\_CFO\_QNA_{i,t-1})$$

All the variables in the above equation are a vector with 100 dimensions. So, TONE_DIFF_DOC2VEC will also have 100 dimensions. The third model will be like this:

$$\begin{aligned} RESTATEMENT = {} & \beta_0 + \beta_1\ TONE\_DIFF\_DOC2VEC\_DIM1 \\ & + \beta_2\ TONE\_DIFF\_DOC2VEC\_DIM2 + \ldots \\ & + \beta_{99}\ TONE\_DIFF\_DOC2VEC\_DIM99 \\ & + \beta_{100}\ TONE\_DIFF\_DOC2VEC\_DIM100 + controls + \varepsilon_{i,t} \end{aligned}$$

The fourth model is a base model by using the abnormal audit fees like followings:

$$RESTATEMENT = \beta_0 + \beta_1\ ABNORMAL\_AUDIT\_FEES + controls + \varepsilon_{i,t}$$

The fifth model contains all the variables with the negative tone changes, Doc2Vec vector at conference call level, and the abnormal audit fees. The detailed model are like following:

$$\begin{aligned} RESTATEMENT = \beta_0 &+ \beta_1\ TONE\_DIFF\_CEO\_CFO\_QnA \\ &+ \beta_2\ ABNORMAL\_AUDIT\_FEES \\ &+ \beta_3\ DOC2VEC\_DIM1 + \beta_4\ DOC2VEC\ \_DIM2 + \ldots \\ &+ \beta_{101}\ DOC2VEC\ \_DIM99 + \beta_{102}\ DOC2VEC\ \_DIM100 \\ &+ controls + \varepsilon_{i,t} \end{aligned}$$

The last sixth model will also use all the variables with the negative tone changes, Doc2Vec, and the abnormal audit fees. Unlike the fifth model, we will use the Doc2Vec model at a job title level only in the Q&A session. The detailed model are like this:

$$\begin{aligned} RESTATEMENT = \beta_0 &+ \beta_1\ TONE\_DIFF\_CEO\_CFO\_QnA \\ &+ \beta_2\ ABNORMAL\_AUDIT\_FEES \\ &+ \beta_3\ TONE\_DIFF\_DOC2VEC\_DIM1 + \ldots \\ &+ \beta_{102}\ TONE\_DIFF\_DOC2VEC\_DIM100 + controls + \varepsilon_{i,t} \end{aligned}$$

3.6.4. Model evaluation metric and comparison results

We have six prediction models and we compare these models. There are multiple ways in evaluating the model's performance. In this research, we employ the misclassification cost minimization to decide the cut-off thresholds in the logistic regression model. In order to evaluate to the performance of the models, we use k-fold cross-validation. We will briefly discuss these two concepts, before exploring the evaluation results.

The various metrics have been proposed in order to evaluate the prediction model. The prior research uses overall accuracy rate, or the area under the ROC (Receiver Operating Characteristic) curve. Price et al. (2011) use seemingly unrelated estimation (SUEST) to compare the performance between two models. [34] Some research points out that the area under the ROC curve is not effective to compare two good models (Marzban 2004; Price et al. 2011). Furthermore, if the dataset is unbalanced, the evaluation becomes difficult. When the dataset is unbalanced, the conventional methods (i.e., threshold of 0.5, or threshold minimizing the overall accuracy) classify all the observations as the majority class (Calabrese 2014).

Calabrese (2014) proposes to use the misclassification cost in order to evaluate the models. She shows that using the misclassification cost approach can be used in detecting rare disease diagnosis, or companies' bankruptcy prediction (i.e., unbalanced dataset). Our dataset is also an unbalanced dataset. Thus, we will use a misclassification cost minimization method. [35]

---

[34] Price et. al (2011) perform the pairwise comparison in order to evaluate the several models.

[35] Some of our models have multiple independent variables of interest. As a result, SUEST is not suitable for the comparison of our models. Price et al. (2011) use SUEST by using one independent variable on each model. They standardize each independent variable for the pairwise comparison purpose. All the

If we use the same dataset in building a model and in evaluating the model, it may bring biased results. [36] One of ways to overcome this limitation is a k-fold cross-validation methodology. [37] The value of K can be any natural number more than one. We use 5-fold cross-validation.[38] We divide our dataset into 5 groups randomly and use them. We will use the logistic model for the classification.

There is one more to consider when we evaluate the ANN model. The randomness intervenes when we implement the Doc2Vec model (specifically, setting the initial weights of ANN). [39] Thus, we repeat five times to reduce the impact of the randomness in our model evaluation process. By completing the above procedures, we can get 25 test datasets and the corresponding misclassification cost values. [40] By averaging these 25 values on each model, we can compare the performance of the models.

In many cases, the misclassification cost between the false positive and the false negative is not the same.[41] Usually, the misclassification cost of false negative is much

---

detection models that Price et al. (2011) use have only one independent variable of interest in each model.

[36] A model is constructed based on a certain dataset. Thus, the model can fit well on that specific same dataset. However, we cannot make sure whether the model will still fit on additional unused dataset. As a result, this evaluation with the same dataset may bear a limitation of generalizability.

[37] This method randomly divides a dataset into K groups. We hold one sub-dataset from K groups for the evaluation purpose (so called 'test dataset'). And we build a model by using the rest of the dataset (training dataset). We iterate this procedure K times by switching the test datasets. So, every observation can be served as a test dataset during the K iteration process (Alpaydin 2014).

[38] 10-fold cross-validation is also frequently used. However, we have a relatively small number of restatement observations. So, in our research, we use 5-fold cross-validation. In the logistic regression model, we need a cut-off point for the classification. We also use the training dataset for deciding the cut-off point on each training dataset. Based on this, we evaluate the test dataset.

[39] In the ANN model, all the initial weights between nodes are randomly assigned (Alpaydin 2014; Rashid 2016). Usually a value near zero is assigned. How to assign initial weights randomly is another huge research area.

[40] 25 = 5 × 5 (i.e., 5-fold cross-validation with 5 iterations)

[41] The false negative is usually defined as a case when we fail to identify a certain event. The false positive is usually defined as a case when we wrongly claim it is a certain event that is actually not that type of event. For example, in a bankruptcy prediction model, if a model fails to predict the real bankruptcy case and the model indicates that case as a non-bankruptcy case. This is a false negative case. Similarly, the

higher than the one of false positive. In our study we assume three cases. We assume 1:1, 10:1, and 100:1 scenarios.[42]

Table 24 shows the expected misclassification costs with multiple models. The first column in Table 24 shows the estimated misclassification cost when we do not differentiate the cost of the false negative and the false positive. It would produce the same results when we decide the cut-off point that maximizes the overall accuracy rate. The model with negative tone changes performs best among the six models (i.e., the lower estimated misclassification cost means the better model.). The column (2) and (3) in Table 24 show the results when we differentiate the misclassification cost ratio with 10:1 or 100:1. The negative tone model and the abnormal audit fee model still outperform the other models. Interestingly, in our experiments, the more independent variables do not necessarily improve the models. The Dov2Vec models or models with all three variables do not perform well in identifying accounting misreporting.

As we discussed in Chapter 2, the Word2Vec model was effective in identifying the semantic meaning of words. However, the Dov2Vec model is less effective in identifying accounting misreporting or measuring the level of negative tone during the conference calls. In this experiment, we produce 100 dimensions on each document (i.e., each conference call transcript). If we change the number of dimensions, one of the

---

false positive is a case that an actual none-bankrupt observation is predicted as a bankruptcy event by the model.

[42] With respect to predicting accounting misreporting, the false negative can cause huge social cost (i.e., the investment losses, and systematic chaos such as market crash, bank-run, various over-reactions, etc.). The false positive can also bring the social cost (i.e., unnecessary usage of the regulatory resources, cost for resolving investigation from the company side, investors' under-investment towards the target company, etc.). At our best knowledge, there is no empirical study about the misclassification cost in accounting misreporting cases. But we may say that the cost of the false negative cases is usually much larger than the cost of false positive cases.

dimensions would correspond to the level of the negative tone. If this works, the modified dimension may work in accounting misreporting identification. [43] [44] Or, the Doc2Vec model might not be an effective methodology to measure the level of negative tone of the earnings conference call transcripts. The application of the Doc2Vec model is a still open and on-going topic.

In this research, we investigate a new possibility of the Dov2Vec model, including its vector operation. At the same time, we evaluate various models, including the model of prior study. The audit fee model and our model of the negative tone change between CEOs and CFOs produce better performance than other models. The performances between two models are statistically insignificant or marginally significant. [45]

Table 24 Misclassification costs with multiple models

Models	Cost Ratio (FN / FP)		
	(1) 1:1	(2) 10:1	(3) 100:1
Negative Tone model	<u>72.4</u>	<u>729.4</u>	1,361.4
Doc2Vec at call level	73.0	758.1	1,486.0
Doc2Vec at personal level (job title)	72.6	802.7	1,554.4
Audit Fee model	72.7	732.4	<u>1,348.2</u>
Negative Tone + Audit_Fee + Doc2Vec (call level)	73.0	764.2	1,502.0
Negative Tone + Audit_Fee + Doc2Vec (job title level)	72.7	786.5	1,602.8

[43] In the original Dov2Vec model (Le and Mikolov 2014), they use 400 dimensions on each document. However, our untabulated result indicates that the 400-dimension model is less effective than the 100-dimension model in our setting.

[44] Kim et al. (2019) experiment the Doc2Vec model with multiple textual datasets. They document that the optimum representation dimensions vary depending on the types of documents (i.e., Tweets, newspaper, medical articles, etc.). And in some cases, a Doc2Vec model with a small dimension works better than the one with a large dimension.

[45] Based on pairwise t-tests in 25 experiments, the tone model outperforms the audit fee models statistically with respect to the expected misclassification costs (p-values with 1:1, 10:1, and 100:1 settings are 0.03, 0.48, and 0.32, respectively).

The less misclassification cost means the better performance.

3.7. Contributions and conclusion

In this chapter, we demonstrate the possibility that earnings conference calls can be used to predict accounting misreporting. This result can be beneficial for all financial statement users as well as auditors.

Previous researchers have found controlling personal characteristics challenging when analyzing earnings conference calls with large datasets (Larcker and Zakolyukina 2012; Burgoon et al. 2016). This is because unobservable variables can have impacts on the tone in earnings conference calls. To solve this problem, we employ the tone change variable over two consecutive years for the same person. We also take advantage of the job title information contained in conference call transcripts. CEOs have been documented to have a different personality type from other employees, including CFOs (Kaplan and Sorensen 2017; Green et al. 2019). We assume that these different personalities can be used in identifying accounting misreporting. This chapter shows the association about this relationship.

We document that the negative tone of CFOs increases more rapidly than that of CEOs during Q&A sessions in earnings conference calls when accounting misreporting exists. If we consider the individual tone changes of CEOs and CFOs separately, we cannot find an association with accounting misreporting; only when we use them simultaneously by measuring the differences between them is the accounting misreporting identifiable. Interestingly, when accounting misreporting exists, the negative tone expressed by CFO increases during Q&A sessions in earnings conference calls, whereas the negative tone of CEOs decreases. This likely stems from the distinct personal characteristics associated with these two job titles and the potential expected loss associated with each position.

In the additional analysis section, Findings also indicate that the tone changes in presentation sessions are associated with accounting misreporting. In addition, we compare the various models, including the existing model and the model with the up-to-date AI NLP model, Doc2Vec. Because this dissertation aims to make a prediction with unbalanced samples, we deploy the misclassification cost in our evaluations of the models (Calabrese 2011). Our results show that the performance of our tone difference model is better than—or similar to—that of the existing model (i.e., the model with abnormal audit fees, which is evaluated as the model that outperforms the other previously proposed models). Collectively, our research can be utilized by investors, auditors, or regulators when assessing the risk level of companies with respect to accounting misreporting or when interpreting earnings conference calls.

Our research has some limitations. First, we deploy the tone changes. As a result, if CEO or CFO is replaced mid-year, our method cannot be applied. In addition, we focus on tone changes when accounting misreporting occurs. As a result, our model is only valid in the first year when the accounting misreporting happens. However, some accounting misreporting spans more than a year. In this case, other models can be used to complement ours to identify the accounting misreporting.

In this research, we only use comments made by CEOs and CFOs, yet other individuals with different job titles (i.e., COO, legal counsel, etc.) participate in earnings conference calls. Including these comments would help to enrichen the earnings conference call research. Unlike the second chapter, which has an effective model with the Word2Vec model, the Doc2Vec model in this chapter does not prove its effectiveness. The Doc2Vec model has a wide variety of implementation options, and more effective hyper-parameter settings can be studied further in future research.

Chapter 4: Determinants of materiality and audit opinion

4.1. Introduction

Materiality is a major key component in controlling audit quality by implementing reasonable assurance. The materiality is used for planning, testing, evaluating, and forming an audit opinion (ISA320 and 450; PCAOB AS 2105 and 2810; Eilifsen and Messier 2015).

Due to its important role, the amount of materiality in the auditing is determined by considering numerous complex factors (ISA320, PCAOB AS 2105; Amiram et al. 2018). The Public Company Accounting Oversight Board's (PCAOB) Accounting Standard No. 2105 (PCAOB 2010) requires auditors to consider "(whether it has) altered 'total mix' of information made available" or perform "delicate assessments" when the auditor decides the materiality level by referring to the Supreme Court of the United States. Because of this complexity, the materiality level has remained the focus of research questions for several decades. Similarly, ISA 320 also defines that "the auditor considers not only the size but also the nature of uncorrected misstatements, and the particular circumstances of their occurrence." (ISA 320 para. 6.)

Extensive research has been conducted on materiality (Holstrum and Messier 1982; Messier et al. 2005). However, most of this research has relied on indirect measurements as proxies for materiality level or analysis of accounting firms' audit manuals.[1] Due to

[1] For more detailed information, please see the two literature review papers about the materiality research, including the experimental research (Holstrum and Messier 1982; Messier et al. 2005).

the difficulty of acquiring an empirical dataset with the materiality of actual audit engagements, there has been only limited research in this are over the past 20 years. Blokdijk et al. (2003) demonstrated that auditors mainly use their clients' size, ROA, and internal control environment to decide the materiality. The researchers found that when the absolute value of ROA is close to zero, auditors set a conservative materiality level (i.e., with a low monetary threshold). More recently Choudhary et al. (2019) document the possibility of the potential relationship between loosened materiality level and low accounting reliability and provided empirical evidence by using their materiality dataset.

In this study, we analyze the actual materiality amount of 843 engagement cases obtained from Spanish accounting firms. We confirm prior findings with respect to the determinants of the materiality level. Additionally, we find that the materiality level is loosened during the early periods. In the initial periods of audit contracts, the materiality decreases with the audit tenure. Interestingly, after a certain point in the year, the level of materiality increases.[2]

This research may be beneficial to investors as well as regulators by providing the determinants and other characteristics of auditors' materiality decisions. First, our sample shows that the materiality level changes over time with a unique trend. Information users in the capital market consider the materiality level to be associated with the financial reporting quality (Amiram et al. 2017), and Choudhary et al. (2019) empirically prove this association. Based on these arguments, this study may suggest that accounting information users need to be especially cautious when they use the accounting

[2] Davis et al. (2009) find similar results by analyzing the relationship between the audit tenure and whether a company meets or beats analysts' forecasts. The likelihood of meeting or beating the forecasts has "U-shape" in their research along with the audit tenure.

information of new auditors. Second, our study provides evidence that new auditors consider previous auditors' audit opinions when they decide the materiality level. This study also suggests that when new auditors have a new client, they attempt to use as much information as possible to decide the materiality level. Interaction between auditors or auditor communities has not been widely studied. This study provides empirical evidence about this research avenue (i.e., communication between predecessor and successor auditors). Lastly, many studies use types of audit opinion as a proxy for audit quality (Krishnan and Krishnan 1996; Craswell et al. 2002; Li 2009; Hope and Langli 2010; Blay and Geiger 2013; Tepalagul and Lin 2015). Our study documents the relationship between the materiality level and the audit opinion. We find a negative association between the materiality level and the issue of a modified opinion when a client is a new one. This finding extends the existing literature by providing empirical evidence of the relationship between the materiality level and the audit quality (i.e. audit opinion).

The following sections of this paper are structured thusly: first, in Section 2, we discuss the auditing guidelines and previous literature on materiality. In Section 3, we develop the research hypotheses. We then discuss the research sample and research design in Section 4. We provide the results of the main and additional analyses in Section 5 and 6. Finally, we conclude with relevant contributions in Section 7.

