Using Narrative Disclosures to Detect Financial Fraud

USING NARRATIVE DISCLOSURES TO DETECT FINANCIAL FRAUD

by

Lee Allen Spitzley

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Lee Spitzley, titled Using Narrative Disclosures to Detect Financial Fraud and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: July 11, 2018
Jay F. Nunamaker

_____ Date: July 11, 2018
Judee K. Burgoon

_____ Date: July 11, 2018
William J. Mayew

_____ Date: July 11, 2018
Bin Zhang

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: July 11, 2018
Dissertation Director: Jay F. Nunamaker

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Lee Allen Spitzley

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of many people through the years.

First, I would like to thank my adviser, Jay Nunamaker, for encouraging me to keep moving forward when things were difficult and really helping me understand the importance and relevance of my work and how to convey this to others. I would like to thank Judee Burgoon for sharing her vast knowledge of human behavior, research design,k and statistics. I thank Bill Mayew for helping me understand the nuances of narrative disclosures and the value of using financial reports to understand human behavior. Most of all, my committee helped me make substantial progress as a researcher.

I would also like to thank the faculty in the University of Arizona MIS department for their feedback on various versions of these essays. The Writing Skills Improvement Program helped me grow as a writer. My writing became concise, and their feedback is the reason why this dissertation has fewer than 100 pages. I would especially like to thank Jen Glass for three great semesters of productive writing groups. I also want to thank everyone at Central Michigan University who helped me get this journey started, especially Anil Kumar for mentoring me and and becoming a friend.

The PhD student cohort in Esquire was tremendously important during this journey. Lunch breaks, happy hours, complaining about things, helping with teaching, and reminding me that there are other things than just work. Brad, Michael, Rich, Steve, I doubt I would have finished without you guys.

There are those who have been there throughout my life that I would like to thank for helping me become the person I am today. My parents, Allen & and Julie, for doing a good job of raising me, teaching me the importance of hard work and persistence, and the importance of relaxing. To my siblings, Andy, Kristi, Keith, and Brad, for all of your support through the years. I also want to thank my mother- and father-in-law for taking me in and treating my as their own son. To my grandparents: Robert & Mary Ellen Thelen for persistently pushing me to (finally!) finish school; to the late Donald & Ruth Spitzley, I especially recall how proud Grandpa was that I have come so far in my education–after he had to complete school after eighth grade to work on the farm; and to Helen Spitzley. To my friends back home, thanks for the many adventures through the years and for making each trip home feel like a celebration.

Most of all: my wife, Dongchen Hou. Thanks for everything.

Contents

Contents – *Continued*

List of Figures

List of Tables

ABSTRACT

This dissertation measures the information content in narrative financial disclosures to identify linguistic differences in manager and analyst language when fraud versus when it is not. The first chapter describes the motivation for this research and an overview of the research domain. Next, I review the literature covering textual analysis of narrative disclosures and present a heuristic and classification scheme for studies in this context. In Chapter 3, I compare the language across two common narrative disclosure types: quarterly earnings calls and the Management's Discussion & Analysis (MD&A) section of quarterly and annual financial statements and find evidence of restricted incremental information from the CFOs of fraudulent companies. Chapter 4 uses a quasi-experiment to compare analyst the frequency and topics of analysts' question during earnings calls. I find that relative to nonfraudulent firms, analysts ask the managers of fraudulent firms more questions overall, and are more persistent in asking questions as a call progresses. Chapter 5 is an exploratory study of dominance and linguistic style matching from managers and analysts when interacting in the question-and-answer portion of an earnings call. The dissertation concludes with a discussion of the work, limitations, and avenues for future research.

Chapter 1

INTRODUCTION

Corporate financial fraud occurs when a company intentionally misstates information in a financial statement. It damages investors, the public, and the companies involved. Firms may use misrepresentation, concealment, or non-disclosure to achieve some material benefit, either for the company or to enrich the individuals personally (Dyck, Morse, & Zingales, 2013). Financial statement fraud comprises about 5% of all accounting fraud cases, but they are the most costly (Association of Certified Fraud Examiners, 2014). Dyck et al. (2013) estimated that 14.5% of firms are committing fraud at any given time. The costs associated with investigating a potentially fraudulent firm can easily exceed $100 million (Henning, 2012).

Identifying financial fraud is a difficult task for investors and regulators, like the Securities and Exchange Commission (SEC). Audit costs are high and false positives consume limited investigative resources, making the decision of choosing companies to investigate extremely important. To inform this decision, researchers have started to consider the language contained in narrative financial disclosures, which contain information about firm performance above what is contained in financial numbers alone (Davis, Piger, & Sedor, 2012; Throckmorton, Mayew, Venkatachalam, & Collins, 2015). One is the Management's Discussion and Analysis of Financial Position and Results of Operations (MD&A). This is a required portion of financial statements (U.S. Securities and Exchange Commission,

2008). The SEC has developed tools to analyze the text in the MD&A (Bauguess, 2016; Eaglesham, 2013).

In addition to the MD&A, most companies host quarterly earnings calls to qualitatively discuss the results of operations and future outlook. These calls tend to have two distinct portions. In the first portion, company managers, such as the CEO and CFO, will discuss the results of the current reporting period. In the second portion, financial analysts will ask questions of management, and use information from managers' responses when making forecasts and investment recommendations.

Earnings calls afford multiple opportunities to identify companies that have misstated financial results. Earnings calls contain spontaneous remarks in the question & answer (Q&A) portion, which may produce greater differences in language cues between deceptive and truthful managers, since managers have less ability to craft language than in the prepared portion of the call.

Earnings calls typically precede the release of the financial statement, giving stakeholders a timely view of operating conditions. Multiple discussions of the same fiscal period gives firms control over where to disclose sensitive items; for example, burying negative financial results in the more opaque MD&A (Davis & Tama-Sweet, 2012) Earnings calls contain spontaneous remarks in the question & answer portion, which may produce greater differences in language cues between deceptive and truthful managers, since managers have less ability to craft language than in the prepared portion of the call. The availability of the MD&A and earnings calls (both are easy to find online), and variations in exposition make these disclosures a potentially rich source for identifying behavioral signals of financial fraud.