4.2. Authoritative guideline and prior literature

The materiality is one of the major mechanisms that control the audit quality throughout the whole auditing procedures. IAASB (International Auditing and Assurance Standards Board) ISA (International Standards on Auditing) 320 and 450 describe the materiality as an essential component during the whole audit process. ISA 320 proclaims that the materiality is involved in "*planning*", "*performing*", and "*evaluating*" the audit (ISA 320 Para. 5). In addition, the materiality is considered when the auditor evaluates audit results at the final stage (ISA 450 para. 11; ISA 700 para. 11 and 17). Auditors issue their audit opinions, considering the accumulated and uncorrected misstatements, and the level of the materiality. The U.S. auditing standards are also similar to the abovementioned (PCAOB AS 2105 and 2810). From a practical perspective, the actual materiality level have a direct impact on the sample size during the test of detail procedures, if the auditor deploys statistical sampling method (AICPA 2017). This is directly related to the auditor's workload during the test of details.

When auditors set the materiality level, auditors are required to consider comprehensive factors of their clients and their specific circumstances. ISA 320 adds that the auditors need to take account of the information user's perspectives–*"the auditor's perception of the financial information needs of users of the financial statements*." The U.S. standards also emphasis that the materiality must be determined by taking into account both "qualitative and quantitative factors" (PCAOB AS 2105 and 2810; SEC SAB 99).

Due to this enormous important role and high complexity of the materiality, there has been extensive prior literature about the materiality (Holstrum and Messier 1982; Messier et al. 2005). Research uses the accounting firms' internal audit manuals for materiality decisions (Steinbart 1987; Eilifsen and Messier 2015). Or, the materiality level is inferred by utilizing the modified opinion or restatement cases (Acito et al. 2009). Keune and Johnstone (2012) use a unique regulation setting which requires to disclose the detected *immaterial* audit findings. However, the actual materiality amount of each audit engagement is hard to access publicly (Amiram et al. 2017; Choudhary et al. 2019). Because of this limited accessibility of the materiality dataset, there have few empirical studies about the materiality level based on actual monetary amount observations. The actual exact amounts of the materiality which are decided and applied to real audit engagements still remain untouched.

The actual materiality level varies across clients because the materiality amount is decided by the complicated judgmental process of auditors both explicitly and implicitly, considering each circumstance of the engagement (ISA320, PCAOB AS 2105; Amiram et al. 2018). By analyzing the real in-practice data, the research can reveal the mechanism of this complicated process, auditor's decision-making process on the materiality.

Among a few studies using the actual monetary amount of each engagement case, Blokdijk et al. (2003) use the 108 audit cases of the Netherlands from 13 audit firms for their research. They find that auditors consider the control environment, profitability, and complexity of the client when auditors decide the materiality level. Choudhary et al. (2019) use PCAOB's internal data that the accounting firms reported to the PCAOB during the inspection process. They document that there is no absolute prevail rules, such as 5% of pre-tax earnings, when auditors set the materiality monetary level. They propose a new measurement to gauge the looseness of each materiality level. By using this new measurement, they document that the loosened materiality is connected to both the restatement incidents (in the most loosened cases) and less proposed audit adjustments by auditors. Eventually, they point out that this loosened materiality may dampen the reliability of the financial reporting.

In the U.K., there were transformative regulatory changes in the audit report (FRC 2013). Certain companies are required to disclose the actual materiality amount in their auditor's report. Amiram et al. (2017) take advantage of the U.K.'s new regulation in order to obtain the monetary amount of actual materiality level. They test the extensive determinant factors of the materiality level. They find that debt financing and the level of inside shareholders are negatively associated with the materiality level. By analyzing the capital market reactions, they also document that investors rely more on the financial statements when the audit process is performed under the lower materiality level.

Our research is similar to the studies of Blokdijk et al. (2003), Amiram et al. (2018), and Choudhary et al. (2019) in that we try to find out the determinants for materiality decision. However, characteristics of our sample are different from theirs. The recent two studies uses the public companies for their samples.[3] In this study, our sample are all private companies. In Spain, it is mandatory to get the external audit if a company meets the any of two conditions out of three: the size of total assets, the revenue amounts, and the number of employees. It is required to be audited by the external auditor even if the company is not listed on the stock market. Given this different setting, we retest the

[3] Blokdijk et al. (2003) used both public companies and private companies. Among 108 their observations, 28.7% are listed companies.

determinants of the materiality. Furthermore, in our dataset, the modified opinions are prevail compared to the prior literature.[4]

Besides the difference of samples, our datasets provide more abundant observations with respect to the audit tenure. The sample of the prior literature is limited to only two consecutive fiscal years (Amiram et al. 2018; Choudhary et al. 2019). Our dataset has longer time spans.[5] Based on this unique characteristics, our dataset enables us to analyze the materiality level, especially, with respect to the audit tenure. We will discuss this more in the following sections.

4.3. Hypothesis development

The authoritative guidelines request auditors to consider quantitative factors as well as qualitative factors when the auditors decide the materiality level (ISA 320; PCAOB AS 2105 and 2810; SEC SAB 99). In other words, every auditor faces a different situation or circumstance and all of these situations are associated with decisions of the materiality level. Audit risk can be one of the qualitative factors which auditors need to consider with respect to the materiality level (Blokdijk et al. 2003; Eilifsen and Messier 2015; Choudhary et al. 2019; ISA 320; PCAOB 2015). [6]

---

[4] For the listed companies, most companies receive the unqualified opinion from the auditors (with or without the going concern opinion), in general. On the contrary, in our datasets that are randomly selected, all companies are private companies, and the ratio of qualified opinion is more than 25%.

[5] Some clients have more than eight-consecutive-year observations in our dataset.

[6] Strictly speaking, there are multiple types of materiality. Performance materiality level is different from overall materiality level (ISA 320 para. 10 and 11; PCAOB AS2105 para. 8 and 9; Eilifsen and Messier 2015). The overall materiality level is decided from the information user's perspective (i.e., whether a certain amount of misstatement is material to information users' decision-making or not). The performance materiality is used to achieve reasonable assurance by auditors. As a result, the performance materiality is less than the overall materiality by considering the audit risk and potential undetected misstatements (ISA

Auditors are required to maintain the overall audit risk at a low level by controlling its detection risk (ISA 200 para. 17; PCAOB AS 1101 para. 2 and para. 4). By doing so, auditors can provide reasonable assurance to the information users (ISA 200 para. 17; PCAOB AS 1101 para. 3). As a result, if the assessed risk (i.e., inherent risk or control risk) is high, the auditor needs to reduce the materiality level in order to achieve the desirable overall audit risk (ISA 200 para. 13; ISA 200 para. 9 and para. 10; PCAOB AS 1101 para. 7; PCAOB AS 2105 para. 8). By decreasing the materiality level, the auditor can enhance the effectiveness of his or her audit procedures. [7] [8] In these contexts, Blokdijk et al. (2003) find that the positive relationship between the materiality level and the quality of control environment as well as the negative relationship between the materiality level and the assessed level of entity complexity. Choudhary et al. (2019) document that material weakness of internal controls or new clients has a negative association with the level of materiality. Eilifsen and Messier (2015) also show that an internal manual from one of the accounting firms recommends auditors to reduce the materiality level when the audit is related to "*high-risk industries, unusually high market pressures, first-year and special purpose financial statements.*"

---

320 para. 10; PCAOB AS2105 para. 8; Eilifsen and Messier 2015). However, the prior literature mainly uses the overall materiality level for its empirical analysis (Blokdijk et al. 2003; Amiram et al. 2018; Choudhary et al. 2019). And they find some factors that are related to the audit risk (e.g., internal control effectiveness, entity complexity, internal material weakness, new client, etc.) have significant associations with the overall materiality level. In this research, we also use the overall materiality for the analysis. For the terminology, the ISA uses the term *performance materiality*. The US standard uses the term *tolerable misstatement* for the same concept.

[7] Empirically, Choudhary et al. (2019) document that the strict materiality is related to less restatement incidents and more proposed audit adjustments.

[8] However, the high materiality level is not necessarily problematic. If the risks of material misstatement (i.e., inherent risk or control risk) is low, the auditor can set the high materiality level. By doing so, auditors can perform the audit procedure efficiently while delivering reasonable assurance to the information users.

As discussed above, Choudhary et al. (2019) find that the auditors set the materiality level relatively low when it comes to new clients.[9] Choudhary et al. (2019) use new clients as a way of measuring the client complexity. They briefly mention that auditors regard new clients as more complex and riskier audits, which lead to the reduced materiality level.

The same arguments—a new client is a riskier one—can be applied to our sample. However, there is still another possibility. A new client can generate a lot of workload and pressure to an auditor. The heavy workload can cause high audit costs. Especially, if a new audit is initiated with a reduced audit fee, it may create economic pressure (i.e., budget constraint) to auditors (Blankley et al. 2012).

There have been studies that investigate the relation between the audit tenure and the audit quality. Johnson et al. (2002), and Carcello and Nagy (2004) find that the quality of financial reports is lower in the initial years with a new client (e.g., in the first three years). In addition, Choudhary et al. (2019) document that the loosened materiality (i.e., higher amount of the materiality level) is related to low accounting reliability. If we take into account these findings together, the low accounting quality of new clients can be associated with the higher materiality levels.

In this study, we investigate determinants of the materiality. More specifically, we test whether the auditors reduce the materiality level when they start an audit with a new client. This has been tested in one of the previous literature (Choudhary et al. 2019). [10]

[9] Blokdijk et al. (2003) find no association with the audit tenure. Amiram et al. (2018) do not include either the new client dummy variable or the audit tenure variable.
[10] Choudhary et al. (2019) document the negative association between materiality and new clients. However, the results are unstable. Among their three models, only one specification produces a statistical

However, as we discussed above, if we consider the low audit quality during the initial years, there are still two opposite possibilities. It can be either higher materiality or lower materiality, although we posit a lower materiality hypothesis consistent with the prior literature.

H1: *The level of the materiality of the new client would be lower than the one of the continuing clients.*

As the prior literature points out, the financial reporting quality can be enhanced as the same auditors continue the auditing (Johnson et al. 2002; Carcello and Nagy 2004). Although there are many factors which may affect the audit quality, the level of materiality would be one of those factors (Choudhary et al. 2019; ISA 320 para. 9, 10, and 11). In other words, if the same auditor continues the auditing with a certain client, the audit quality would be improved over the years. And the audit quality is associated with the level of the materiality. If we rely on this argument, the improved accounting quality over the years can be obtained through the reduced materiality level.

However, there may exist an alternative possibility. As the auditors continue the audit with the same clients, they are able to capture and measure the inherent risk and the control risk more precisely. In this case, the auditor's initial conservativeness can be reduced. Choudhary et al. (2019) document that new clients are assigned with lower

significance less than 1%. One has a less than 10% p-value, and the other has more than 10% significance level on the new client variable.

materiality levels, because the auditor considers the new clients as relatively complex clients (i.e., risky audit engagement). If we rely on this finding, the perceived risk level by auditors can be reduced as the same auditor continues the auditing over the years by accumulating the understanding about their clients. This would lead to the increase in the materiality level as time goes by. However, prior literature (Blokdijk et al. 2003) finds no association with the audit tenure. The direction between the materiality level and the audit tenure is an open empirical question.

H2*: When the same auditor conducts the auditing over multiple years, the materiality level would be increased over the years.*

If H2 is true, we can think about the timing of these increases in the materiality level. As mentioned before, there has been research between the audit tenure and the audit quality.[11] This prior research divides the audit tenure into three parts (i.e., short, medium, and long). Some use the first three years as an initial period (Johnson et al. 2002; Carcello and Nagy 2004; Jenkins and Velury 2008). Meanwhile, Davis et al. (2009) find auditor's behavior changes when the same auditor continues auditing for a long period (i.e., longer than 14 years). The specific point in time may vary, depending on the samples or the specific settings. However, all of them are pointing out that

[11] Tepalagul and Lin's literature review paper (2014) summarizes the research about the tenure and the audit quality. Please see this paper for more details.

auditors' behavior changes at a certain point in time. If H2 is valid, we can further investigate to find out when this increase in the materiality level happens.

H2a: *When the materiality level increases over the years, those increases will occur after a certain period of time.*

4.4. Sample description and research design

4.4.1. Sample

In this study, we deploy the Spanish small and mid-sized accounting firms' audit engagement data. The dataset consists of 843 observations with 259 clients from 11 accounting firms. The fiscal year periods are from the year 2001 to 2009. The clients are randomly selected by considering their size and industry. From the original data obtained from the Spanish accounting firms, the voluntary audit engagements are excluded in order to eliminate the selection bias. We obtain the financial variables from the Iberian Balance sheet Analysis System (SABI) database. The 857 observations which are not matched with the SABI database are removed from the final sample.

The accounting firms provide the auditor's own evaluation regarding the integrity of the management and the effectiveness of the internal control on each audit engagement, which is documented in their audit working papers. In addition, the accounting firms provide us the monetary amount of the materiality on each engagement. Table 25 and Table 26 illustrate the sample of this study.

Table 25 Sample description per year

Year	Number of observations	Percentage
2001	3	0.40%
2002	15	1.80%
2003	76	9.00%
2004	117	13.90%
2005	122	14.50%
2006	147	17.40%
2007	137	16.30%
2008	114	13.50%
2009	112	13.30%
Total	843	100.00%

Table 26 Sample description per industry

The first digit of US SIC code	Industry Name	Number of observations	Percentage
0	Agriculture, Forestry, and Fishing	26	3.1%
1	Mining, Construction	216	25.6%
2	Manufacturing (Consumer goods)	92	10.9%
3	Manufacturing (Tools and machinery)	78	9.3%
4	Transportation, Communications, Electric, and Gas service	49	5.8%
5	Wholesale and Retail	309	36.7%
6	Finance, Insurance and Real Estate	32	3.8%
7	Services (Accommodation, Communication, and Entertainment)	31	3.7%
8	Services (Health and Legal)	10	1.2%
	Total	843	100.0%

* SABI database also provides the industry classification with the US SIC codes.

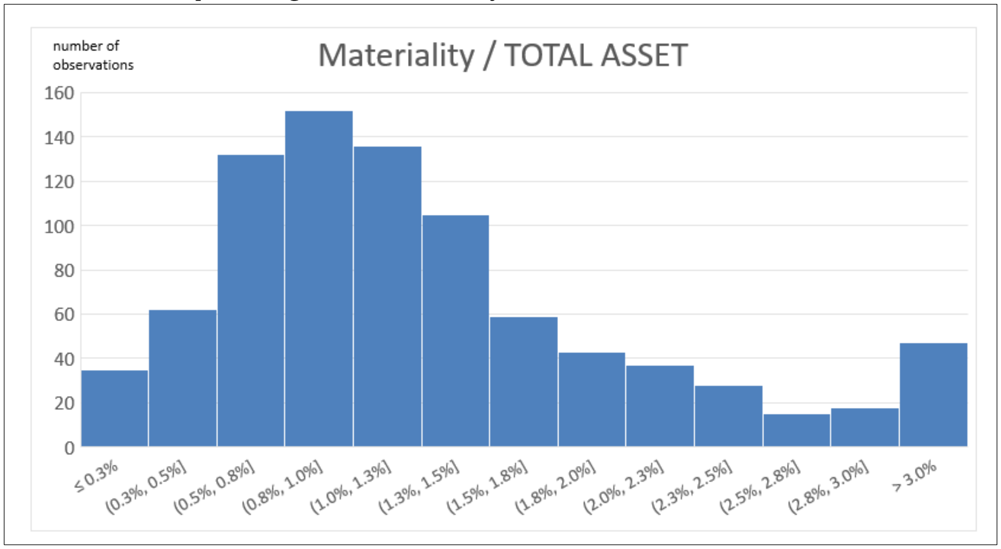
4.4.2. Testing the general rules of the materiality decision

The materiality level is determined by multiplying a certain percentage value with a benchmark account, the line item in the financial statements. (Blokdijk et al. 2003; Eilifsen and Messier 2015; Amiram et al. 2018; Choudhary et al. 2019). In this section, we briefly explore the materiality level decision in terms of using benchmarks.