Deception theories and empirical work show that a person's verbal patterns and style

change when someone is being deceptive (DePaulo et al., 2003; Hauch, Blandon-Gitlin, Masip, & Sporer, 2015). These linguistic cues can be classified into two distinct categories: strategic and non-strategic. In this context of financial reporting, there are prepared and spontaneous remarks, language from managers and analysts, and structured interactions. My overarching research question is: *Can linguistic behaviors in narrative financial disclosures be used to identify financial misstatements?*

These behaviors have often been measured using dictionaries that relate to specific cognitive and behavioral constructs. The relatively static nature of the dictionary-based approach limits the amount of domain-specific knowledge one can use when studying financial reports. A method to mitigate this is to create domain-specific dictionaries (Loughran & McDonald, 2011), though this is costly and time-consuming.

This dissertation utilizes natural language processing methods that can model a domain without the costs associated with creating domain-specific dictionaries. These state-of-the-art text analysis methods from information retrieval and computational linguistics will help understand how managers and financial analysts change their behavior in the presence of fraud.

I conduct three studies to analyze these documents at multiple levels. The first study compares the earnings calls to the MD&A to test for differences in incremental information between frauds and non-frauds. The second study covers investment analyst utterances in earnings calls to compare behavior when fraud is present to when fraud is not present. The third study investigates the manager-analyst dyads in earnings calls to better understand the evolving dynamics of interpersonal deception in a high-stakes scenario.

Chapter 2

LITERATURE REVIEW

2.1   Motivation

Current methods to identify potentially fraudulent companies use quantitative information (i.e. financial ratios, Abbasi, Albrecht, Vance, & Hansen, 2012; Beneish, 1999; Gaganis, 2009), and qualitative information, such as text from narrative disclosures, document comparisons, and non-verbal behavior. Many studies of qualitative data use the text from the Management's Discussion and Analysis (MD&A) section of 10-K and/or 10-Q statements (Dechow, Ge, Larson, & Sloan, 2011; Goel & Gangolly, 2012; Goel, Gangolly, Faerman, & Uzuner, 2010; Hoberg & Lewis, 2014; Humpherys, Moffitt, Burns, Burgoon, & Felix, 2011). Other studies consider manager statements during earnings calls (Larcker & Zakolyukina, 2012; Lee, 2016), and several have analyzed combined verbal and non-verbal measures to identify differences between fraudulent and non-fraudulent managers (Burgoon, Mayew, et al., 2016; Hobson, Mayew, & Venkatachalam, 2012; Throckmorton et al., 2015).

Recent literature reviews have summarized the use of NLP tools in accounting and finance research (Fisher, Garnsey, & Hughes, 2016; Loughran & McDonald, 2016). These reviews focus on text analysis methods in the financial reporting context. One important area that these and other reviews that focus on NLP methods (Barman et al., 2016; Hajek & Henriques, 2017; Kumar & Ravi, 2016; Ngai, Hu, Wong, Chen, & Sun, 2011) do not discuss is the alignment between research questions and the level of text analysis. This

review differentiates itself by instead focusing on narrative report levels and the types of research questions that can be answered at each level. With this line of thinking, researchers can align the type of question they seek to answer to the appropriate level of text analysis. I ask two questions to guide this review:

Where does information in a narrative report exist? How can we find that information, and at what cost?

I address the first question by breaking down financial texts into a hierarchy. I address the second question by identifying the tradeoffs of annotating data each level in the hierarchy, and then classifying existing literature by levels of inference and analysis. I find that research using narrative financial reports typically studies phenomena at the annual or quarterly level. There is vast potential to conduct research at the level of paragraphs, sentences, or utterances; however, the annotation costs for this granular data are potentially high. Information systems researchers can address this by developing automated and reproducible annotation methods.

2.2 Defining the structure of narrative reports

There are many levels on which to examine these narrative reports, creating the challenge of choosing an appropriate analysis strategy. In financial fraud research, one must define the level where fraud exists and then measure some attribute of that level. For example, when labeling fraud at the fiscal-quarter level, Larcker and Zakolyukina (2012) use the dictionaries from Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) to predict whether a firm had committed fraud in a quarter. Each company issues reports (and/or hosts earnings calls) periodically. A report (earnings call) has sections (call

segments), and each section (call segment) has paragraphs (turns-at-talk), sentences, and words. Table 2.1 shows these levels, how they compare across the MD&A and earnings calls, and statistical controls that are heterogeneous at each level.

Table 2.1: Levels by report type and common non-linguistic control variables

Level	MD&A	Calls	Example Covariates
1	Company	Company	Industry
2	Period	Period	Financials
3	Section	Segment	Time of day
4	Paragraph	Utterance	Speaker type
5	Sentence	Sentence	Sequence number
6	Word	Word	N/A

As data quality and text mining capabilities improve, the distinction between levels of analysis and levels of inference will grow in importance. Viewing narrative disclosures in this way also clarifies areas that have not been addressed in the current literature, and limitations to data collection and analysis at these levels.

To further elucidate the importance of document levels in research design, I use an analogy. In project management, there is a heuristic known as the "iron triangle" (Figure 2.1). A project manager can choose only two of the following: *good*, *cheap*, and *fast*. Quickly completing a project while keeping low costs is likely to be low quality. A high-quality, low-cost project will likely be slow to complete, and a high-quality, quickly-completed project will not come cheap. This "iron triangle" has an analog in psycho-linguistic analysis, which I demonstrate here. I will define the terms *precision*, *price*, and *power* in this heuristic, with contextual examples from financial fraud research. I will refer to these three terms as the Psycholinguistic Data Quality Triangle (PDQT; Figure 2.2).

Each document level has pros and cons that can be evaluated with the PDQT, and

Figure 2.1: The Project Management "Iron Triangle"

the feasibility of a study is dictated by the research question. "Annotation" refers to the meta information a researcher adds to a document (e.g. labeling a document as fraudulent, a sentence as containing forward-looking statements, or associating financial performance results to disclosures in a fiscal period). For each term, I define and identify the research-relevant questions and trade-offs that affect data collection and analysis:

Precision is the level of annotation accuracy. A key question is "Is the annotation level too broad or too narrow for the research question?" For example: using a full document to look for fraud might be too broad, since only some aspects of a business are reported fraudulently. This may introduce noise to the analysis. The annotation method a researcher uses may also affect precision: human coding, dictionaries, and other deterministic methods should all be accurate and reliable; probabilistic and latent models will be less accurate (noisy), and possibly difficult to reproduce.