'Pre-tax income' is one of frequently referred benchmarks (Blokdijk et al. 2003; Choudhary et al. 2019). Total assets amount, or total sales amount is also widely used (Eilifsen and Messier 2015; Amiram et al. 2018; Choudhary et al. 2019). Figure 4 shows the histogram of the ratio between the materiality amount and the total assets amount on each engagement. As conventionally mentioned, 1% of total assets amount is the most frequent in our sample distribution. However, there are other cases. For example, there are considerable observations whose materiality is more than 3% of its total assets.

Figure 4 Histogram of materiality to total assets

The most frequent range of the materiality lies on around 1% of total asset amounts.



Similarly, Figure 5 shows the histogram by using the ratio between the materiality amount and the sales amount. In Figure 5, around 1% of the total sales is the most frequent. However, there do exist cases that have less than 0.3% or above 3% of the sales amount for the materiality level. Figure 6 uses the pre-tax income amount. With respect to the pre-tax income amount, the range between 5% and 10% is the most frequent one. However, observations, which are less than -10% or more than 80% of pre-tax income amount, are also prevailing in our sample.

Based on the histograms, we reconfirm the finding of the previous literature (Blokdijk et al. 2003; Choudhary et al. 2019). There is no single dominant determinant of the materiality level (e.g., pre-tax income amount). Each materiality level is decided depending on the circumstance of each audit engagement. In the next section, we will look further into this materiality determinants.

Figure 5 Histogram of materiality to total sales amount

The most frequent range of the materiality lies on around 1% of total sales amounts.

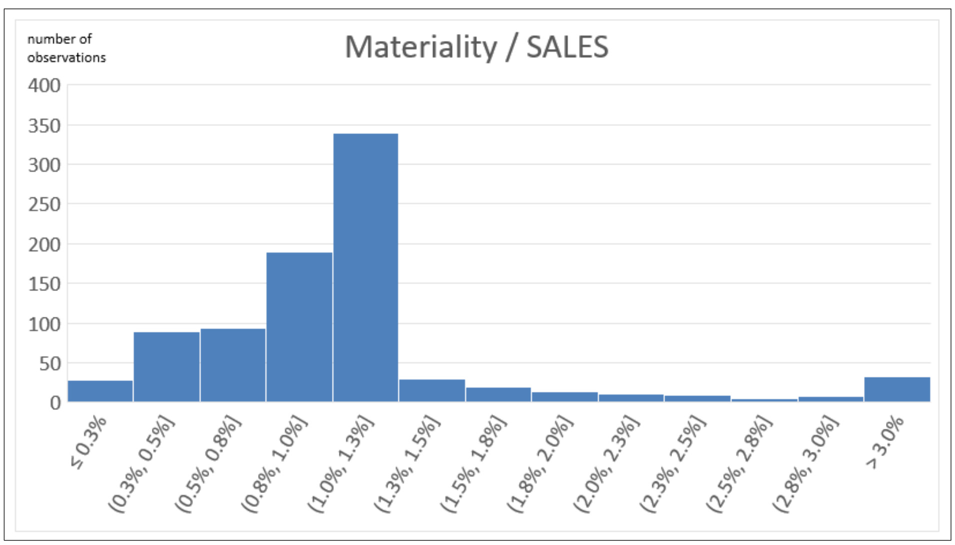
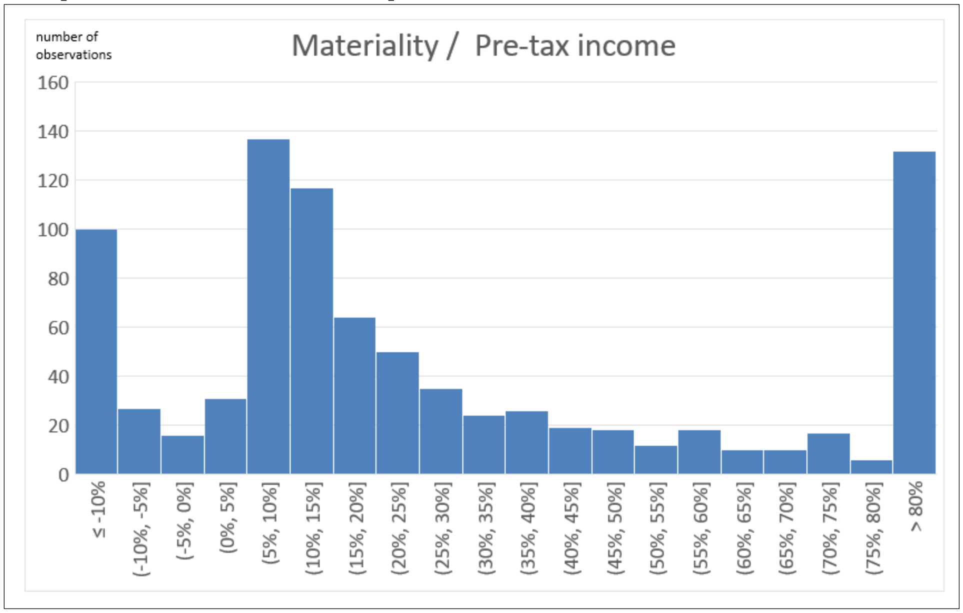


Figure 6 Histogram of materiality to pre-tax income

Materiality level is decided around 5%~10% of total net income. However, there are other cases. For example, some cases are 80% of its pre-tax income.



4.4.3. Research Design

This study tries to find determinants of the materiality in practice. We use a similar model that is used in the previous research of the materiality (Blokdijk et al. 2003; Choudhary et al. 2019). [12] In order to test our Hypothesis 1 and 2, we used the following model:

$$
\begin{aligned}
\text{Log}(\text{MATERIALITY}_{it}) = {} & \beta_0 + \beta_1\,\text{NEW\_CLIENT}_{it} + \beta_2\,\text{AUDIT\_TENURE}_{it} \\
& + \beta_3\,\text{LOG\_TOTAL\_ASSET}_{it} + \beta_4\,\text{LOG\_SALES}_{it} \\
& + \beta_5\,\text{LOG}\left(|\text{Pretax income}|_{it}\right)
\end{aligned}
$$

[12] Blokdijk et al. (2003) document that the materiality level is positively associated with the size variables (i.e., total assets or sales amount). However, it marginally diminishes on the scale. Based on this, they use the logarithm transformation on the materiality amount and the size variable.

$$+ \beta_6\, CURR\_RATIO_{it} + \beta_7\, DEBT\_RATIO_{it} + \beta_8\, ROA_{it}$$
$$+ \beta_9\, MANAGEMENT\_INTEGRITY_{it}$$
$$+ \beta_{10}\, INTERNAL\_CONTROL\_EFFECTIVENESS_{it}$$
$$+ \beta_{11}\, SMALL\_ROA\_DUMMY_{it}$$
$$+ \text{Fixed effects (industry and year)} + \varepsilon_{it} \quad (4)$$

MATERIALITY is the amount of the materiality. NEW_CLIENT is a dummy variable indicating whether the client is new to the accounting firm or not. If it is a new client to the auditor, the value is 1. If it is a continuing client, the NEW_CLIENT value is 0. AUDIT_TENURE is an audit tenure year variable. This starts from 1, thus, the value of this variable with a new client observation would be 1. TOTAL_ASSET is the amounts of total assets at the fiscal year end. SALES is the amount of total revenue. ABS_PRETAX_INCOME is the absolute value of the ordinary pretax income. We did not include the extraordinary incomes in this variable, because the prior literature documents that the auditors exclude the not-recurring numbers when they decide the materiality base amount (Eilifsen and Messier 2015; Choudhary et al. 2019). And Choudhary et al. (2019) also find that many of *negative* income observations in their samples indicate the absolute amount of income account as its base account (benchmark account) when auditors decide the materiality.[13] DEBT_RATIO is the leverage ratio, and ROA is the return on assets. [14] MANAGEMENT_INTEGRITY is the audit engagement team's own assessment of the integrity of management. The range of this variable varies

[13] Choudhary et al. (2019) use their privileged dataset obtained from the regulator (PCAOB). Their dataset contains the information about which base account the auditor used for deciding the materiality amount.
[14] We use return with pretax-ordinary activity income. Blokdijk et al. (2003) find that there is an association between the materiality and the ROA in their sample. They used the ROA by using net income before tax for its calculation.

from 1 to 3. If the integrity of the management is high, the variable would be 3. If it is low, the value would be 1. INTERNAL_CONTROL_EFFECTIVENESS is the audit team's assessment of the internal control's effectiveness of the client. 3 means effective, and 1 means less effective. All the financial variables are obtained from the SABI database. Appendix I shows the definition of all variables.

Choudhary et al. (2019) apply the three size financial variables (total asset, sales, and pretax income) at the same time in one single equation, mentioning that their correlations are between 0.73 and 0.82 and their VIFs are not exceeding 5. In our sample, the correlation among these three variables are much less than the ones of Choudhary et al. (2019); VIFs of these variables are less than 4 except for one regression.[15] Our sample is much less correlated with each other than previous literature. We apply the same approach as the previous literature (Blokdijk et al. 2003; Choudhary et al. 2019).

4.5. Empirical results

4.5.1. Descriptive statistics

Table 27 and Table 28 show the descriptive statistics of the sample. The sample contains 9.6% of the newly started engagement cases (new clients) among the 843 observations. The average contract year of the sample is 7.1 years. In the sample, there are 259 distinct clients and some clients are included multiple times with different years. The average level of the management integrity is 2.68 (out of 3). Thus, many

[15] The regression in Table 32 Column 2 has 6.5 VIF value with the logarithm of the pre-tax income variable. The regression result is robust without this pre-tax income variable. Except for this case, all the other analysis, VIFs values are not exceeding 3.9 on these variables in our study.

engagements are assessed as high integrity levels (There are 646 observations that are assessed as high management integrity). The correlation between the level of management integrity and the effectiveness of internal controls confirms general perceptions. The integrity of management has a significantly positive correlation with the effectiveness of internal control.

Table 27 Descriptive Statistics

Variables	N	Mean	St.Dev	p1	p25	Median	p75	p99
LOG_MA	843	4.621	0.798	2.597	4.111	4.588	5.160	6.705
NEW_CLIENT	843	0.096	0.295	0	0	0	0	1
AUDIT_TENURE	843	7.097	5.375	1	3	5	11	20
LOG_TA	843	9.191	0.888	7.321	8.553	9.101	9.71	11.85
LOG_SALES	843	9.302	0.836	6.179	8.841	9.301	9.81	11.35
LOG_ABS_PRETAX_INCOME	843	5.957	1.512	1.587	5.006	6.072	6.98	9.144
CURR_RATIO	843	1.785	1.646	0.219	0.987	1.251	1.84	10.42
DEBT_RATIO	843	0.641	0.237	0.082	0.473	0.68	0.84	1.014
IC_EFFECTIVENESS	843	2.344	0.693	1	2	2	3	3
MANAGEMENT_INTEGRITY	843	2.684	0.617	1	3	3	3	3
ROA	843	0.045	0.085	-0.29	0.008	0.036	0.08	0.317
ABS_ROA	843	0.071	0.074	0.001	0.019	0.05	0.09	0.416
SMALL_ABS_ROA_DUMMY	843	0.503	0.5	0	0	1	1	1
PRIOR_QUALIFIED_OPINION	779	0.316	0.465	0	0	0	1	1
MA_RESISUAL	843	0	0.447	-1.33	-0.24	0.028	0.3	1.082
QUALIFIED_DUMMY	843	0.276	0.447	0	0	0	1	1
AUDIT_TENURE SQ	843	79.23	104.184	1	9	25	121	400
SHORT_TENURE	843	0.331	0.471	0	0	0	1	1
LONG_TENURE	843	0.222	0.416	0	0	0	0	1

All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

Table 28 Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1) LOG_MA	1																		
(2) ABS_ROA	0.100***	1																	
(3) AUDIT_TENURE	0.214***	-0.099***	1																
(4) AUDIT_TENURE_SQ	0.226***	-0.067*	0.968***	1															
(5) CURR_RATIO	0.051	0.105***	0.114***	0.107***	1														
(6) DEBT_RATIO	-0.047	-0.261***	-0.251***	-0.244***	-0.554***	1													
(7) IC_EFFECTIVENESS	0.140***	0.076**	0.032	0.034	0.129***	-0.125***	1												
(8) LOG_ABS_PRETAX_INCOME	0.451***	0.629***	0.060*	0.071**	0.177***	-0.208***	0.066*	1											
(9) LOG_SALES	0.650***	0.043	0.191***	0.190***	-0.169***	0.081**	0.035	0.399***	1										
(10) LOG_TA	0.640***	-0.039	0.198***	0.181***	0.213***	-0.009	0.009	0.563***	0.504***	1									
(11) LONG_TENURE	0.179***	-0.03	0.864***	0.901***	0.102***	-0.233***	0.032	0.065*	0.155***	0.147***	1								
(12) MANAGEMENT_INTEGRITY	0.066*	0.064*	0.024	0.032	0.070**	-0.04	0.610***	0.016	0.021	-0.068**	0.023	1							
(13) MA_RESI_RATIO	0.618***	0.093***	-0.003	-0.004	-0.021	-0.023	-0.008	0.019	0.043	0.003	-0.005	0.026	1						
(14) NEW_CLIENT	-0.059*	0.087**	-0.370***	-0.245***	-0.056*	0.077**	-0.022	-0.02	-0.059*	-0.115***	-0.174***	-0.003	0.006	1					
(15) PRIOR_QUALIFIED_OPINION	-0.108***	-0.038	-0.297***	-0.265***	-0.061*	0.053	-0.021	-0.035	-0.122***	-0.054	-0.245***	-0.065*	-0.048	0.138***	1				
(16) QUALIFIED_DUMMY	-0.068**	-0.015	-0.268***	-0.247***	-0.087**	0.129***	-0.05	-0.026	-0.047	-0.019	-0.241***	-0.088**	-0.046	0.240***	0.574***	1			
(17) ROA	0.213***	0.357***	-0.032	-0.014	0.130***	-0.323***	0.031	0.341***	0.252***	0.066*	0	0.049	0.003	0.046	-0.028	-0.078**	1		
(18) SHORT_TENURE	-0.111***	0.139***	-0.658***	-0.502***	-0.097***	0.136***	-0.033	-0.03	-0.116***	-0.183***	-0.376***	-0.024	0.013	0.464***	0.287***	0.202***	0.074**	1	
(19) SMALL_ABS_ROA_DUMMY	-0.064*	-0.671***	0.061*	0.044	-0.079**	0.241***	-0.078**	-0.619***	-0.097***	0.022	0.028	-0.055	-0.001	-0.03	0.037	0.026	-0.365***	-0.062*	1

*, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

4.5.2. Main results for testing hypothesis

Table 29 shows the regression results, confirming the findings of prior literature. Column (1) Table 29 includes all variables that Blokdijk et al. (2003) examine. [16] Blokdijk et al. (2003) do not find statistical significance with current ratio, audit tenure, and management integrity. Similarly, in our sample, current ratio, management integrity, and audit tenure have no significant relationships with the materiality level by using the same model specification. Unlike Blokdijk et al. (2003), in our sample, the materiality level has no significant relationship with small ROA companies (i.e., the absolute value of ROA is less than 5 percent). [17] They find that the materiality level is reduced when the company has an ROA value close to zero.

Choudhary et al. (2019) have a privileged sample which is obtained from the regulator, the US PCAOB. Their sample includes the benchmark information—the line item account that each auditor used in deciding the materiality level. More than half (59.7%) in their samples use pretax income amount as the materiality base account. However, Blokdijk et al. (2003) comment that only 31 percent of their samples exist within the range from 5% to 10% of the absolute pretax income. Blokdijk et al. (2003) use Netherlands audit engagement observations. Our Spanish audit sample is more like Blokdijk et al. (2003)'s sample. Among 21.7% of our sample, the materiality amount exists within the range from 5% to 10% of absolute pretax income. It seems that, in our

[16] Blokdijk et al. (2003) use a size variable with total assets and the total sales. In our research, we use them individually like Choudhary et al. (2019)'s approach. It is not clear whether Blokdijk et al. (2003) include the fixed effects in their analysis. We include fixed effects to control the year and industry effects as Choudhary et al. (2019) perform.

[17] Among 3, 4, and 5 percent ROA indicator variables, Blokdijk et al. (2003) document the 5-percent ROA dummy variable shows the highest t-statistics.

sample, the auditors utilized other accounts rather than the pretax income when they decided the materiality level. Column (1) Table 29 confirms this argument.

Regarding our hypothesis H1, the materiality levels of new clients are higher than the ones of the continued clients. Column (2) Table 29 shows the result with respect to new clients and ROA values. And Column (3) Table 29 is the result with only statistically significant variables. With respect to the audit tenure, the materiality level increases over the years if the same auditor continues the auditing. This confirms our H2. The magnitudes of these two coefficients are interesting. The coefficient of new clients is 0.1339, and the one of audit tenure is 0.0102. Thus, it takes 13.1 years if a continuing client gets the same level of the materiality of a new client (13.1 = 0.1339 ÷ 0.0102). The coefficients of other control variables are consistent with our prediction in that the effectiveness of internal control and ROA have a positive association with the materiality level. Interestingly, the current ratio is not significant with the winsorized dataset. However, with the raw dataset, it has a positive relationship. [18]

[18] This may suggest that the auditor set a low materiality when the current ratio of his or her client has an extremely low value of current ratio.

Table 29 Determinants of Materiality

Variables	Prior Literature (Blokdijk 2003) (1)	Full variables (2)	Significant variables (3)	No winsorized dataset (4)
	LOG_MATERIALITY			
NEW_CLIENT		0.1314*	0.1339*	0.1544*
		(1.7740)	(1.7873)	(1.9573)
AUDIT_TENURE	0.0042	0.0086	0.0102*	0.0077
	(0.8910)	(1.6318)	(1.8466)	(1.4658)
LOG_TA	0.3518***	0.3558***	0.3520***	0.3605***
	(5.9540)	(4.8023)	(5.7766)	(5.8805)
LOG_SALES	0.4235***	0.3907***	0.3938***	0.3992***
	(6.4298)	(5.9415)	(5.9660)	(6.0673)
LOG_ABS_PRETAX_INCOME		0.0063		
		(0.2071)		
CURR_RATIO	-0.0193	-0.02		0.0081***
	(-0.7905)	(-0.8427)		(3.7039)
DEBT_RATIO	-0.2875*	-0.1736		-0.1595
	(-1.9034)	(-1.1228)		(-1.3713)
IC_EFFECTIVENESS	0.1078**	0.1197***	0.1409***	0.1385***
	(2.4526)	(2.7212)	(4.2490)	(4.1094)
MANAGEMENT_INTEGRITY	0.0506	0.0404		
	(0.9359)	(0.7469)		
ABS_ROA	0.2263			
	(0.7534)			
SMALL_ABS_ROA_DUMMY	0.0038	0.022		
	(0.0829)	(0.3642)		
ROA		0.8599**	0.9735***	0.4204**
		(2.5821)	(3.1111)	(2.1522)
Obs.	843	843	843	843
R-squared	0.6131	0.6199	0.6179	0.6101
INDUSTRY FE	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES
WINSORIZING	1%, 99%	1%, 99%	1%, 99%	No
SD CLUSTERING	By client	By client	By client	By client

Standard errors are clustered by client. t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

The results of Table 29 shows that the materiality level is higher when an auditor starts a new audit, compared to continuing audits. [19] At the same time, the materiality level is increasing along with the audit tenure. This is interesting as we mentioned in the previous paragraph. In order to investigate further the relationship between materiality level and audit tenure, we examine Equation 5 based on our findings in Table 29.

$$
\begin{aligned}
\text{Log (Materiality}_{it}) = \beta_0 &+ \beta_1 \text{ AUDIT\_TENURE}_{it} \\
&+ \beta_2 \text{ LOG\_TOTAL\_ASSET}_{it} + \beta_3 \text{ LOG\_SALES}_{it} \\
&+ \beta_4 \text{ ROA}_{it} + \beta_7 \text{ INTERNAL\_CONTROL}_{it} \\
&+ \text{Industry Fixed effects} + \text{Year Fixed effects} + \varepsilon_{it} \qquad (5)
\end{aligned}
$$

Column (2) Table 30 shows the estimation result of Equation 5. Figure 7 illustrates the relationship between the residuals of Equation 5 and the audit tenure. Based on the residuals, we draw a trendline in Figure 7 by using Microsoft Excel's trendline function. The trendline is estimated by setting a polynomial line with the second order (i.e., a parabola line with the second order of x-axis, the audit tenure).

[19] The untabulated result with client level fixed effect has robust results. The new client dummy variable has a significantly positive coefficient (1.8% of p-value without clustering; 10.2% of p-value with client level clustering).

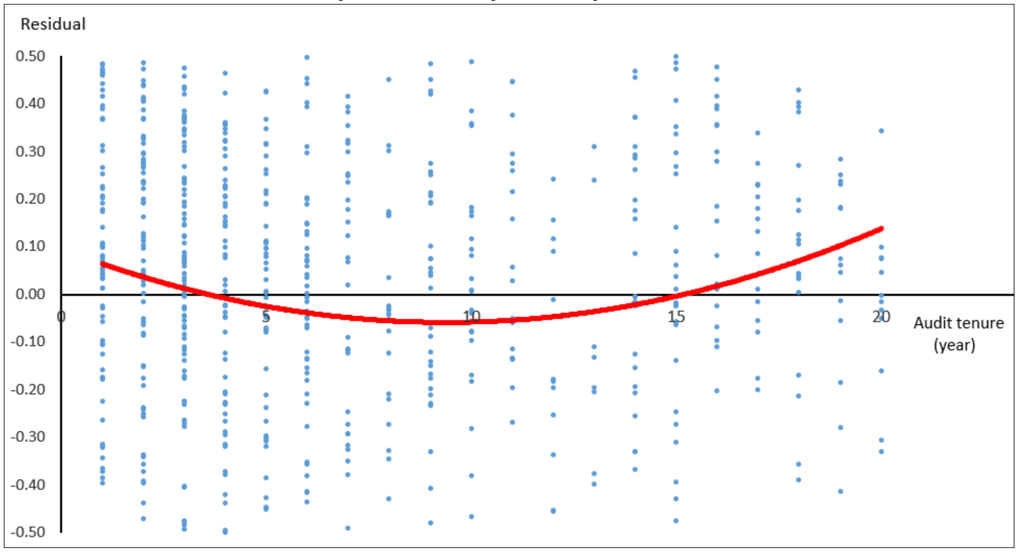
Table 30 Materiality over audit tenure

Variables	(1) LOG_MA	(2) LOG_MA	(3) LOG_MA	(4) LOG_MA
AUDIT_TENURE_SQ			0.0018**	
			(2.4979)	
AUDIT_TENURE		0.0076	-0.0265*	
		(1.5357)	(-1.8809)	
SHORT_TENURE				0.0847
				(1.4583)
LONG_TENURE				0.1114*
				(1.8039)
LOG_TA	0.3604***	0.3500***	0.3506***	0.3581***
	(5.8921)	(5.7288)	(5.7697)	(5.9200)
LOG_SALES	0.3964***	0.3947***	0.3962***	0.3919***
	(5.9852)	(5.9821)	(6.0668)	(5.9487)
IC_EFFECTIVENSS	0.1429***	0.1400***	0.1404***	0.1421***
	(4.2065)	(4.2157)	(4.2825)	(4.2412)
ROA	0.9619***	0.9821***	0.9316***	0.9563***
	(3.1270)	(3.1411)	(2.9946)	(3.0618)
Obs.	843	843	843	843
R-squared	0.6138	0.6159	0.6193	0.6177
INDUSTRY FE	YES	YES	YES	YES
YEAR FE	YES	YES	YES	YES

Standard errors are clustered by client. t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

Figure 7 Residual and audit tenures by the same auditor

Residual are estimated by using the first order equation with audit year term. Residuals decrease around 9 years. After 9 years, they increases.



Based on Figure 7, we introduce a polynomial equation for estimating the materiality level like Equation (6). [20]

$$
\begin{aligned}
LOG\_MATERIALITY_{it} = \beta_0 + \beta_1 (AUDIT\_TENURE_{it})^2 \\
+ \beta_2 AUDIT\_TENURE_{it} \\
+ \beta_3 LOG\_TOTAL\_ASSET_{it} + \beta_4 LOG\_SALES_{it} \\
+ \beta_5 ROA_{it} + \beta_8 INTERNAL\_CONTROL_{it} \\
+ \text{Industry fixed effects} + \text{Year fixed effects} + \varepsilon_{it} \qquad (6)
\end{aligned}
$$

Column (3) Table 30 shows the results of estimating Equation 6. Both β_1 and β_2 are statistically significant. Based on this result, we can estimate the AUDIT_TENURE

[20] This second-order polynomial equation is used in the prior literature of the audit tenure study (Davis et al. 2009). They studied the association between the audit tenure and the meeting or beating analyst earnings forecast. They find the U-shape behavior over the audit tenure.

that minimizes the materiality level by using the first order condition (i.e., $\frac{\partial \text{ Log (Materiality)}}{\partial \text{ Audit\_Tenure}} = 0$). The materiality becomes minimized when the same auditor continues auditing for approximately 7.4 years ($= \frac{0.0265}{0.0018 \times 2}$). Figure 8 shows distributions of the residuals from Equation 6 with respect to the audit tenure. The pattern of the residuals which we can see in Figure 7 disappears in Figure 8. By comparing Figure 7 with Figure 8, we may say that Equation 6 is closer to the underlying structure with respect to the auditor's decision about the materiality level. This result suggests that the auditor's behavior changes around eight years. This finding may be related to some of the prior literature (Deis and Giroux 1992; Carey and Simnett 2006; Davis et al. 2009). This prior literature finds that the audit quality becomes low when the same auditor continues the audit. [21] Our finding may suggest a possibility that the low quality with a longer auditor in prior research is related to the increased materiality level of their audit engagements. In addition, Column (4) Table 30 shows the result with another model specification. We use binary variables for the audit tenure. If the audit tenure is less than four years, we classified them as the earlier period. If the audit year is longer than eleven years, we categorized those observations as the long period. Column (4) Table 30 shows the consistent results with other models. In the earlier and the later periods, the materiality level is higher than the materiality of the middle period (i.e., audit tenure from four to eleven years). For the earlier periods, it is marginally significant (t-value: 1.56, p-value: 12%).

[21] There is also other research that documents that there is no audit quality change between medium-tenure audits and long-tenure audits. Please see Tepalagul and Lin (2016) for more details.

Figure 8 Residual and audit tenures by the same auditor

Residual are estimated by using the second-order equation with audit year term.
Residual has no significant trend in terms of audit years.

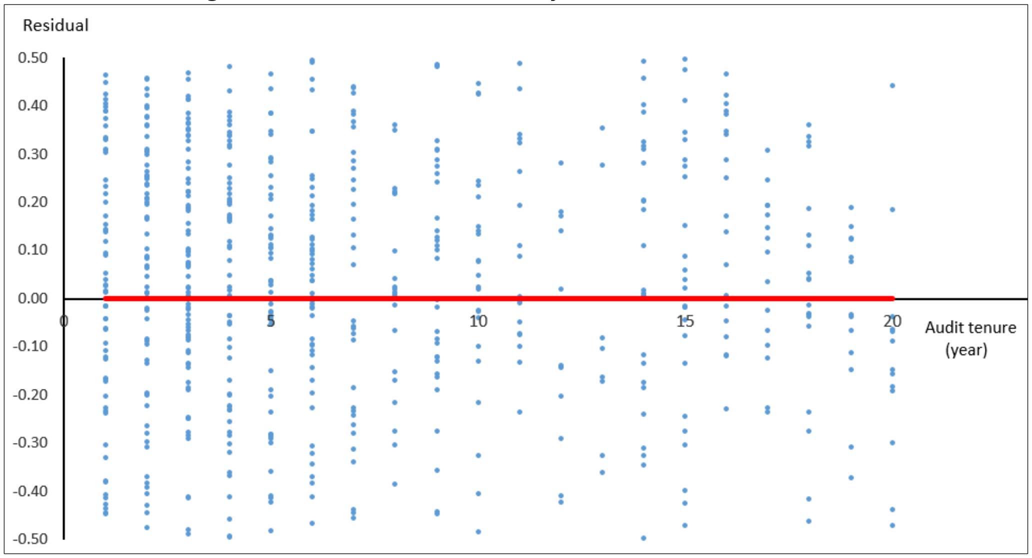


Table 30 shows the comprehensive results with respect to the materiality and audit tenure. There exists a non-linear relationship between the materiality and audit tenure. Blokdijk et al. (2003) do not find the significant relationship between the materiality level and audit tenure. Table 30 may explain why they do not find the association. Column (2) Table 30 also shows that no statistical significance when we include only the first-order audit tenure variable in the model as Blokdijk et al. (2003) do. The underlying non-linear relationship of audit tenure might prevent Blokdijk et al. (2003) from finding this association. Figure 5, Figure 6, and Table 30 support this argument.

4.6. Additional analysis

In the previous section, we examined the determinants of the materiality level. In this section, we investigate the relationship between the materiality level and the audit opinion. As we briefly mentioned in the dataset description section, our dataset consists of all private companies which are required to be audited by independent auditors. It is required by the Spanish legislation. One of the interesting aspects in our dataset is that this dataset has a considerable portion of the modified opinions (i.e., qualified opinion, or disclaimer opinion). In the previous section, we found that the materiality level of new clients is higher than continuing ones. This may suggest that the auditors have difficulty in assessing the risk level of their new clients. In this section, we investigate this further by using audit opinion.

Table 31 shows the results between the materiality level and the prior year's audit opinion. The result of column (1) contains the prior audit opinion variable as an independent variable. The coefficient of the prior audit opinion is negative, but it is not statistically significant. We add the interaction terms of the prior year's audit opinion. Column (2) Table 31 shows the result with these interaction terms. The auditor, especially the auditor who starts auditing a new client, significantly uses the prior audit opinion when they decide the materiality level. This may be due to the fact that a new auditor has a lack of information about his or her new client. As a result, the new auditor tries to use as much information as possible, including the predecessor auditor's opinion, in order to make the best assessment of his or her new client. When the auditor's opinion of the prior year is a qualified opinion, a new auditor reduces the materiality level compared to the unqualified opinion. [22]

Table 31 Regression results with the audit opinion of the prior year

	(1)	(2)
Variable	LOG_MA	LOG_MA
PRIOR_QUALIFIED_OPINION	-0.0471	0.0534
	(-0.8984)	(0.6161)
PRIOR_QUALIFIED_OPINION * NEW_CLIENT		-0.3896*
		(-1.7161)
PRIOR_QUALIFIED_OPINION * AUDIT_TENURE		-0.0124
		(-1.4316)
NEW_CLIENT	0.1455	0.3538*
	(1.2692)	(1.7316)
AUDIT_TENURE	0.0108*	0.0139**
	(1.7820)	(2.0188)
LOG_TA	0.3536***	0.3521***
	(5.2583)	(5.2787)
LOG_SALES	0.3795***	0.3809***
	(5.4070)	(5.4703)
IC_EFFECTIVENSS	0.1412***	0.1416***
	(4.0330)	(4.0731)
ROA	0.8927***	0.8976***
	(2.7749)	(2.7569)
Obs.	779	779
R-squared	0.6136	0.6168
INDUSTRY FE	YES	YES
YEAR FE	YES	YES

Standard errors are clustered by client. t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

Lastly, we test the relationship between the materiality level and the current audit opinion. Prior research uses the accounting reporting reliability by deploying the restatement incidents and the proposed adjustments by the auditor (Choudhary et al.

[22] Based on the estimation of Table 31, a new client with the clean opinion from the prior year's auditor would have 0.3677 materiality level in this year (=0.3538 + 0.0139 × 1). However, a new client with the modified opinion from the prior year's auditor would have 0.0191 materiality level in the first year (=0.0534 - 0.3896 - 0.0124 × 1 + 0.3538 + 0.0139 × 1).

2019). Those types of information are not available to our dataset. However, under the mandatory auditing setting in Spain, our datasets have a considerable portion of modified opinion. We can use the type of audit opinions as a proxy for the audit quality (Krishnan and Krishnan 1996; Craswell et al. 2002; Li 2009; Hope and Langli 2010; Blay and Geiger 2013; Tepalagul and Lin 2015).

Materiality is used in the final stage of auditing when the auditors evaluate their findings and when they form their audit opinions (ISA 450 and 700; PCAOB AS 2810 and 3105). The lower materiality level can lead to the more likelihood of the modified opinion, as long as the auditors use the materiality level as the final criterion to decide his or her final audit opinion. This is obvious; however, it has not been empirically documented.