Figure 2.2: The Psycholinguistic Data Quality Triangle

Precision

Pick two

Price   Power
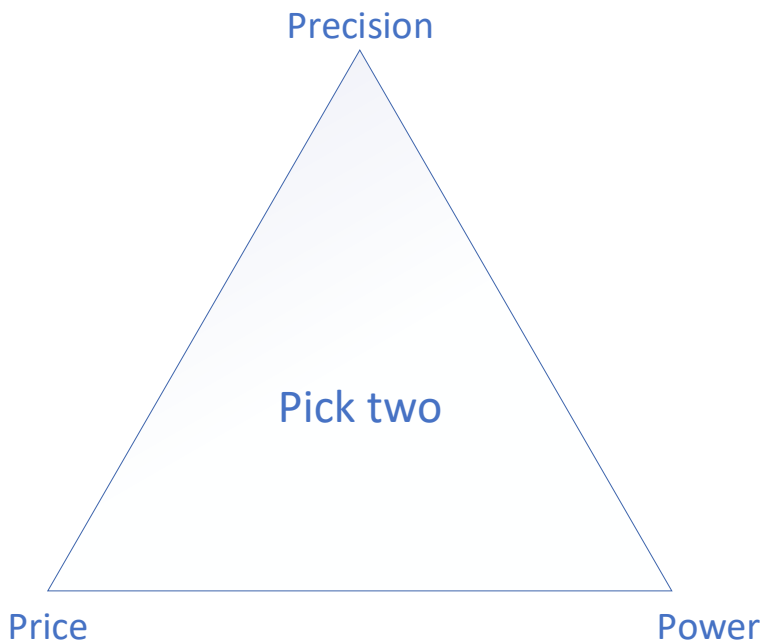
Power is the number of annotated observations for analysis. Key questions are "Will a study with high-resolution data lack statistical power to detect differences? Will low-resolution data introduce too much noise?"

Price is the cost of obtaining appropriately and accurately annotated data. Key questions include "Are there enough resources to get data that is precise enough to match the research question, yet also large enough detect the phenomenon of interest? Are there automated ways of annotating the data, or must it be done manually?"

It is highly unlikely that everything reported by a company committing fraud is untrue. Treating an entire fiscal quarter, and all of the disclosures that come with it, as fraudulent captures the noise from disclosures not related to fraud. This noise may lead to a lack of *precision* in results. Annotating data at a more granular level can consume significant resources (e.g. the tone of sentences in an MD&A (Li, 2010) or whether individual sentences

from managers in earnings calls pertain to fraud (Burgoon, Mayew, et al., 2016)). That is, finer-resolution comes at a higher *price*. *Precise* data that comes at a low *price* is likely to result in a small sample size, sacrificing *power*.

At the period level, the coding of fraud is generally cheap and easy, treating the entire document as fraudulent. It also introduces noise, most of the document likely discusses topics unrelated to fraud. Aggregation from the word level to the period level loses considerable information about the sub-structure of the document. Most studies of fraud use a bag-of-words or dictionary-based approach to summarize narrative reports for a company-period. Bag-of-words generally annotates at the period level, counting the words in each annual or quarterly report (Goel & Gangolly, 2012; Goel et al., 2010; Humpherys et al., 2011) or earnings call (Larcker & Zakolyukina, 2012). This aggregation strategy treats the entire period-report as fraudulent, trading granularity for annotation cost.

In earnings calls, the hierarchy is similar. Notable differences are the "segment" level and the "utterance" level, since earnings calls have just two segments, and each segment is comprised of individual utterances. Because this is spoken text, the next level is utterances instead of paragraphs. Another difference exists between the call section and utterance levels: the dialog (or dyad) level. This only exists in the Q&A portion of earnings calls, but analyst-manager interactions can reveal information at a precise scale (Mayew, Sethuraman, & Venkatachalam, 2016).

With this framework of document classification in place, I can evaluate existing studies in this area and note the levels of inference, levels of text analysis, and the aggregation strategy researchers use when the level of text analysis is more granular than the level of inference.

2.3 Categorizing prior work by level

In this section, I will categorize studies that use narrative financial reports to predict fraud or other future performance measures. There are many studies that use the text in 10-K or 10-Q filings that are not included in this list. I strategically included studies that produced a novel combination of levels and analysis methods. See Table 2.2 for a summary of the studies included in this review.

I categorize each study by the level that it makes inferences. If a study uses bag-of-words to predict annual performance, it belongs at the period level, not the word level. In general, the level of text analysis is deeper than the level of inference; therefore researchers must aggregate the text analysis results to match the level of inference. For example, Li (2010) builds a model to categorize each forward-looking statement from quarterly and annual MD&As as positive (+1), neutral (0), or negative (-1) tone. He then gets the mean tone from all sentences in a document to achieve a period-level measure. The study uses this measure as an indicator of future performance.

Most of these studies tend to focus on period-level analyses. There are several possible explanations for this. First, financial reporting and evaluation typically occurs at a quarterly or annual level. When accounting measures are a dependent variable, the period level is often an appropriate fit. Second, the SEC EDGAR database makes MD&A texts easy for researchers to access, and open-source 10-K/10-Q parsing tools make text extraction trivial. A third reason may be convention, publishers and reviewers are familiar with this venue which leads to selection bias.

Studies at lower levels are significantly less common, but they yield interesting results.

Table 2.2: Financial text analysis since 2010

This table lists recent studies that have conducted text analysis on narrative financial reports. The **Venue** column is the type of document the study used [MDA = Management's Discussion & Analysis, EC = earnings call, EPR = earnings press release]. The **inference** column captures the level where the study draws conclusions. The **Analysis** column shows the level where the study derived raw text measures. **Method** shows the type(s) of text analysis the study used. **Aggregation** describes how each study consolidated the text measures from the analysis level up to the inference level.