The materiality level is the relative measurement which is determined by considering the size of the company and other factors (Blokdijk et al. 2003; Eilifsen and Messier 2015; Amiram et al. 2018; Choudhary et al. 2019). [23] As a result, we need to measure whether the materiality level is relatively high or low. Choudhary et al. (2019) use the possible list of candidates (i.e., 5% of net income, 1% of total assets, or 1% of total revenue, etc.). And then, they calibrate the relative location among these possible candidates in order to measure the materiality looseness of each audit engagement. In this paper, we used a simple approach. By using Equation 6, we can estimate the average

[23] In some sense, the audit fee is similar to the materiality level. Usually, the audit fee or the materiality level is proportional to the size of the client.

expected materiality level. Based on this, we use the residual as a measurement for the materiality looseness. [24]

By using this materiality looseness, we can examine the relationship between the materiality looseness and the final audit opinion with Equation 7.[25]

$$\begin{aligned} QUALIFIED\_DUMMY_{it} = \alpha_0 + \alpha_1\, Materiality\_Looseness_{it} \\ + \text{Other control variables}_{it} \\ + \text{Industry Fixed effects} + \text{Year Fixed effects} + \varepsilon_{it} \end{aligned} \quad (7)$$

Our logistic regression result is documented in Table 32. Column (1) Table 32 shows the results with the full sample. The association between the materiality looseness and the audit opinion of the current year has marginally significant (p-value is 16.4%). Column (2) Table 32 is analyzed with new client observations only. It shows a significant negative association. In other words, the lower materiality level increases the probability of issuing the qualified opinion. And Column (3) Table 32 is investigated with the continued clients. Based on the results of Column (2) and Column (3), we may say that the audit opinion is more sensitive to the level of the materiality when it is a new client. [26]

[24] This is also similar to the audit fee model and the abnormal audit fee. The abnormal audit fee is measured by using the estimated audit fee model and its residual (Hribar et al. 2014)

[25] For the control variables, we include the variables that we used in Equation (4). We argue that these variables are related to the audit risk. The binary dependent variable is the type of audit opinion.

[26] For the continuing clients, we may still say that the auditors use the materiality level to form the final audit opinion. However, the materiality level does not have marginal effects on the types of the audit opinion (i.e., statistically insignificant). This might be because the disagreement between the auditor and the management is relatively large when the modified opinion is issued to the continuing clients. As a result, the materiality level may not have statistically significance on the types of audit opinion. However,

Table 32 Logistic Regression Results between the materiality level and the type of audit opinions

Variables	Qualified_OPINION_DUMMY (1) Full sample	(2) New client sample	(3) On-going client sample
MA_RESIDUAL	-0.2609	-2.1438**	-0.1242
	(-1.3170)	(-2.1620)	(-0.5826)
NEW_CLIENT	1.1574***		
	(3.8367)		
AUDIT_TENURE	-0.1401***		-0.1486***
	(-4.9001)		(-5.0314)
LOG_TA	0.2793	0.8019	0.2985
	(1.3340)	(1.1799)	(1.3440)
LOG_SALES	-0.2792	-1.7103**	-0.2372
	(-1.4077)	(-2.1914)	(-1.1259)
LOG_ABS_PRETAX_INCOME	0.0568	0.6857	0.0254
	(0.4733)	(1.3918)	(0.2204)
CURR_RATIO	-0.1525	-1.3089**	-0.1342
	(-1.4965)	(-2.4395)	(-1.3573)
DEBT_RATIO	0.0388	0.8500	-0.3703
	(0.0592)	(0.4454)	(-0.5418)
IC_EFFECTIVENSS	0.1685	0.8557	0.1227
	(0.6052)	(1.0211)	(0.4082)
MANAGEMENT_INTERGRITY	-0.4034	0.1687	-0.4617
	(-1.3661)	(0.1930)	(-1.5101)
ROA	-1.9752	1.5217	-3.2146**
	(-1.5604)	(0.4230)	(-2.0290)
SMALL_ABS_ROA_DUMMY	0.0587	-1.3395	0.2542
	(0.1961)	(-0.9929)	(0.8227)
Obs.	843	69	763
Pseudo R^2	0.1351	0.3929	0.1057
INDUSTRY FE	YES	YES	YES
YEAR FE	YES	YES	YES

Standard errors are clustered by client. z-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
Due to the fixed effects, some of the observations are dropped in the subsampling analysis.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

in order to support this argument empirically, additional information (i.e., the amount of the uncorrected misstatements) is necessary.

As an additional analysis, we perform a sub-sample analysis by dividing the sample based on the current year's audit opinion. Interestingly, when auditors issue the qualified opinions, there is no significant difference between new clients and continuing clients with respect to the level of the materiality. However, when auditors issue unqualified opinions, the materiality of new clients is higher than the materiality of the continuing clients (Table 33). [27]

---

[27] Table 33 may suggest that the current audit opinion has an association with the current year's materiality level. An untabulated result shows that the interaction term between the current year audit opinion dummy and the new client indicator has a significant association with the materiality level. And auditing standards require that the materiality level should be determined at the beginning of the auditing procedure. In some sense, the audit fees research is similar to our research. Palmrose (1989) shows that the audit fee with the fixed amount contract type is more prevalent than the audit fee with the cost-reimburse type. In her research, 51% of her samples are the fixed fee contracts. In later research, the portion of the fixed audit fees is more dominant (Thorne [2001] is 59% and Ettredge [2003] is 83% among the total observations). In the fixed audit fee contract, the fee is determined before the final audit opinion. Some of the audit fee research uses the lagged audit opinion as a control variable for controlling the level of client risk when implementing the audit fee model. However, the majority of the audit fee research has used the current year's audit opinion (including going-concern opinion) for the proxy for the client's risk level. Hay et al. (2006) conduct extensive literature review on the audit fee research. They found that 57 literature used the current year auditor's opinion in their audit fee models as an independent variable. Similarly, Table 33 may suggest that the current audit opinion has an association of the current year's materiality level. (In addition, the auditing standards request the auditors to update the materiality level, if it is necessary during the audit procedures. If the updated materiality level information is available, it would be possible to investigate further this relationship between the audit opinion and the materiality level.)

Table 33 Sub-sample Analysis based on the audit opinions

	LOG_MATERIALITY	
Variables	(1) Qualified opinion subsample	(2) Unqualified opinion sample
NEW_CLIENT	-0.0297	0.3710***
	(-0.3452)	(2.7396)
AUDIT_TENURE	-0.0069	0.0131**
	(-0.8205)	(2.0288)
LOG_TA	0.2135***	0.3912***
	(2.6585)	(5.2608)
LOG_SALES	0.6268***	0.3386***
	(5.8749)	(5.1098)
IC_EFFECTIVENSS	0.0999**	0.1636***
	(2.5546)	(3.9674)
ROA	-0.0393	1.3758***
	(-0.0956)	(3.7637)
Obs.	233	610
R-squared	0.6654	0.6298
INDUSTRY FE	YES	YES
YEAR FE	YES	YES

Standard errors are clustered by client. t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
All continuous variables are winsorized at every analysis level at the top and bottom 1 percent.
All variables are defined in Appendix I.

4.7. Implications and conclusion

Due to the difficulty of acquiring actual monetary data about the audit materiality level, few studies in the existing literature have presented empirical evidence (Amiram et al. 2017; Choudhary et al. 2019). In this study, we identify the major determinants of the audit materiality level. Unlike in the previous literature (sparse though it may be), we focus on the mandatory audit setting for private firms under the Spanish legal system. In this context, this research enriches the audit materiality in general as well as under the specific legal environment of mandatory auditing.

This research presents several forms of empirical evidence. First, the auditor sets a higher materiality level with new clients compared to the continuing clients. Second, the materiality level has a significant association with audit tenures. More specifically, the materiality level decreases over the initial seven years of auditing, after which it begins to increase. Third, auditors use the previous year's audit opinion to decide their own materiality level, especially with new clients. Lastly, the materiality level has a negative association with the likelihood of a qualified audit opinion being issued when auditors perform audits with their new clients. These determinants of the materiality level are fundamental, but previous studies have provided little empirical evidence on the materiality determinants themselves.

Our study shows that auditors' behavior of setting materiality with new clients differs from their behavior with the continuing clients. That may imply that auditors have difficulty in properly assessing the risk of new clients, compared to continuing clients. Unlike our results, the prior literature (Choudhary et al. 2019) documents that auditors set lower materiality levels with new clients. The different characteristics of the research samples (i.e., public firms versus private firms) may be the cause of these opposing results. How to approach auditing with a new client while maintaining effectiveness and achieving efficiency remains a subject for further research.

Our study can contribute to the existing body of literature by enriching the empirical evidence in the materiality research. For market participants and regulators, this study provides information about auditors' behaviors in the setting of materiality with new clients. Based on this, market participants and regulators can use our results to inform their decision-making. Furthermore, this study provides useful grounds for the recent arguments about disclosing the materiality level to the public (FRC 2013).

Our dataset contains only one final materiality level for each audit engagement. If more detailed data such as the initial planning materiality, the updated materiality amounts, and the performance materiality amount, become available, the analysis can be more rigorously tested by using these additional data. The materiality can be updated with the audit procedures if necessary (ISA 320 para 12 and 13). It would be another interesting research topic to examine auditors' judgments throughout the audit procedures within an engagement with respect to the materiality level updates; we leave this as a subject of future research.

Chapter 5: Conclusion

In this dissertation, we document that textual analysis techniques can be effectively applied to accounting. In particular, they can be used to predict company's future financial performance or identify accounting misreporting. In the third essay, focusing on a traditional auditing topic, we examine the determinants of materiality levels in auditing. The materiality level is regarded as being closely related to audit quality. In this context, our results suggest that accounting information users need to be especially cautious when using the accounting information provided by a new auditor.

Taking all three essays into account, it is evident that textual analysis can be an effective tool with which information users and auditors may interpret accounting information, both in general and in specific circumstances (i.e., with new clients). With the advancement of new technologies, an increasing number of effective tools will be applicable to accounting and auditing (Moffitt and Vasarhelyi 2013; Brown-Liburd and Vasarhelyi 2015; Vasarhelyi et al. 2015). This research contributes additional possibility to the existing literature.

According to Turney and Pantel (2010), "[c]omputers understand very little of the meaning of human language." To overcome this limitation, NLP has recently been studied extensively. In this dissertation, we attempt to capture more accurate semantic meanings from the textual material, earnings conference call transcripts. To achieve this goal, we take advantage of sentence structure (i.e., the subjects and verbs of the complex hierarchical structure in sentences), and individuals' personal characteristics based on their job titles in their companies. These approaches can help researchers to appreciate textual information more accurately.

In addition, we also introduce some of the NLP models that employ the AI techniques such as the artificial neural networks (i.e., Word2Vec and Dov2Vec). NLP techniques have significant potential, as demonstrated here. Recently, more sophisticated AI methods have been actively combined with NLP tasks. This dissertation contributes to an emergent literature with NLP task in accounting domain, employing textual analysis from simple techniques to up-to-date AI models.

Bibliography

Acito, A. A., J. J. Burks, and W. B. Johnson, 2009. Materiality decisions and the correction of accounting errors. *The Accounting Review*, 84 (3): 659–688.

Akerlof, G. A. 1978. The market for “lemons”: Quality uncertainty and the market mechanism. In Uncertainty in economics, 235–251. Academic Press.

Alpaydin, E. 2014. *Introduction to Machine Learning*. Third edit.

American Institute of Certified Public Accountants (AICPA). 2017. *Audit Guide - Audit Sampling*

Amiram, D., J. Chircop, W. R. Landsman, and K. V. Peasnell. 2017. Mandatorily disclosed materiality thresholds, their determinants, and their association with earnings multiples. *Columbia Business School Research Paper*, 15–69.

Appelbaum, D., and R. A. Nehmer. 2017. Using drones in internal and external audits: An exploratory framework. *Journal of Emerging Technologies in Accounting*, 14 (1): 99–113.

Argamon, S., S. Dhawle, M. Koppel, and J. W. Pennebaker. 2005. Lexical predictors of personality type. In *Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America.* 1-16.

Asthana, S. C., and J. P. Boone. 2012. Abnormal audit fee and audit quality. *Auditing: A Journal of Practice and Theory* 31 (3): 1–22.

Bandura, A. 1986. Social foundation of thought and action: A social-cognitive view. *Englewood Cliffs, NJ*: Prentice-Hall.

Bandura, A. 2000. Cultivate self-efficacy for personal and organizational effectiveness. *Handbook of principles of organization behavior*, 2: 11–21.

Barney, J. 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17 (1): 99–120.

Beneish, M. D. 1999. Incentives and penalties related to earnings overstatements that violate GAAP. *The Accounting Review*, 74 (4): 425–457.

Bengio, Y., R. Ducharme, P. Vincent, and C Jauvin. 2003. A neural probabilistic language model. *Journal of machine learning research*, 3 (Feb): 1137–1155.

Bennett, G. B. and R. C. Hatfield, 2013. The Effect of the Social Mismatch between Staff Auditors and Client Management on the Collection of Audit Evidence. *The Accounting Review*, 88 (1): 31–50.

Blankley, A. I., D. N. Hurtt, and J. E. MacGregor. 2012. Abnormal audit fees and restatements. *Auditing: A Journal of Practice and Theory*, 31 (1): 79–96.

Blau, B. M., J. R. DeLisle, and S. M. Price. 2015. Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. *Journal of Corporate Finance*, 31: 203–219.

Blay, A. D., and M. A. Geiger. 2013. Auditor fees and auditor independence: Evidence from going concern reporting decisions. *Contemporary Accounting Research*, 30 (2): 579–606.

Blokdijk, H., F. Drieenhuizen, D.A. Simunic, and M.T. Stein. 2003. Factors affecting auditors' assessments of planning materiality. *Auditing: A Journal of Practice & Theory* 22 (2): 297–307.

Bond, G. D., and A. Y. Lee. 2005. Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, 19 (3): 313–329.

Bovee, M., A. Kogan, K. Nelson, R. P. Srivastava, and M. A. Vasarhelyi. 2005. Financial reporting and auditing agent with net knowledge (FRAANK) and extensible business reporting language (XBRL). *Journal of Information Systems*, 19 (1): 19–41.

Braun, M. T., L. M. Van Swol, and L. Vang. 2015. His lips are moving: Pinocchio effect and other lexical indicators of political deceptions. *Discourse Processes*, 52 (1): 1–20.

British Council (n.d.). Present perfect. Retrieved January 10, 2020, from https://learnenglish.britishcouncil.org/english-grammar-reference/present-perfect

Buller, D. B., and J. K. Burgoon. 1996. Interpersonal deception theory. *Communication theory*, 6 (3): 203–242.

Burgoon, J., W. J. Mayew, J. S. Giboney, A. C. Elkins, K. Moffitt, B. Dorn, ... and L. Spitzley 2016. Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *Journal of Language and Social Psychology*, 35 (2): 123–157.

Brown-Liburd, H., and M. A. Vasarhelyi. 2015. Big Data and audit evidence. *Journal of Emerging Technologies in Accounting*, 12 (1): 1–16.

Calabrese, R. 2014. Optimal cut-off for rare events and unbalanced misclassification costs. *Journal of Applied Statistics*, *41* (8): 1678–1693.

Carcello, J. V., and A. L. Nagy. 2004. Audit firm tenure and fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 23 (2): 55–69.

Carey, P., and R. Simnett. 2006. Audit partner tenure and audit quality. *The accounting review*, 81 (3): 653–676.

Chen, D., and C. D. Manning. 2014. A fast and accurate dependency parser using neural networks. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP): 740–750.

Cheong, A., S. Cho, W. No, and M. A. Vasarhelyi. 2019. *If You Cannot Measure It, You Cannot Manage It: Assessing the Quality of Cybersecurity Risk Disclosure through Textual Imagification*. Working paper, Rutgers, The State University of New Jersey. Available at SSRN: https://ssrn.com/abstract=3474575

Choi, J. H., J. B. Kim, and Y. Zang. 2010. Do abnormally high audit fees impair audit quality? *Auditing: A Journal of Practice and Theory* 29 (2): 115–140.

Choudhary, P., K. J. Merkley, and K. Schipper. 2019. Auditors' Quantitative Materiality Judgments: Properties and Implications for Financial Reporting Reliability. *Journal of Accounting Research*, 57 (5): 1303–1351.

Craswell, A., D. J. Stokes, and J. Laughton. 2002. Auditor independence and fee dependence. *Journal of Accounting and Economics*, 33 (2): 253–275.

Davis, L. R., B. S. Soo, and G. M. Trompeter. 2009. Auditor tenure and the ability to meet or beat earnings forecasts. *Contemporary Accounting Research*, 26: 517–548.

DeAngelo, L. E. 1981. Auditor independence, 'low balling', and disclosure regulation. *Journal of Accounting and Economics* 3 (2): 113–127.

Dechow, P. M., and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77: 35–59.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28 (1): 17–82.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193–225.

Deis Jr, D. R., and G. A. Giroux. 1992. Determinants of audit quality in the public sector. *Accounting Review*: 462–479.

Eilifsen, A. and W. Messier. 2015. Materiality guidance of the major public accounting firms. *Auditing: A Journal of Practice & Theory* 34 (2): 3–26.

Eshleman, J. D., and P. Guo. 2014. Abnormal audit fees and audit quality: The importance of considering managerial incentives in tests of earnings management. *Auditing: A Journal of Practice and Theory*, 33 (1): 117–138.

Ettredge, M. L., J. C. Bedard, and K. M. Johnstone. 2003. An empirical study of audit effort budgeting dynamics. *School of Business University of Kansas*.

Falato, A., D. Li, and T. Milbourn. 2015. Which skills matter in the market for CEOs? Evidence from pay for CEO credentials.*Management Science* 61 (12): 2845–2869.

Feltz, D. L., and C. D. Lirgg. 1998. Perceived Team and Player Efficacy in Hockey. *Journal of Applied Psychology* 83 (4): 557–564.

Financial Reporting Council (FRC), 2013. International Standard on Auditing (UK and Ireland) (ISA) 700. *The independent auditor's report on financial statements*.

Forbes, D.P., 2005. The effects of strategic decision making on entrepreneurial self-efficacy. *Entrepreneurship Theory and Practice* 29 (5): 599–626.

Francis, J. R. 1984. The effect of audit firm size on audit prices: A study of the Australian market. *Journal of accounting and economics*, 6 (2): 133–151.

Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. Journal of Accounting Research 32 (2): 137–164.

Francis, J. R., and D. T. Simon. 1987. A test of audit pricing in the small-client segment of the US audit market. *Accounting Review*: 145–157.

Francis, J. R., and D. J. Stokes. 1986. Audit prices, product differentiation, and scale economies: Further evidence from the Australian market. *Journal of Accounting Research*: 383–393.

Gensim (n.d.). Doc2Vec Model. Retrieved July 2, 2020, from https://radimrehurek.com/gensim/auto_examples/tutorials/run_doc2vec_lee

Gibson, C. B. 1999. Do they do what they believe they can? Group efficacy and group effectiveness across tasks and cultures. *Academy of Management Journal*, 42 (2): 138–152.

Gist, M. E. 1989. The influence of training method on self-efficacy and idea generation among managers. *Personnel psychology*, 42 (4): 787–805.

Gist, M. E., and T. R. Mitchell. 1992. Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management review*, 17 (2): 183–211.

Gong, Y., J. C. Huang, and J. L. Farh. 2009. Employee learning orientation, transformational leadership, and employee creativity: The mediating role of employee creative self-efficacy. *Academy of management Journal*, 52 (4): 765–778.

Green, T. C., R. Jame, and B. Lock. 2019. Executive extraversion: Career and firm outcomes. *The Accounting Review*, *94* (3): 177–204.

Gully, S. M., A. Joshi, K. A. Incalcaterra, and J. M. Beaubien. 2002. A Meta-Analysis of Team-Efficacy, Potency, and Performance: Interdependence and Level of Analysis as Moderators of Observed Relationships. *Journal of Applied Psychology*, 87 (5): 819–832.

Hauch, V., I. Blandón-Gitlin, J. Masip, and S. L. Sporer. 2015. Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and social psychology Review*, 19 (4): 307–342.

Harrison, A. W., R. K. Rainer Jr., W. A. Hochwarter, and K. R. Thompson. 1997. Testing the Self-Efficacy-Performance-Linkage of Social-Cognitive Theory. *Journal of Social Psychology*, 137 (1): 79–87.

Hay, D. C., W. R. Knechel, and N. Wong. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary accounting research*, 23 (1): 141–191.

Healy, P., J. Wahlen, 1999. A review of the earnings management literature and its implications for standard setting. Accounting Horizons 13: 365–383.

Hennes, K. M., A. J. Leone, and B. P. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review* 83: 1487–1520.

Hermann, M. G. 1999. Assessing leadership style: A trait analysis. Social Science Automation. Available at: https://socialscience.net/docs/LTA.pdf

Hill, T., N. D. Smith, and M. F. Mann. 1987. Role of efficacy expectations in predicting the decision to use advanced technologies: The case of computers. *Journal of applied psychology*, 72 (2): 307.

Hirshleifer, D., A. Low, and S. H. Teoh. 2012. Are confident CEOs better innovators? *Journal of Finance* 67 (4): 1457–1498.

Hmieleski, K. M., and A. C. Corbett. 2008. The contrasting interaction effects of improvisational behavior with entrepreneurial self-efficacy on new venture performance and entrepreneur work satisfaction. *Journal of business venturing*, 23 (4): 482–496.

Hollander, S., M. Pronk, and E. Roelofsen. 2010. Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, 48 (3): 531–563.

Holstrum, G. L., and W. F. Messier, Jr. 1982. A review and integration of empirical research on materiality. *Auditing: A Journal of Practice & Theory* 2 (1): 45–63.

Hope, O. K., and J. C. Langli. 2010. Auditor independence in a private firm and low litigation risk setting. *The Accounting Review*, *85* (2): 573–605.

Hribar, P., T. Kravet, and R. Wilson. 2014. A new measure of accounting quality. *Review of Accounting Studies* 19 (1): 506–538.

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

Instituto de Censores Jurados de Cuentas de España (ICJCE). 2014. GUÍA DE ACTUACIÓN SOBRE IMPORTANCIA RELATIVA O MATERIALIDAD.

International Auditing and Assurance Standards Board (IAASB), 2018a. International Standard on Auditing (ISA) 200, Overall Objectives of the Independent Auditor and the Conduct of an Audit in Accordance with International standards on Auditing, available at: https://www.ifac.org/system/files/publications/files/IAASB-2018-HB-Vol-1.pdf

International Auditing and Assurance Standards Board (IAASB), 2018a. International Standard on Auditing (ISA) 320, Materiality in Planning and Performing an Audit, available at: https://www.ifac.org/system/files/publications/files/IAASB-2018-HB-Vol-1.pdf

International Auditing and Assurance Standards Board (IAASB), 2018b. International Standard on Auditing (ISA) 450, Evaluation of Misstatements Identified During the Audit, available at: https://www.ifac.org/system/files/publications/files/IAASB-2018-HB-Vol-1.pdf

International Auditing and Assurance Standards Board (IAASB), 2018b. International Standard on Auditing (ISA) 700, Forming an Opinion and Reporting on Financial Statements, available at: https://www.ifac.org/system/files/publications/files/IAASB-2018-HB-Vol-1.pdf

Jenkins, D. S., and U. Velury. 2008. Does auditor tenure influence the reporting of conservative earnings. *Journal of Accounting and Public Policy*, 27: 115-132.

John, O. P., and S. Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research, 2: 102–138.

Johnson, V. E., I. K. Khurana, and J. K. Reynolds. 2002. Audit-firm tenure and the quality of financial reports. *Contemporary Accounting Research*, 1: 637–660.

Jones, J. J. 1991. Earnings management during import relief investigations. Journal of Accounting Research 29: 193–228.

Kaplan, S., M. Klebanov, and M. Sorensen. 2012. Which CEO characteristics and abilities matter? Journal of Finance 67 (3): 973–1007.

Kaplan, S. N., and M. Sorensen. 2017. *Are CEOs different? Characteristics of top managers* (No. w23832). National Bureau of Economic Research.

Katz-Navon, T. Y., and M. Erez. 2005. When collective-and self-efficacy affect team performance: The role of task interdependence. *Small group research*, 36 (4): 437–465.

Keila P. S., and D. B. Skillicorn. 2005. Detecting unusual and deceptive communication in email. *In CASCON '05 Proceedings of the 2005 Conference of the Centre for Advanced Studies on Collaborative Research*. IBM Press: 17–20.

Keune, M. B., and K. M. Johnstone. 2012. Materiality judgments and the resolution of detected misstatements: The role of managers, auditors, and audit committees. *The Accounting Review*, 87 (5): 1641–1677.

Kim, D., D. Seo, S. Cho, and P. Kang. 2019. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Information Sciences*, 477: 15-29.

Kipfer, B. 1993. Roget's 21st Century Thesaurus. *New York*, *NY*: Dell Publishing.

Krishnan, J., and J. Krishnan. 1996. The role of economic trade-offs in the audit opinion decision: An empirical analysis. *Journal of Accounting*, *Auditing & Finance*, 11 (4): 565–586.

Lancaster, G. L., A. Vrij, L. Hope, and B. Waller. 2013. Sorting the liars from the truth tellers: The benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology*, 27 (1): 107–114.

Larcker, D. F., and A. A. Zakolyukina. 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50 (2): 495–540.

Le, Q., and T. Mikolov. 2014. Distributed representations of sentences and documents. In International conference on machine learning: 1188–1196.

Lee, J. 2016. Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review*, *91* (1): 229–250.

Lent, R. W., S. D. Brown, and K. C. Larkin. 1987. Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking. *Journal of counseling psychology*, 34 (3): 293.

Lester, S. W., B. M. Meglino, and M. A. Korsgaard. 2002. The antecedents and consequences of group potency: A longitudinal investigation of newly formed work groups. *Academy of Management Journal*, 45 (2): 352–368.

Levitt, A. 1998. Remarks by Chairman Arthur Levitt of Securities and Exchange Commission, 'The Number Game'. NYU Center for Law and Business, New York, NY. available at: https://www.sec.gov/news/speech/speecharchive/1998/spch220.txt

Li, C. 2009. Does client importance affect auditor independence at the office level? Empirical evidence from going-concern opinions. *Contemporary Accounting Research*, *26* (1): 201–230.

Li F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3): 221–247.

Li F. 2010a. Textual analysis of corporate disclosures: a survey of the literature. *Journal of Accounting Literature* 29: 143–165.

Li, F. 2010b. The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research*, 48 (5): 1049–1102.

Li, Q., and J. Liu. 2020. Development of An Intelligent NLP-Based Audit Plan Knowledge Discovery System. *Journal of Emerging Technologies in Accounting*, 17 (1): 89–97.

Liu, Y., and K. C. Moffitt. 2016. Text mining to uncover the intensity of SEC comment letters and its association with the probability of 10-K restatement. *Journal of Emerging Technologies in Accounting*, 13 (1): 85–94.

Lindsley, D. H., D. J. Brass, and J. B. Thomas. 1995. Efficacy-performing spirals: A multilevel perspective. *Academy of management review,* 20 (3), 645–678.

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

Loughran T., and B. McDonald. 2014. Measuring readability in financial disclosures measuring readability in financial disclosures. *The Journal of Finance* 69 (4): 1643–1671.

Loughran, T., and B. McDonald 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research,* 54 (4): 1187–1230.

Luthans, F., and S. J. Peterson. 2002. Employee engagement and manager self-efficacy. *Journal of Management Development*, 21 (5): 376–387.

Lobo, G. J., and Y. Zhao. 2013. Relation between audit effort and financial report misstatements: Evidence from quarterly and annual restatements. *The Accounting Review*, 88 (4): 1385–1412.

Mairesse, F., M. A. Walker, M. R. Mehl, and R. K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research* 30: 457–500. https://doi.org/10.1613/jair.2349

Malmendier, U., and G. Tate. 2005. CEO overconfidence and corporate investment. *Journal of Finance* 60 (6): 2661–2700.

Malmendier, U., and G. Tate. 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics* 89 (1): 20–43.

Marzban, C. (2004). The ROC curve and the area under it as performance measures. *Weather and Forecasting*, *19* (6), 1106–1114.

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

Mayew, W. J. 2008. Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research* 46 (3): 627–659.

Mayew, W., C. Parsons, and M. Venkatachalam. 2013. CEO voice pitch predicts labor market success. *Evolution and Human Behavior* 34: 243–248.

Merkley, K., R. Michaely, and J. Pacelli. 2017. Does the scope of the sell-side analyst industry matter? An examination of bias, accuracy, and information content of analyst reports. *The Journal of Finance*, 72 (3): 1285–1334.

Messier, W. F. Jr., N. Martinov, and A. Eilifsen, 2005. A review and integration of empirical research on materiality: Two decades later. *Auditing: A Journal of Practice & Theory* 24 (2): 153–187.

Mikolov, T., W. T. Yih, and G. Zweig. 2013a. Linguistic regularities in continuous space word representations. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: *Human language technologies*: 746–751.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013b. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems: 3111–3119.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013c. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mitchell, T. R., H. Hopper, D. Daniels, J. George-Falvy, and L. R. James. 1994. Predicting self-efficacy and performance during skill acquisition. *Journal of Applied Psychology*, 79: 506 –517.

Moffitt, K. C., and M. A. Vasarhelyi. 2013 AIS in an Age of Big Data. *Journal of Information Systems*, 27 (2): 1 –19.

Moores, T. T., and J. C. J. Chang 2009. Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. *Information & Management*, *46* (2): 69–76.

Nelson, M. W., J. A. Elliott, and R. L. Tarpley. 2002. Evidence from auditors about managers' and auditors' earnings management decisions. *The accounting review*, 77 (s-1): 175–202.

Nelson, K. M., A. Kogan, R. P. Srivastava, M. A. Vasarhelyi, and H. Lu. 2000. Virtual auditing agents: the EDGAR Agent challenge. *Decision Support Systems*, 28 (3): 241–253.

Oberlander, J., and S. Nowson, 2006. Whose thumb is it anyway? Classifying author personality from weblog text. In Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL).

Palmrose, Z. V. 1986. Audit fees and auditor size: Further evidence. *Journal of accounting research*: 97–110.

Penrose, E. T. 1959. The theory of the growth of the Firm, *Wiley*. New York.

Perols, J. L., R. M. Bowen, C. Zimmermann, and B. Samba. 2017. Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review*, 92 (2): 221–245.

Peterson, E., T. R. Mitchell, L. Thompson, and R. Burr. 2000. Collective Efficacy and Aspects of Shared Mental Models as Predictors of Performance over Time in Work Groups. *Group Processes & Intergroup Relations*, 3 (3): 296–316.

Price III, R. A., N. Y. Sharp, and D. A. Wood. 2011. Detecting and predicting accounting irregularities: A comparison of commercial and academic risk measures. *Accounting Horizons*, *25* (4): 755–780.

Prussia, G. E., and A. J. Kinicki. 1996. A Motivational Investigation of Group Effectiveness Using Social -- Cognitive Theory. *Journal of Applied Psychology*, 81 (2): 187–198.

Purda, L., and D. Skillicorn. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 32 (3): 1193–1223.

Public Company Accounting Oversight Board (PCAOB), 2010a. Auditing Standard (AS) 1101, Audit Risk

Public Company Accounting Oversight Board (PCAOB), 2010a. Auditing Standard (AS) 2105, Consideration of Materiality in Planning and Performing an Audit

Public Company Accounting Oversight Board (PCAOB), 2010b. Auditing Standard (AS) 2810, Evaluating Audit Result

Public Company Accounting Oversight Board (PCAOB), 2010b. Auditing Standard (AS) 3105, Departures from Unqualified Opinions and Other Reporting Circumstances

Rashid, T. 2016. Make your own neural network. CreateSpace Independent Publishing Platform.

Rehurek, R., and P. Sojka. 2010. Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

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

Rowley-Jolivet, E., and S. Carter-Thomas. 2005. Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting. English for Specific Purposes 24: 41–64.

Shah, A. 2019. ASDOC: Stata module to create high-quality tables in MS Word from Stata output.

Securities and Exchange Commission (SEC), 1999. Staff Accounting Bulletin (SAB) 99, Materiality

Simunic, D. A. 1980. The pricing of audit services: Theory and evidence. *Journal of accounting research*: 161–190.

Shin, D., S. He, G. M. Lee, A. B. Whinston, S. Cetintas, and K. C. Lee. 2019. Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. *Forthcoming at MIS Quarterly*.