Study	Journal	Venue	Inference	Analysis	Method	Aggregation
N. C. Brown et al. (2018)	Working	MDA	Annual	Annual	LDA	Topic dist over document
Cecchini et al. (2010b)	DSS	MDA	Annual	Sentence	WordNet Ontology	Concept scores per 10K
Dong et al. (2016)	PACIS	MDA	Annual	Annual	LDA + Dict	10K level feat + word feat
Glancy and Yadav (2011)	DSS	MDA	Annual	Annual	LSA/SVD	Topic dist over 10K
Hoberg and Lewis (2017)	JCF	MDA	Annual	Annual	LDA	Topic dist over 10K
Humpherys et al. (2011)	DSS	MDA	Annual	Word/Sentence	SPLICE	Counts & ratios per 10K
Karapandza (2016)	JBF	MDA	Annual	Word	Tense	Frequency of future-tense & aux verbs
Lehavy et al. (2011)	TAR	MDA	Annual	Sentence	Readability	Score per document
Li (2010)	JAR	MDA	Quarter	Sentence	Tone	sum by category (-1,0,1) per 10X
Mayew et al. (2015)	TAR	MDA	Annual	Word	Tone	Pos and neg tone ratios
Purda and Skillicorn (2014)	CAR	MDA	Quarter	Quarter	BoW	Weighted word counts from 10X
Larcker and Zakolyukina (2012)	JAR	EC	Quarter	Word	LIWC	LIWC word counts by call
Throckmorton et al. (2015)	DSS	EC	Quarter	Word	LIWC	Counts & ratios per doc
Allee and Deangelis (2015)	JAR	EC	Segment	Segment	Tone	Dispersion measure per segment
Lee (2016)	TAR	EC	Segment	Speaker	Style	Style similarity per manager in Q&A
Mayew et al. (2016)	Working	EC	Utterance	Utterance	Tone	Utterance-level tone
Burgoon et al. (2016)	JLSP	EC	Sentence	Sentence	SPLICE	Sentence level features
Huang et al. (2014)	TAR	EPR	Annual	Word	LM tone	Abnormal tone per doc
Davis and Tama-Sweet (2012)	CAR	EPR, MDA	Quarter	Word	DICTION	Percent pessimistic/optimistic per doc

Research at the granular level can more closely match the phenomenon of interest and the behavioral correlates or causes of dependent variable change.

2.4   Chapter Summary

By breaking down existing research according to document level, the pros and cons of each level of analysis become clearer. While most research focuses on quarterly or annual financial statements, there are tremendous possibilities for work at more granular levels. These levels directly influence design choices based on the theory being tested. Systems-oriented researchers can contribute by designing methods that reduce the costs associated with analysis at the utterance or sentence level by incorporating machine learning, topic modeling, and emergent text analysis methods.

Chapter 3

INCREMENTAL INFORMATION BETWEEN EARNINGS CALLS AND FINANCIAL

STATEMENTS

3.1   Motivation

Studies that use narrative disclosures to identify fraud tend to focus on a single reporting

venue (Hoberg & Lewis, 2017; Humpherys et al., 2011; Larcker & Zakolyukina, 2012). One

limitation of this approach is that it does not consider the relationships between the multiple

disclosures that cover a single fiscal period. Firms often release these disclosures at different

times, leaving managers the opportunity to adjust their story between disclosures (Davis &

Tama-Sweet, 2012). Whether or not fraudulent firms modify their stories differently than

non-fraudulent firms remains unknown. Dissimilarity between narratives is an indicator of

greater information disclosure (S. V. Brown & Tucker, 2011; Davis & Tama-Sweet, 2012; Lee,

2016), which may increase attention from investors and regulators. Fraudulent managers

may use greater similarity between narratives to reduce the risk of presenting contradictory

information that comes from maintaining a complex deception, particularly during the Q&A

of the calls. To reduce inconsistencies and create a favorable impression, it is possible that

fraudulent companies will disclose less information across reporting venues. Repetition can

lead to a more persuasive message (Cacioppo & Petty, 1989; Petty & Cacioppo, 1979).

Language similarity may also indicate that managers take a more active role in the document

preparation process. This would reduce the number of people who interact with the area of

fraud. This leads to my research question:

Relative to non-fraudulent companies, do the CEOs and CFOs of fraudulent companies use language that is more similar between their conference calls and MD&A sections than the CEOs and CFOs of non-fraudulent companies?

To answer this question, I present a novel approach to identifying fraudulent financial statements by considering content modifications in narratives covering a single reporting period (i.e. fiscal quarter). This study measures language similarity between the earnings calls and subsequent MD&A section. Earnings calls present an interesting opportunity to investigate prepared statements and spontaneous remarks from the question and answer session with analysts.

Figure 3.1: Relation to other narrative comparison studies
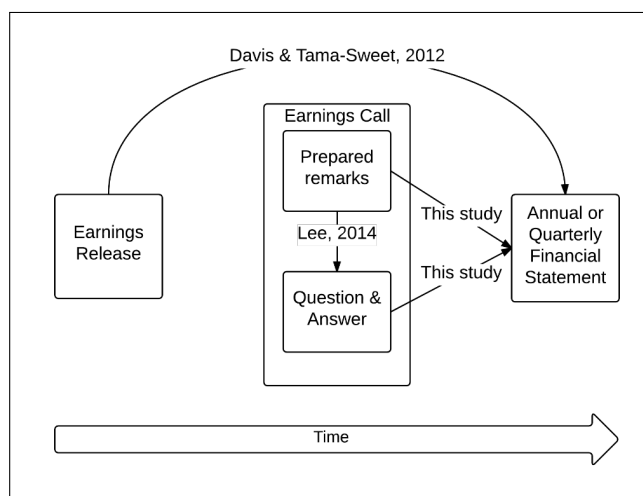
Answering this question has implications for both the deception and accounting domains. In deception research, it is often difficult to create high-stakes deception in an experimental setting, and obtaining real-world data with known cases of truth and deception is challenging. Also of interest to deception researchers are story-consistency strategies by deceptive parties when there is ample time to prepare.

From an accounting and fraud investigation perspective, this research expands on knowledge of strategic corporate reporting. Analyzing multiple disclosures provides a chance to examine variations between truth-tellers and deceivers at multiple times. This will also improve understanding of strategic corporate financial reporting by learning how fraudulent and non-fraudulent firms differ in their information disclosure strategies between venues.

From a practical perspective, there is a growing interest in the analysis of qualitative information to identify fraud in public companies. This non-financial information provides a complementary source of information to the quantitative analysis of financial measures. For example, the SEC is increasing its focus on the MD&A section of annual reports because of their ability to distinguish fraudulent activities (Association of Certified Fraud Examiners, 2013).

The remaining portion of this chapter is as follows: In Section 3.2, I summarize the relevant background literature on financial fraud and prior methods used to identify fraudulent companies, describes the research questions. In Section 3.3, I describe the data collection process and my resulting sample. Following that is an analysis of the data (Section 3.4), and I conclude with a discussion of the results and avenues for future study (Section 3.5).

3.2   Research Development

3.2.1   Disclosure strategy and impression management

Impression management is the "process by which people control the impressions others form of them" (Leary & Kowalski, 1990). There are two major components of this process: impression construction and impression motivation. Impression construction is the type of

impression one is trying to create, and impression motivation is the level of how motivated someone is to control how others see them. While this original description was defined in terms of individuals, this definition of impression management is also useful in understanding how companies strategically present information through their disclosures (Lo & Rogo, 2014; Merkl-Davies & Brennan, 2007; Merkl-Davies, Brennan, & McLeay, 2011).

Narrative disclosures present an opportunity for impression management as a way to advance their goals (Merkl-Davies & Brennan, 2007), and firms do use these opportunities to present a positive image (Merkl-Davies et al., 2011). When there is bad news to report, for example, firms have a tendency to withhold this information from a relatively transparent reporting venue (earnings press release) and place it in the text of the financial statement (Davis & Tama-Sweet, 2012). Companies that experience a change in CEOs will strategically use the presentational graphs in financial statements to create a favorable image of their performance (Godfrey, Mather, & Ramsay, 2003). In the earnings calls with analysts, managers in firms that are performing poorly or are at risk of lawsuits tend to script their responses to analysts during the question and answer portion of the call, presumably to mitigate the risk of accidental information disclosure (Lee, 2016). Another way to portray a desired identity is by maintaining consistency; the language in financial statements seems to be consistent with a firms reported financial state (Merkl-Davies et al., 2011).

In the case of fraud, managers not only need to use impression management to portray their desired image–they must also portray a persuasive image that leads their audience to believe that what they are reporting reflects the true results of their performance. This additional complexity in impression construction should lead to differences in the ways that fraudulent and non-fraudulent companies present information in narrative disclosures.

3.2.2   Differences between groups

When executives are trying to deceive in this context, I assume that they are motivated to convince investors and analysts that their financial information is accurate through impression management techniques. While both fraudulent and non-fraudulent firms engage in impression management, deceptive parties should be more motivated to create a consistent story. Prior research on deception has revealed that unrehearsed liars tend to have more inconsistencies their language (Walczyk, Mahoney, Doverspike, & Griffith-Ross, 2009), and these inconsistencies may increase suspicion by other parties involved in the interaction (Buller & Burgoon, 1996).

There are several reasons to suspect that fraudulent firms will have language that is more similar between disclosure venues. In this setting, there are two options for persuading the audience of their integrity: discussing only the accurate items (and ignoring the fabricated items, deception by limiting disclosure) or by disclosing fabricated items (deception by fabrication) (Hoberg & Lewis, 2017). In the first instance, there should be an increase in similarity because of the limited pool of items to discuss. In the second case, deception by fabrication, there should be increased similarity because of attempts to maintain consistency in the fabricated story.

Deceivers tend to use less-diverse language in spontaneous communication (Zhou & Zhang, 2008). However, the use of less-diverse language also appears in the thoroughly-prepared MD&As of fraudulent companies (Humpherys et al., 2011) Like the MD&A, the prepared portion of the call is also thoroughly prepared. Less-diverse language may reflect a strategy to keep a manufactured story consistent. Although theories of cognitive load are

not likely to apply to the prepared portion of the MD&A, there may be an incentive to keep this portion simple to avoid increasing the difficulty of answering questions during the Q&A. To mitigate the potential of disclosing information that may later be used against the company, at-risk companies tend to use higher amounts of scripting during the Q&A portion of the call (Lee, 2015). If this is the case, then scripting should lead to higher similarity between Q&A and MD&A. Another possibility is that managers convey consistency across narratives by incorporating off-script remarks into the MD&A.

The strategies of either limiting information or disclosing fabricated information lead to the following research question:

RQ1: Do fraudulent managers have more similar language than non-fraudulent companies between the conference call and the MD&A (a) for the prepared portion, and (b) for the Q&A portion?

3.2.3 Differences between CEOs and CFOs

Larcker and Zakolyukina (2012) separate the data by CEO and CFO. Their results show greater predictive ability from CFOs than CEOs. I also separate my data to compare findings from CEOs and CFOs.

3.2.4 Predictive ability of the prepared portion vs. the Q&A

Larcker and Zakolyukina (2012) revealed a counter-intuitive finding: that the prepared remarks and Q&A remarks in an earnings call display similar linguistic variations between fraudulent and non-fraudulent firms. This is interesting, as one might expect a less-rehearsed scenario to have greater predictive ability due to greater cognitive load. It is also possible

that the prepared remarks will have a greater predictive ability, since analysts lead the discussion in during the Q&A. Bloomfield (2012) calls for further investigation into this issue; therefore, I propose the following exploratory research question:

RQ2: Are there differences in predictive ability between the prepared remarks and Q&A?

3.3 Method

3.3.1 Rationale

This study investigates quarterly earnings calls, rather than earnings press releases, because they contain prepared and spontaneous remarks and provide information above what is contained in the accompanying press release (Matsumoto, Pronk, & Roelofsen, 2011). The MD&A is a suitable comparison document because companies often release it sometime after the occurrence of the earnings call, giving managers ample time to manage strategically any new information in the MD&A.

3.3.2 Fraud identification

There are varying approaches to determining which companies to label as fraudulent (Karpoff, Koester, Lee, & Martin, 2012). One approach is to use firms SEC identified as fraudulent through an AAER (Cecchini, Aytug, Koehler, & Pathak, 2010a; Dechow et al., 2011; Humpherys et al., 2011; Purda & Skillicorn, 2014). Larcker and Zakolyukina (2012) use a wider approach, using companies that had a disclosure of material weakness, an auditor change, a late filing, or a Form 8K filing. Another potential source of fraud identification is

the Stanford Securities Class Action Clearinghouse (SSCAC) dataset, which contains many instances of shareholder lawsuits initiated because of fraud. However, this data also includes many other lawsuits not related to fraud, and after eliminating frivolous lawsuits and those occurring in small firms, the size of the sample drops below 250 observations (Dyck et al., 2013).

I use companies identified as fraudulent by the SECs AAER database because of its general acceptance of identifying fraud, and the fact that it has very few false positives (i.e., companies labeled as fraud that were not fraudulent). I used the database from Dechow et al. (2011), which contained the reason for the AAER and the relevant time periods. This database was current through September 2013. To increase sample size, I followed their procedures and added AAERs from September 2013 through January 2015. The final sample contains fraudulent statements from 2006 through 2013. The selected companies had at least one quarterly statement affected by the fraud. Like many prior studies, I assumed that companies not identified by an AAER were not fraudulent.

3.3.3 Data

Seeking Alpha (seekingalpha.com; hereafter SA) hosts earnings call transcripts and makes them publicly available. For a few companies with transcripts on SA, the transcribed earnings calls begin appearing for events that took place around 2006, with many others beginning at some time before 2008 (See 3.2). For each firm in the AAER database, I attempted to locate the URLs for the earnings call transcripts that take place during the period of the fraud. There were a total of 139 fraudulent earnings calls available for this period.