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

Steinbart, P. J., 1987. The construction of a rule-based expert system as a method for studying materiality judgments. *The Accounting Review* 62 (1): 97–116.

Stumpf, S. A., A. P. Brief, and K. Hartman. 1987. Self-efficacy expectations and coping with career-related events. *Journal of Vocational Behavior*, 31 (1): 91–108.

Taylor, M. S., E. A. Locke, C. Lee, and M. E. Gist. 1984. Type A behavior and faculty research productivity: What are the mechanisms?. *Organizational Behavior and Human Performance*, 34 (3): 402–418.

Tepalagul, N., and L. Lin. 2015. Auditor independence and audit quality: A literature review. *Journal of Accounting, Auditing & Finance*, 30 (1): 101–121.

Tetlock, P. C. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, 62 (3): 1139–1168.

Thorne, J., S. A. Holmes, A. S. McGowan, C. A. Strand, and R. H. Strawser. 2001. The relation between audit pricing and audit contract type: A public sector analysis. *Journal of Accounting and Public Policy*, 20 (3): 189–215.

Turney, P. D., and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, *37*: 141–188.

Vancouver, J. B., C. M. Thompson, E. C. Tischner, and D. J. Putka. 2002. Two studies examining the negative effect of self-efficacy on performance. *Journal of applied psychology, 87* (3): 506.

Vasarhelyi, M. A., and F. B. Halper. 1991. The continuous audit of online systems. *Auditing: A Journal of Practice and Theory*.

Vasarhelyi, M. A., A. Kogan, and B. M. Tuttle. 2015. Big Data in accounting: An overview. *Accounting Horizons*, 29 (2): 381–396.

Vrij, A., K. Edward, K. P. Roberts, and R. Bull. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal behavior*, *24* (4): 239–263.

Vrij, A. 2004. Why professionals fail to catch liars and how they can improve. *Legal and criminological psychology*, *9* (2): 159–181.

Wernerfelt, B. 1984. A resource-based view of the firm. *Strategic Management Journal*, 5 (2) : 171–180.

Wood, R. E., and A. Bandura. 1989. Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of Personality and Social Psychology*, 56 : 407–415.

Young, T., D. Hazarika, S. Poria, and E. Cambria. 2018. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, *13* (3) : 55–75.

Zhaokai, Y., and K. C. Moffitt. 2019. Contract analytics in auditing. *Accounting Horizons*, 33 (3): 111–126.

Zhou, L., J. K. Burgoon, J. F. Nunamaker, and D. Twitchell. 2004. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation*, 13 (1): 81–106.

Appendices

Appendix A: Development of the new word list for measuring the management self-efficacy

We develop the new word list through the validation process. We name it as *'KevinNBen'* word list. These words work effectively for capturing the management self-efficacy statements.

approach / approaches / approached / approaching / capacity / capacities /
complete / completed /grow / grows / grew / growing / grown / growth / growths /
milestone / milestones / ready / patent / patents / quality / qualities /
ability / abilities / able / capability / capabilities / capable

Appendix B: Validation process for refining the rules

We iterated three times to refine the rules for identifying the management self-efficacy statements. Each process is carried out with randomly selected 200 management sentences. We applied our initial rules and obtained the confusion matrix that shows the classification results with those 200 sentences. Table 34 shows this confusion matrix.

Table 34 Classification results with the first 200 sentences (confusion matrix)

	Predicted as self-efficacy sentences	**Predicted as NOT self-efficacy sentences**	**Total**
Actual: Self-efficacy Sentences	17	6	23
Actual: Not Self-efficacy Sentences	43	134	177
Total	60	140	200

Overall accuracy: 75.5% = ([17 + 134] / 200)

Based on this result, we improved the initial rules by modifying the initial rules. Table 34 shows a summary of the rule updates. Table 35 shows the improved results with the application of the updated rule.

Table 35 Summary of rule updating with the first 200 sentences

Word list	Type of action	Contents
KevinNBen Word list	Adding new words	grow / grows / grew / growing / grown / growth / growths / approach / approaches / approached / approaching / ready / patent / patents
LIWC	Removing existing words	best / better / earn* (only excluding when the full word is 'earning' or 'earnings') / first / importan* / plan / produc* / requir* / responsib* / strength* / strong*
Synonym words of capability	Removing existing words	responsible / strong / strength / strengths / stronger / strongest / strongly / means (only excluding when it is used as verb)

Table 36 Classification results with the first 200 sentences with the updated rule

	Predicted as self-efficacy sentences	Predicted as NOT self-efficacy sentences	Total
Actual: Self-efficacy Sentences	22	1	23
Actual: Not Self-efficacy Sentences	31	146	177
Total	53	147	200

Overall accuracy: 84.0% = ([22 + 146] / 200)

We iterated the same process again with another randomly selected 200 sentences. We improved the updated rule again. Table 17 shows this confusion matrix with these second 200 sentences. Table 18 shows the summary of the second rule refinement. Table 19 shows the improved results with the application of the second updated rule.

Table 37 Classification results with the second 200 sentences (confusion matrix)

	Predicted as self-efficacy sentences	Predicted as NOT self-efficacy sentences	Total
Actual: Self-efficacy Sentences	24	11	35
Actual: Not Self-efficacy Sentences	41	124	165
Total	65	135	200

Overall accuracy: 74.0% = ([24 + 124] / 200)

Table 38 Summary of rule updating with the second 200 sentences

Word list	Type of action	Contents
KevinNBen Word list	Adding new words	capacity / capacities / complete / completed / quality / qualities / milestone / milestones
LIWC	Removing existing words	conclud* / conclus* / perform* (only excluding when the full word is 'performance') / working* (only excluding when the full words are 'working capital')
Synonym words of capability	Removing existing words	might (only excluding when it is used as modal)

Table 39 Classification results with the second 200 sentences with the second updated rule

	Predicted as self-efficacy sentences	Predicted as NOT self-efficacy sentences	Total
Actual: Self-efficacy Sentences	28	7	35
Actual: Not Self-efficacy Sentences	37	128	165
Total	65	135	200

Overall accuracy: 78.0% = ([28 + 128] / 200)

We randomly select another 200 sentences. Table 20, Table 21, and Table 21 show the classification results with the initial rule, the second rule, and the final rule, respectively. The overall accuracy has been improving along with the rule updates. We finalize our rule with this stage. Table 22 shows our final rule.

Table 40 Classification results with the third 200 sentences with the initial rule

	Predicted as self-efficacy sentences	**Predicted as NOT self-efficacy sentences**	**Total**
Actual: Self-efficacy Sentences	26	9	35
Actual: Not Self-efficacy Sentences	43	122	165
Total	69	131	200

Overall accuracy: 74.0% = ([26 + 122] / 200)

Table 41 Classification results with the third 200 sentences with the updated rule

	Predicted as self-efficacy sentences	**Predicted as NOT self-efficacy sentences**	**Total**
Actual: Self-efficacy Sentences	27	8	35
Actual: Not Self-efficacy Sentences	40	125	165
Total	67	133	200

Overall accuracy: 76.0% = ([27 + 125] / 200)

Table 42 Classification results with the third 200 sentences with the final rule

	Predicted as self-efficacy sentences	**Predicted as NOT self-efficacy sentences**	**Total**
Actual: Self-efficacy Sentences	27	8	35
Actual: Not Self-efficacy Sentences	39	126	165
Total	66	134	200

Overall accuracy: 76.5% (= [27 + 126] / 200)

Appendix C: Analyzing sentence structure

We use the Stanford CoreNLP package for analyzing the sentence structures. In this section, we illustrate the example of sentence analyzing. Those sentences are two examples of the management statement during the earnings conference calls.[28]

"I will" sentence

I'm confident that the intensive defining and preparation efforts **we** conducted during 2014 **will** lead to success for the company at both projects as we enter into construction phases during 2015.

```
"(ROOT
 (S
  (NP (PRP I))
  (VP (VBP am)
   (ADJP (JJ confident)
    (SBAR (IN that)
     (S
      (NP
       (NP (DT the)
        (ADJP (JJ intensive))
        (VBG defining))
       (CC and)
       (NP
        (NP (NN preparation) (NNS efforts))
        (SBAR
         (S
          (NP (PRP we))
          (VP (VBD conducted)
           (PP (IN during)
            (NP (CD 2014))))))))
      (VP (MD will)
       (VP (VB lead)
        (PP (TO to)
         (NP
          (NP (NN success))
          (PP (IN for)
           (NP
            (NP (DT the) (NN company))
            (PP (IN at)
             (NP (DT both) (NNS projects)))))))
        (SBAR (IN as)
         (S
          (NP (PRP we))
          (VP (VBP enter)
           (PP (IN into)
            (NP (NN construction) (NNS phases)))
           (PP (IN during)
            (NP (CD 2015)))))))))))
  (. .)))"
```

[28] The conference call transcript is provided by seekingAlpha.com (https://seekingalpha.com/)

"I have done" sentence

As we mentioned before, achieving **our** intermediate objective of 6.5 EBITDA target which was targeted for year end 2017 **has been** accelerated and we would certainly expect continued EBITDA improvements through both internal growth and development lease-up as well as reaping the benefits of a higher quality portfolio.

```
"(ROOT
 (S
  (SBAR (IN As)
   (S
    (NP (PRP we))
    (VP (VBD mentioned)
     (ADVP (RB before)))))
  (, ,)
  (S
   (S
    (VP (VBG achieving)
     (NP
      (NP (PRP$ our) (JJ intermediate) (NN objective))
      (PP (IN of)
       (NP
        (NP (CD 6.5) (NN EBITDA) (NN target))
        (SBAR
         (WHNP (WDT which))
         (S
          (VP (VBD was)
           (VP (VBN targeted)
            (PP (IN for)
             (NP (NN year) (NN end) (CD 2017)))))))))))
   (VP (VBZ has)
    (VP (VBN been)
     (VP (VBN accelerated)))))
  (CC and)
  (S
   (NP (PRP we))
   (VP (MD would)
    (ADVP (RB certainly))
    (VP
     (VP (VB expect)
      (NP (JJ continued) (NN EBITDA) (NNS improvements))
      (PP (IN through)
       (NP (DT both) (JJ internal) (NN growth)
        (CC and)
        (NN development) (NN lease-up))))
     (CONJP (RB as) (RB well) (IN as))
     (VP (VBG reaping)
      (NP
       (NP (DT the) (NNS benefits))
       (PP (IN of)
        (NP (DT a)
         (ADJP (JJR higher) (NN quality))
         (NN portfolio)))))))))
  (. .)))"
```

Appendix D: Variable Descriptions in Chapter 2

Variable (Source)	Definition
SELF_EFFICACY_PT (SeekingAlpha.com)	Measured level of management self-efficacy in the presentation session (= the number of management self-efficacy sentences in the presentation session / the number of total sentences in the presentation session)
SELF_EFFICACY_QNA (SeekingAlpha.com)	Measured level of management self-efficacy in the Q&A session (= the number of management self-efficacy sentences in the Q&A session / the number of total sentences in the Q&A session)
SELF_EFFICACY_ BOTH (SeekingAlpha.com)	Measured level of management self-efficacy in both the presentation session and the Q&A session (= the number of management self-efficacy sentences in a call / the number of total sentences in a call)
SELF_EFFICACY_DIFF (SeekingAlpha.com)	Adjusted level of management self-efficacy by using the difference between two session (= SELF_EFFICACY_QNA - SELF_EFFICACY_PT)
ROA (Compustat)	Return on Assets (net-income divided by the total assets at the end of the fiscal year)
D_ROA_t (Compustat)	Change of ROA at year t (i.e., the difference between the ROA of the current year (year t) and the ROA of the last year (year t-1), = ROA_t - ROA_{t-1})
$CONTI\_D\_ROA_{it+2}$ (Compustat)	Dummy variable for the ROA increasing trend (1 if Both D_ROA_{t+1} and D_ROA_{t+2} are greater than zero, 0 otherwise)
Big_4_Auditor (Compustat)	1 when the external auditor is one of the big 4 auditors, 0 otherwise
LOG_T_ASSETS (Compustat)	Natural logarithm of the total assets at the end of the fiscal year
DEBT_RATIO (Compustat)	Debt ratio (Total liability / Total assets at the end of the fiscal year)
CUR_RATIO (Compustat)	Current ratio (= Current assets / Current liability at the fiscal year end year)
RET_VOL (CRSP)	The stock return volatility of past 12 months based on the fiscal year end
ROA_VOL (Compustat)	The ROA volatility of the past 5 years
ABS_T_ACCRUALS (Compustat)	Absolute value of the total accruals (The difference between net income and the cash flow from the operation activity scaled by the total asset of the fiscal year end)
BTM(Compustat)	Book to market ratio at the end of the fiscal year (= [total assets – total liability] / capital market value)
D_DIVIDEND (Compustat)	1 if there was a cash dividend during the fiscal year, 0 otherwise
SEG_CNT (Compustat)	Number of businesss segments. If the nuber is not provided by the Compustat, the value of 1 is assigned.
FIRM_AGE (Compustat)	Firm age (The number of years between the first observation fiscal year in Compustat and the year of the conference call fiscal year)

SELF_EFFICACY_PT_Excluding_I_my	Measured level of management self-efficacy in the presentation session by using the first person pronouns we and our, and not including 'I' and 'my' pronoun. Additionally, 'the company' is used as a reference to themselves in a sentence.
LM_POSITIVE_PT	Positive sentence ratio in the presentation session (A positive sentence is defined as a sentence that has any of positive words [Loughran and McDonald, 2011] and does not have any of negative words [Loughran and McDonald, 2011])
LM_NEGATIVE_PT	Negative sentence ratio in the presentation session (A negative sentence is defined as a sentence that has any of negative words [Loughran and McDonald, 2011] and does not have any of positive words [Loughran and McDonald, 2011])
Word2Vec_CBOW_25_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec CBOW 25 word list
Word2Vec_CBOW_50_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec CBOW 50 word list
Word2Vec_CBOW_100_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec CBOW 100 word list
Word2Vec_CBOW_200_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec CBOW 200 word list
Word2Vec_SG_25_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec SG 25 word list
Word2Vec_SG_50_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec SG 50 word list
Word2Vec_SG_100_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec SG 100 word list
Word2Vec_SG_200_PT	Measured level of management self-efficacy in the presentation session by replacing the KevinNBen word list with Word2Vec SG 200 word list

Appendix E: Management self-efficacy word list from the Word2Vec method

We build word lists by deploying the Word2vec method. The words are identified by using six core efficacy words— *capability, capabilities, ability, abilities, able*, and *capable*. We iterate the 10 times with the same conference call transcripts by changing the seed value from 1 to 10. A word located in the earlier position in each word list is the word that is identified as closer to those core efficacy words by the Word2Vec model. We include any words if the words included in the any of 10 iterations.

The actual implementation is performed based on Python Gensim (version 3.8.0).[29] The closest words are identified by using Gensim's 'most_similar' function (i.e., cosine similarity between two words).[30]

We include the words that are used more than 50 times in the entire conference call transcripts. This threshold eliminates the words with typos and less frequently used words. Some words are closed to the efficacy concepts. However, they are not included because of our 50 threshold criteria (e.g., the word 'skillful' with only 13 times, 'mastery' with only 10 times, 'resourcefulness' with only 10 times, etc.).[31] The total number of words in our entire conference call sample is more than 106 million words.

Based on our analysis, the Skip-gram 200 word list have a better performance with the most statistical significance. (Original KevinNBen words are included for the comparison purpose.)

Word Category	**Words**	**Total frequency in the transcripts**
KevinNBen (22 words)	approach, approaches, approached, approaching, capacity, capacities, complete, completed, grow, grows, grew, growing, grown, growth, growths, milestone, milestones, ready, patent, patents, quality, and qualities	580,079
Word2Vec #1 CBOW (25 words)	expertise, skills, agility, credentials, flexibility, desire, resources, knowhow, competencies, strengths, processes, tools, talent, versatility, algorithms, specialization, scalability, solutions, talents, functionality, credibility, knowledge, commitment, vulnerabilities, infrastructures	98,891

[29] The detail hyper parameter are followings: for CBOW method, we set window size as 2, minimum word count as 10. For Skip-Gram method, we set window size as 4, minimum word count as 10. The rest of the setting is assigned the default setting by the Gensim model.
[30] Using Euclidean distance can be an alternative for the identification of the closest words.
[31] Also the unique entity names (i.e., company names) are removed in the word list.