Figure 3.2: Seeking Alpha Call Coverage

Seeking Alpha began transcribing earnings calls around 2006. The oldest data in my sample is limited by when Seeking Alpha began transcribing a firm's calls.
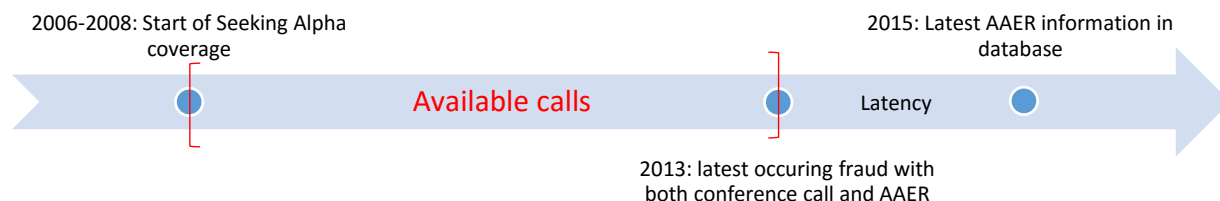
For each call, a web scraper separated the utterances from the call by speaker, position, and company (if an analyst). It also separated the calls by part (prepared remarks or Q&A). I collected financial statements from the SEC EDGAR database. A script separated the MD&A sections from the rest of the statement and removed tables from the text.

Control companies

Because the transcript database has limited coverage, many companies lack enough pre-fraud data to compare within firms. Therefore, I create a comparison group using companies with similar financial characteristics. Table 3.1 shows the similarity criteria. I used the first two digits of the SIC number to find companies from a similar industry, and then narrowed the results further by matching firms with similar market valuations. All firms were matched using data from the quarter when the fraud starts. In the case of multiple firms meeting the criteria, the control company was randomly selected. If the selected firm had no transcripts available, a new control company was selected at random until one was found with available transcripts.

Table 3.2 shows the breakdown of firms by industry. Approximately half of all fraudulent

Table 3.1: Matching criteria for control companies

Variable	Description	Matching criteria
SIC	Standard Industry Code	Same two-digit SIC
ln(ASSETS)	Natural log of the firm's assets	+/- 15%

firm-quarters were from depository institutions (i.e. commercial banks) and business services (largely software development companies).

Table 3.2: Number of calls per industry

Top-level SIC	Description	Fraud	Non-fraud
16	Heavy construction	2	2
20	Food	1	1
23	Apparel	5	6
35	Industrial machinery	4	5
36	Electrical equipment	2	2
37	Transportation equipment	4	4
38	Instrument mfg	5	7
59	Misc. retail	12	12
60	Depository institution	30	24
61	Non-depository credit institution	4	2
62	Security & commodity brokers	2	4
64	Insurance agents	3	3
67	Holding	3	5
73	Business services	31	34
87	Engineering	6	6
		114	117

3.3.4 Measuring language similarity

To compare the earnings calls to their respective MD&A, I first weight the word counts for each document using the term frequency-inverse document frequency (tf-idf) weighting algorithm and then measure similarity using cosine similarity (Manning, Raghavan, & Schutze, 2008, pp.117-125).

Term frequency represents the number of times a term appears in each document, represented as $tf_{t,d}$, which is a $t \times d$ matrix containing the number of times each term appears for each document. The idf formula is

$$idf_t = \log \frac{N}{df_t} \tag{3.1}$$

where idf_t is the weighting of term t, N is the number of documents in the corpus, and df_t is the number of documents containing term t.

The multiplication of $tf \times idf$ yields the weighted matrix for the documents in the corpus. Each row vector in this matrix represents a document, and these vectors can be compared to each other using cosine similarity. Because these vectors only contain non-negative values, the value of the cosine similarity will be between 0 and 1, where 0 represents two completely orthogonal vectors (completely dissimilar) or 1, which represents the same vector (completely the same). The formula for cosine similarity is

$$\cos \theta = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} \tag{3.2}$$

where θ is the angle between vectors v_1 and v_2, the numerator is the dot product of the two vectors, and the denominator is the product of the vector lengths (i.e. $\|v_1\|$ represents the length of v_1). This function returns a value between 0 and 1.

Financial literature has used tf-idf and cosine similarity measure to compare MD&A sections for modifications over time (S. V. Brown & Tucker, 2011). One of the primary benefits of tf-idf is that it gives more weight to rare terms in a corpus and greatly diminishes

the importance of function words like "the" or "and". Many words have connotations in an everyday setting, such as "debt" or "gain", but mean little in a financial setting, where these terms are frequent (Loughran & McDonald, 2011); tf-idf can mitigate this issue.

3.3.5   Text preparation

For all earnings call turns-at-talk belonging to a manager, I identified whether the speaker was the CEO or CFO. I then merged each manager's language for each part of the call. For example, all CEO remarks during a Q&A session would merge into one text document. The Porter stemming algorithm in the Natural Language Toolkit Python package (NLTK; Bird, Klein, & Loper, 2009) reduced words in earnings calls and MD&As to their root form.

After cleaning the text, I computed the idf weights for each word. For each call portion and the MD&A, I created the term frequency matrix, and then computed the tf-idf matrix. I then used cosine similarity to compare the tf-idf vectors of the prepared remarks and Q&A of each call to the corresponding MD&A section. For each combination, I merged the similarity measures with the same-period financial variables from Compustat.

3.3.6   Document length correction

S. V. Brown and Tucker (2011, pp. 343-344) showed that similarity will increase as document length increases. To correct for this, they estimated a polynomial regression to get the expected similarity score for a pair of documents. From the expected similarity score they subtracted the actual similarity score to get the residual value. The residualized value is free of the effects of document length. On this new measure, higher values still have the same interpretation: the higher the value, the greater the similarity (respective to what we

would have expected for a document of the given length). I also use this adjustment in my analysis.

3.4 Analysis

3.4.1 Descriptive statistics

Table 3.3 presents the sample descriptive statistics. There were fourteen industry groups represented, with many companies coming from the categories of depository institutions and business services (such as IT firms). There were more observations near the middle of the sampling period, reflecting the limited coverage of Seeking Alpha in the earlier years and the latency between fraud occurring and the filing of an AAER in the later years.

There were 275 transcripts on SA. After eliminating transcripts that did not parse properly, quarters that did not have a matched MD&A, and those without complete financial data, the total usable number of observations was 230. The final sample consists of 114 fraudulent quarters from 31 companies, and the non-fraudulent sample contains 116 quarters from 31 companies. I selected control variables that may have influenced the level of similarity or were important in prior studies. Appendix B contains a list of these variables and descriptions of each.