Word Category	Words	Total frequency in the transcripts
Word2Vec #2 CBOW (50 words)	expertise, skills, agility, credentials, flexibility, desire, resources, knowhow, competencies, strengths, processes, tools, talent, versatility, algorithms, specialization, scalability, solutions, talents, functionality, credibility, knowledge, commitment, vulnerabilities, infrastructures, capacities, willingness, technology, freedom, ecosystems, platform, mission, teams, backbone, wherewithal, workflow, autonomy, creativity, continuity, relevance, efforts, relevancy, needs, roadmaps, relationships, infrastructure, designed, prowess, profitably, skill	270,734
Word2Vec #3 CBOW (100 words)	expertise, skills, agility, credentials, flexibility, desire, resources, knowhow, competencies, strengths, processes, tools, talent, versatility, algorithms, specialization, scalability, solutions, talents, functionality, credibility, knowledge, commitment, vulnerabilities, infrastructures, capacities, willingness, technology, freedom, ecosystems, platform, mission, teams, backbone, wherewithal, workflow, autonomy, creativity, continuity, relevance, efforts, relevancy, needs, roadmaps, relationships, infrastructure, designed, prowess, profitably, skill, workflows, platforms, responsiveness, aim, technologies, dashboards, inability, interfaces, recipes, optionality, ingenuity, adaptability, sophistication, cultures, intelligently, qualities, vision, culture, backgrounds, offerings, solution, advantages, inventions, strategy, efficiencies, aims, freshness, opportunities, protocols, resiliency, passion, technicians, techniques, flexibly, innovations, footprint, architectures, reputation, workforces, ultracapacitors, competency, designers, intent, competitiveness, functionally, innovate, specifications, firepower, engineers, trying	498,562
Word2Vec #4 CBOW (200 words)	expertise, skills, agility, credentials, flexibility, desire, resources, knowhow, competencies, strengths, processes, tools, talent, versatility, algorithms, specialization, scalability, solutions, talents, functionality, credibility, knowledge, commitment, vulnerabilities, infrastructures, capacities, willingness, technology, freedom, ecosystems, platform, mission, teams, backbone, wherewithal, workflow, autonomy, creativity, continuity, relevance, efforts, relevancy, needs, roadmaps, relationships, infrastructure, designed, prowess, profitably, skill, workflows, platforms, responsiveness, aim, technologies, dashboards, inability, interfaces, recipes, optionality, ingenuity, adaptability, sophistication, cultures, intelligently, qualities, vision, culture, backgrounds, offerings, solution, advantages, inventions, strategy, efficiencies, aims, freshness, opportunities, protocols, resiliency, passion, technicians, techniques, flexibly, innovations, footprint, architectures, reputation, workforces, ultracapacitors, competency, designers, intent, competitiveness, functionally, innovate, specifications, firepower, engineers, trying, effectiveness, muscles, methodologies, restaurateurs, creditability, missions, ambition, superiority, interoperable, fluidity, simplicity, equipped, foundation, interactivity, capacity, assortments, reliability, disciplines, responsibly, proficiency, foundations, empowered, breadth, resilience, interconnects, leadership, patented, methods, vendors, innovation, opportunity, need, efficiently, jointly, innovative, competence, seamlessly, organization, database, breath, silos, audiences, scale, partners, workload, approaches, excellence, features, attempting, usability, content, stickiness, analytics, motivation, destiny, scalable, strategies, dedicated, aspiring, agile, dedication, networks, objective, haptics, organizations, marketers, ways, systems, images, try, professionalism, architecture, effort, resourcing, uniqueness, diligently, functionalities, needed, position, insights, necessary, practices, optimized, proprietary, enzymes, tasks, uniformity, eager, unparalleled, customers, helping, positioned, functions, personalization, experiences, products, fanatical, designs, wanting, differentiated	1,200,355
Word2Vec #5 Skip-Gram (25 words)	expertise, knowhow, prowess, flexibly, skills, competencies, scalability, agility, platform, wherewithal, adaptability, flexibility, knowledge, scalable, versatility, unparalleled, enables, tools, autonomy, scale, unmatched, enable, technology, ecosystems, intimacy	131,966
Word2Vec #6 Skip-Gram (50 words)	expertise, knowhow, prowess, flexibly, skills, competencies, scalability, agility, platform, wherewithal, adaptability, flexibility, knowledge, scalable, versatility, unparalleled, enables, tools, autonomy, scale, unmatched, enable, technology, ecosystems, intimacy, decisioning, processes, unrivaled, enabling, strengths, unparallel, competence, solutions, talent, enabled, sophistication, differentiated, talents, breadth, proficiency, competency, efficiently, resources, firepower, capacities, bioinformatics, optimized, adaptable, advantages, nurture	205,868

Word Category	Words	Total frequency in the transcripts
Word2Vec #7 Skip-Gram (100 words)	expertise, knowhow, prowess, flexibly, skills, competencies, scalability, agility, platform, wherewithal, adaptability, flexibility, knowledge, scalable, versatility, unparalleled, enables, tools, autonomy, scale, unmatched, enable, technology, ecosystems, intimacy, decisioning, processes, unrivaled, enabling, strengths, unparallel, competence, solutions, talent, enabled, sophistication, differentiated, talents, breadth, proficiency, competency, efficiently, resources, firepower, capacities, bioinformatics, optimized, adaptable, advantages, nurture, possesses, credentials, defensible, proven, acumen, intelligently, functionality, platforms, solution, responsiveness, roadmaps, uniquely, fortify, continuity, optimally, leverages, empower, analytic, specialization, empowering, algorithms, qualities, productively, profitably, leveraging, skill, exploit, interfaces, position, empowered, fullest, workflows, incented, optionality, dedicated, strives, efficiencies, automate, superior, credibility, arsenal, workforces, aim, unify, innovate, designed, seamlessly, ensure, incumbency, deliverability	333,396
Word2Vec #8 Skip-Gram (200 words)	expertise, knowhow, prowess, flexibly, skills, competencies, scalability, agility, platform, wherewithal, adaptability, flexibility, knowledge, scalable, versatility, unparalleled, enables, tools, autonomy, scale, unmatched, enable, technology, ecosystems, intimacy, decisioning, processes, unrivaled, enabling, strengths, unparallel, competence, solutions, talent, enabled, sophistication, differentiated, talents, breadth, proficiency, competency, efficiently, resources, firepower, capacities, bioinformatics, optimized, adaptable, advantages, nurture, possesses, credentials, defensible, proven, acumen, intelligently, functionality, platforms, solution, responsiveness, roadmaps, uniquely, fortify, continuity, optimally, leverages, empower, analytic, specialization, empowering, algorithms, qualities, productively, profitably, leveraging, skill, exploit, interfaces, position, empowered, fullest, workflows, incented, optionality, dedicated, strives, efficiencies, automate, superior, credibility, arsenal, workforces, aim, unify, innovate, designed, seamlessly, ensure, incumbency, deliverability, backbone, equipped, enrich, harness, enhance, cushions, allows, frameworks, fourthly, toolbox, infrastructures, needed, desire, proprietary, creatively, creditability, repository, effectively, possess, innovative, equip, innovated, inventions, allowing, backgrounds, computational, globalizing, unleash, functionalities, caliber, ingenuity, teams, enlarge, wisely, utilizing, technologies, secure, localize, offerings, flexible, reputation, effectiveness, vulnerabilities, creativity, commitment, globalize, fanatical, retool, multiproduct, workflow, proposition, utilize, primed, help, programmable, commonality, enhanced, automated, infrastructure, integrated, optimizes, resiliency, resourcing, analytics, need, operationalize, energize, customized, needs, unique, reengineered, enhances, resourced, constructs, uniformity, vertically, reliably, culture, robustness, network, complimentary, retooled, capacity, modernized, footprint, allow, assimilate, helping, sophisticated, ensuring, responsibly, relationships, customizable, grasp, cultures, demands, analytical, valuable, interoperable, adapt	712,455

Appendix F: Variable Descriptions in Chapter 3

Variable (Source)	Definition
RESTATEMENT (AuditAnalytics)	1 if there is a restatement in the financial statements in that specific year, 0 otherwise (Fraud related restatements, GAAP application failure, and errors related restatement categories based on AuditAnalytics classification)
Negative sentence (SeekingAlpha.com)	A sentence that has one of negative words (Loughran and McDonald, 2011) and does not have any of positive words (Loughran and McDonald, 2011).
Negative sentence ratio (SeekingAlpha.com)	The number of negative sentences / the total number of sentences
NEGATIVE _CHANGE_CEO_PT (SeekingAlpha.com)	Change of negative sentence ratio in the presentation session in the CEO's comment between the current year and the prior year in the earnings conference call (= Negative sentence ratio of current year - Negative sentence ratio of prior year) The CEO needs to be the same person across two years.
NEGATIVE_CHANGE_CFO_PT (SeekingAlpha.com)	Change of negative sentence ratio in the presentation session in the CFO's comment between the current year and the prior year (The CFO needs to be the same person across two years.)
NEGATIVE _CHANGE_CEO_QNA (SeekingAlpha.com)	Change of negative sentence ratio in the Q&A session in the CEO's comment between the current year and the prior year (The CEO needs to be the same person across two years.)
NEGATIVE _CHANGE_CFO_QNA (SeekingAlpha.com)	Change of negative sentence ratio in the Q&A session in the CFO's comment between the current year and the prior year (The CFO needs to be the same person across two years.)
NEGATIVE_CHANGE_CEO_CFO_PT (SeekingAlpha.com)	Negative tone change difference between the CEO and the CFO in the presentation session (= NEGATIVE_PT_CHANGE_CEO - NEGATIVE_PT_CHANGE_CFO)
NEGATIVE_CHANGE _CEO_CFO_QNA (SeekingAlpha.com)	Negative tone change difference between the CEO and the CFO in the Q&A session (= NEGATIVE_QNA_CHANGE_CEO - NEGATIVE_QNA_CHANGE_CFO)
SELF_EFFICACY_CHANGE_CEO_CFO_P T (SeekingAlpha.com)	Management Self-efficacy tone change difference between the CEO and the CFO in the presentation session

SELF_EFFICACY_CHANGE_CEO_CFO_QNA (SeekingAlpha.com)	Management Self-efficacy tone change difference between the CEO and the CFO in the Q&A session
BIG4 (Compustat)	1 when the external auditor is one of the big 4 auditors, 0 otherwise
LOG_TA (Compustat)	Natural logarithm of the total assets at the end of the fiscal year
ROA (Compustat)	Return on Assets (= Net-income / the total assets at the end of the fiscal year)
DEBT RATIO (Compustat)	Debt ratio (= Total liability / Total assets at the end of the fiscal year)
CURRENT RATIO (Compustat)	Current ratio (= Current assets / Current liability at the fiscal year end year)
T ACCRUALS (Compustat)	Total accruals (= [Net-income - Cash flow from the operation activity] / Total asset of the fiscal year end)
ABNORMAL AUDIT FEE (AuditAnalytics and Compustat)	Residual from the estimated audit fee model (Price et al. 2011)
BTM (Compustat)	Book to market ratio at the end of the fiscal year (=[total assets – total liability] / capital market value)

Appendix G: Audit fee models and OLS regression result

In order to include as many observation as possible, we use the audit fee model in the prior literature (Price et al. 2011). Their model omits some variables while preserving explanatory power of the audit fee model (Price et al. 2011). The followings are the audit fee model and its regression result.

$$\begin{aligned} LogAuditFees = {} & IndustryIndicators + \beta_1\, BigN + \beta_2 \log(Assets) \\ & + \beta_3\, Inventory / AvgAssets + \beta_4\, Receivables / AvgAssets \\ & + \beta_5\, LTD / AvgAssets + \beta_6\, Earn / AvgAssets \\ & + \beta_7\, Loss + \beta_8\, Qualified + \beta_9\, AuditorTenure + \varepsilon \end{aligned}$$

Variables	Log_AuditFee
BigN	0.5256***
	(81.1519)
LOG_TOTAL_ASSET	0.4426***
	(359.9050)
INVENTORY_RATIO	0.0605***
	(2.6008)
RECEIVABLE_RATIO	-0.0532***
	(-3.2498)
LTD_RATIO	0.0004***
	(3.8138)
EARN_RATIO	-0.0008***
	(-12.8541)
D_Loss	0.0979***
	(15.1171)
D_Qualified_OPINION	0.1986***
	(38.7957)

Variables	Log_AuditFee
Tenure	0.0119***
	(20.8730)
Obs.	89,722
R-squared	0.8104
INDUSTRY FE	YES

t-statistics are in parentheses and *, **, *** represent two-tailed statistical significance at the p-value, 0.10, 0.05, and 0.01 levels, respectively.
BigN is an indicator of Big4 auditor. LOG_TOTAL_ASSET is the logarithm of the average total assets between the beginning of the fiscal year and the end of the fiscal year. INVENTORY_RATIO is the ratio between the inventory (at the end of the fiscal year) and the average total assets. RECEIVABLE_RATIO s the ratio between the receivables (at the end of the fiscal year) and the average total assets. LTD_RATIO is the ratio between the long term debt (at the end of the fiscal year) and the average total assets. EARN_RATIO is the ratio between the operating income after depreciation (at the end of the fiscal year) and the average total assets. D_Loss is an indicator variable of the loss in the operating income in the current year (1 when there is a loss). D_Qualified_OPINION is an indicator variable of the modified opinion in the current year, including the going-concern opinion. Tenure is the number of years of audit tenure with the same client.

Appendix H: Hyper parameter setting for the implementation of the Doc2Vec

There are multiple parameters in the Dov2Vec implementation. We use the following settings.

1) **Dov2Vec implementation at each earnings conference call level**

Variables	Setting
Number of dimensions in a vector	100
Minimum count	10
Epochs	10
Seed	1 to 5
Word window	Gensim default (5)
Training algorithm #1	Gensim default (Distributed Bag of Words)
Training algorithm #2	Gensim default (training only documents)

2) **Dov2Vec implementation at each job title and section level**

(i.e., CEO comments in the presentation session, CFO comments in the Q&A session, etc.)

Variables	Setting
Number of dimensions in a vector	400
Minimum count	10
Epochs	10
Seed	1 to 5
Word window	8
Training algorithm #1	Distributed Memory
Training algorithm #2	Training both words and documents

Appendix I: Variable Descriptions in Chapter 4

Variable Name (Source)	Description
ABS_ROA (SABI)	Absolute value of ROA (Return on total asset)
AUDIT_TENURE (Accounting firms)	The number of years with the same auditor. 1 = a new client, 2 = the second year client, 3 =the third year client, and so on.
AUDIT_TENURE_SQ	Square of AUDIT_TENURE (= AUDIT_TENURE^2)
SHORT_TENURE	Dummy variable. 1 = if the AUDIT_TENURE is between 1 and 3. 0= otherwise.
LONG_TENURE	Dummy variable. 1 = if the AUDIT_TENURE is larger than 11. 0= otherwise.
CURR_RATIO (SABI)	Current asset / Current liability at the end of fiscal year
DEBT_RATIO (SABI)	Total liability / total asset at the end of fiscal year
IC_EFFECTIVENESS (Accounting firms)	Level of internal control effectiveness: 1= low internal control, to 3 = highly effective internal control
LOG_ABS_PRETAX_INCOME (SABI)	Natural logarithm of absolute value of the pretax income from ordinary activities.
LOG_MA (Accounting firms)	Natural logarithm of the amount of materiality
LOG_SALES (SABI)	Natural logarithm of total Sales amounts
LOG_TA (SABI)	Natural logarithm of total assets at the end of the fiscal year
MANAGEMENT_INTERGRITY (Accounting firms)	Level of management integrity: 1= low integrity, to 3 = high integrity
MA_RESIDUAL (Accounting firms, SABI)	Materiality looseness (=LOG_MA – logarithm of the expected materiality)
NEW_CLIENT (Accounting firms)	Dummy variable. 1 = a new client, 0 = the continuing clients
PRIOR_QUALIFIED_OPINION (SABI)	Dummy variable. 0 = unqualified audit opinion in the previous year, 1 = other opinion, or audit report is pending.
QUALIFIED_DUMMY (SABI)	Dummy variable. 0 = unqualified audit opinion, 1 = other opinion
ROA (SABI)	Return on total asset (= Income before tax / total assets at the end of the fiscal year)
SMALL_ABS_ROA_DUMMY (SABI)	Dummy variable. 1 = if the ROA_NIBT is less than 0.05. 0= otherwise.