The similarity scores (SIMSCORE) are separated by part of the call. In relation to other studies that use cosine similarity measurements, S. V. Brown and Tucker (2011) report a mean similarity score of 0.845 when comparing current-period MD&As to prior period MD&As. The similarities of the prepared portion of the calls in the current studies is lower than this figure, likely due to the spoken nature of the prepared part of the call. Lee (2016)

reports a mean similarity of 0.872 when comparing CEO speech from the Q&A portion of conference to the CEO speech in the prepared portion of the call. The similarity of the Q&A and MD&A in the current study are considerably lower; however, Lee uses function words (e.g. a, an, the, or), which makes the similarity scores between these studies difficult to compare.

I include MD&A length (measured in number of words; MDA_LEN) in the event that a longer MD&A section represents a more complex story and therefore requires the executives to maintain more consistency in their disclosures. SIMSCORE correlates with MDA_LEN, which supports the inclusion of this control variable in my regression and classification models.

Table 3.3: Similarity between manager speech and the corresponding MD&A

	Fraud			Non-fraud		
Item	N	Mean	SD	N	Mean	SD
SIMSCORE						
CEO Prepared	106	0.281	0.187	111	0.307	0.168
CEO Q&A	107	0.105	0.093	113	0.104	0.077
CFO Prepared	100	0.349	0.193	106	0.314	0.157
CFO Q&A	100	0.088	0.084	106	0.087	0.070

3.4.2 Regression model

Because of the considerable difference in means of SIMSCORE between parts of the calls and the interaction of analysts in the Q&A, I ran separate regressions for each part of the call, and for CEO and CFO, resulting in four regressions. I use a random-effects model to estimate the time-invariant coefficient of FRAUD and its effect on the similarity between the two documents. This model is estimated using cluster-robust standard errors, with

clustering at the firm level. LOSS is dummy-coded, with 0 representing net income equal to or greater than \$0, and 1 for net income less than \$0. QUARTER is a dummy variable to control for the effects that reporting in a specific period may cause. The model below is the same for both parts of the call in most respects; however, the number of observations between CEO and CFO regressions vary based on participation.

$$SIMSCORE = \beta_0 + \beta_1 FRAUD + \beta_2 LOSS + \beta_3 MEET\_BEAT$$
$$+ \beta_4 \ln(ATQ) + \beta_5 QUARTER + \beta_6 TDSIC$$
$$+ \beta_7 \ln(AGE) + \beta_8 SOFT.ASSETS + \beta_9 SCH.REC$$
$$+ \beta_{10} SCH.INV + \beta_{12} SCH.ROA + \beta_{13} CAPMKT$$

The results of these regressions are in Table 3.4 (CEO) and Table 3.5 (CFO). For CEO language, there were no differences in similarity scores between the CEOs of fraudulent companies and the CEOs of non-fraudulent companies for both the prepared remarks ($p = 0.18$) and the Q&A remarks ($p = 0.38$). For CFO language, the CFOs in fraudulent firms had significantly higher language similarity between their prepared remarks and the same-period MD&A than non-fraudulent CFOs.

3.4.3   Classification accuracy

To test for the predictive ability of the variables in the regression (RQ1), I used several different classification algorithms in Weka (Frank, Hall, & Witten, 2016). The calls were again separated by call segment and speaker, and the regression variables, including SIMSCORE,

Table 3.4: CEO language similarity regressions

	Prepared	Q&A
(Intercept)	2.6203	0.7887
	(2.4010)	(0.7000)
fraud2	0.0236	0.0022
	(0.0344)	(0.0166)
loss	−0.0228	−0.0011
	(0.0276)	(0.0189)
log(atq)	0.0169	0.0066
	(0.0112)	(0.0060)
as.factor(quarter)2	0.0294	**0.0230***
	(0.0203)	(0.0111)
as.factor(quarter)3	0.0154	−0.0101
	(0.0253)	(0.0121)
as.factor(quarter)4	**0.0640***	0.0239
	(0.0246)	(0.0146)
log(age)	−0.4081	−0.1038
	(0.4161)	(0.1148)
soft.assets	**−0.1883***	−0.0379
	(0.0865)	(0.0746)
sch.rec	0.4037	0.0665
	(0.3636)	(0.1535)
sch.inv	0.1852	0.1161
	(0.3610)	(0.1448)
sch.roa	−0.1045	−0.2306
	(0.2175)	(0.1376)
capmkt	−0.0142	−0.0165
	(0.0222)	(0.0097)
meetbeat	**0.0322***	0.0120
	(0.0138)	(0.0096)
R^2	0.2447	0.2061
Adj. R^2	0.1255	0.0807
Num. obs.	199	199

$^{***}p < 0.001$, $^{**}p < 0.01$, $^*p < 0.05$

Table 3.5: CFO language similarity regressions

	Prepared	Q&A
(Intercept)	2.7194	0.4292
	(1.9899)	(0.6932)
fraud2	**0.0901***	0.0024
	(0.0395)	(0.0175)
loss	−0.0284	−0.0113
	(0.0191)	(0.0133)
log(atq)	0.0059	−0.0042
	(0.0158)	(0.0067)
as.factor(quarter)2	−0.0109	0.0288
	(0.0228)	(0.0155)
as.factor(quarter)3	−0.0020	**0.0305***
	(0.0250)	(0.0145)
as.factor(quarter)4	−0.0478	0.0194
	(0.0316)	(0.0141)
log(age)	−0.5060	−0.0878
	(0.3503)	(0.1183)
soft.assets	0.0197	0.0491
	(0.1050)	(0.0488)
sch.rec	**0.4668***	0.0928
	(0.2139)	(0.0917)
sch.inv	**0.4420***	0.0137
	(0.1879)	(0.1078)
sch.roa	−0.0564	0.2553
	(0.1836)	(0.1424)
capmkt	0.0004	−0.0176
	(0.0193)	(0.0118)
meetbeat	0.0266	**0.0275****
	(0.0141)	(0.0085)
R^2	0.2215	0.1752
Adj. R^2	0.0791	0.0243
Num. obs.	182	182

$^{***}p < 0.001$, $^{**}p < 0.01$, $^*p < 0.05$

were used as the feature sets. Each algorithm attempts to predict fraud using the variables included in the regressions. The results are reported in Table 3.6. I use multiple algorithms because of the possibility that any single method may vary considerably in its performance (Gaganis, 2009). Each algorithm was tested using 10-fold cross validation.

Table 3.6: Fraud classification accuracy when using similarity score measures for CEOs and CFOs, and financial period control variables

Method	Prepared	QA
Logistic	57.70%	56.82%
C4.5	65.63%	61.20%
Naïve Bayes	57.26%	59.90%
SVM	58.10%	58.10%
Bagging	75.33%	74.44%
Means	62.80%	62.09%

The prepared remarks tended to have marginally higher accuracy results. For C4.5 and Bootstrap Aggregation (bagging), the results were significantly higher than chance. Bagging uses multiple random samples and averages decision trees to improve results.

The logistic regressions and support vector machines (SVM) showed little difference between the parts of calls. These results do provide some insight into RQ2. The prepared remarks, at the very least, perform as well as the Q&A, and may perform better. These results were similar to other studies that tested classification accuracy using language-based models: Larcker and Zakolyukina (2012) achieve accuracy between 56-66% and Humpherys et al. (2011) achieve between 58-67%.

3.5   Discussion

In this study, I explored the possibility that executives in fraudulent companies would use more similar language between their earnings call and MD&A than non-fraudulent companies. To test this, I collected earnings calls and MD&A texts from fraudulent and non-fraudulent firms. I used similarity scores to measure marginal information disclosure between venues, where lower scores indicate more information. The CFOs of fraudulent firms had significantly greater similarity scores than their non-fraudulent peers during the prepared remarks. This may be the result of attempting to maintain a more consistent story across disclosure venues. These differences did not show up for CEOs, which may be caused by a lower amount of involvement in preparing the MD&A than the CFO, or less knowledge of the financial manipulations. There is an interesting observation that may corroborate this conjecture. In the Q&A portion of a call, the CEOs of fraudulent firms averaged nearly eight more turns at talk than the CFOs (22.6 to 14.9); whereas in nonfraudulent firms, the CEOs averaged just over three more turns at talk than the CFOs (19.3 to 16.1). There were minimal differences between CEOs and CFOs in the prepared remarks.

There were no statistically significant differences in similarity between fraudulent and non-fraudulent companies in the Q&A portion of the call, although the fraud coefficients were greater than zero. Analysts generally control the dialog in this setting, which may limit the ability of managers to use strategic behavior. This finding is counterintuitive, since most spontaneous communication should show greater signs of leakage. It is, however, consistent with Larcker and Zakolyukina (2012), who also find the prepared remarks useful for identifying fraud.

This study has several limitations. While AAERs can reasonably establish the ground truth of fraud and non-fraud, it is often unclear if those preparing and presenting the information in the earnings calls and financial statements are aware of the fraud. The sample is somewhat small; however, this number is similar to other fraud research based on AAER data. Humpherys et al. (2011) used 101 fraudulent MD&As from 10-Ks, and Larcker and Zakolyukina (2012) used conference calls from 274 fraudulent firm-quarters. The use of tf-idf and cosine similarity as a measurement tool may be too coarse to capture the meaningful modifications between documents. Deeper semantic analysis may perform better by identifying topical subsets of financial statements and earnings calls. Lastly, it does not consider information contained in the earnings press release.

Future research in this area should investigate the role of financial analysts in uncovering (or facilitating) fraud through their interaction with executives in the earnings calls. It is known that executives choose to speak with analysts who are more favorable to the company, and they may select those who minimize their risk of being caught. Analysts drive the conversation in the Q&A, and therefore have the potential to mitigate strategic information disclosure by managers.

Chapter 4

ANALYST QUESTIONS

4.1 Introduction

Current methods to identify potentially fraudulent companies use quantitative information (i.e. financial ratios), and qualitative information, such as text from voluntary disclosures, document comparisons, and non-verbal behavior. The SEC's Division of Economic and Risk Analysis (DERA) has developed advanced text analytics capabilities to identify potential risks and help decision-makers choose where to allocate investigative resources (Bauguess, 2016; Wilczek, 2014). While these methods have shown promise, they are potentially limited by the fact that fraudulent firms can manage the narrative and make anomaly detection difficult.

Text and behavioral analytics currently rely on firm disclosures, such as the text from the Management's Discussion and Analysis (MD&A) section of 10-K and/or 10-Q statements (Dechow et al., 2011; Goel & Gangolly, 2012; Goel et al., 2010; Hoberg & Lewis, 2017; Humpherys et al., 2011), earnings call text (Larcker & Zakolyukina, 2012; Lee, 2016), and multi-modal methods (Burgoon, Mayew, et al., 2016; Hobson et al., 2012; Throckmorton et al., 2015). This research will address the limitations of firm-generated data by investigating analyst turns-at-talk in quarterly earnings calls–a novel information source in the fraud detection domain. Fraudulent managers must deceive market participants in order to be successful in their fraud. In particular, they must satisfy the analysts who study the company

that what they say is accurate and represents the true state of the business. If an analyst perceives management's statements on a topic as deceptive, uninformative, or otherwise unsatisfactory, presumably analysts will ask questions to resolve their concerns. Such a strategy makes intuitive sense, given the analyst's limited interactions and potential to maximize information gain through their questions.

There are several reasons that this information source can lead to new insights into fraudulent behavior. Current fraud detection methods do not consider the ability of analysts, who report approximately 17% of fraud cases. This is equal to the number discovered by auditors and the SEC combined (Dyck, Morse, & Zingales, 2010). Analysts have substantial interactions with managers, and these interactions guide the discussion and extract information. It is possible that analysts' frequent interactions with management and intimate knowledge of the firms and industries they cover may give them insights into when a firm may be fraudulent. Some analysts receive FBI training to understand behavioral cues (L. D. Brown, Call, Clement, & Sharp, 2015). Analysts control the direction of the narrative during the Q&A, reducing managers' ability to steer away from fraudulent subjects. From a methodological perspective, using analysts to identify differences caused by fraudulent reporting gets much closer to random condition assignment than using managers. If we assume there are no analyst traits that would correlate to covering fraudulent firms, presence of one or more fraudulent firms in their coverage basket is random. This results in a quasi-experimental design, free of the self-selection bias present in any study that relies on manager-generated information.

If analysts suspect deceptive reporting from managers, Interpersonal Deception Theory (IDT; Buller & Burgoon, 1996) posits that they will adjust their interactions with the

Figure 4.1: Research domain

This figure shows the transfer of internal information to the public through earnings calls. Analysts use the earnings call venue to question managers and extract internal information. This research investigates analysts' ability to provide information about fraud by studying the content of analyst utterances in the earnings calls (marked in blue fill) to learn about analysts' suspicions (red dotted line).
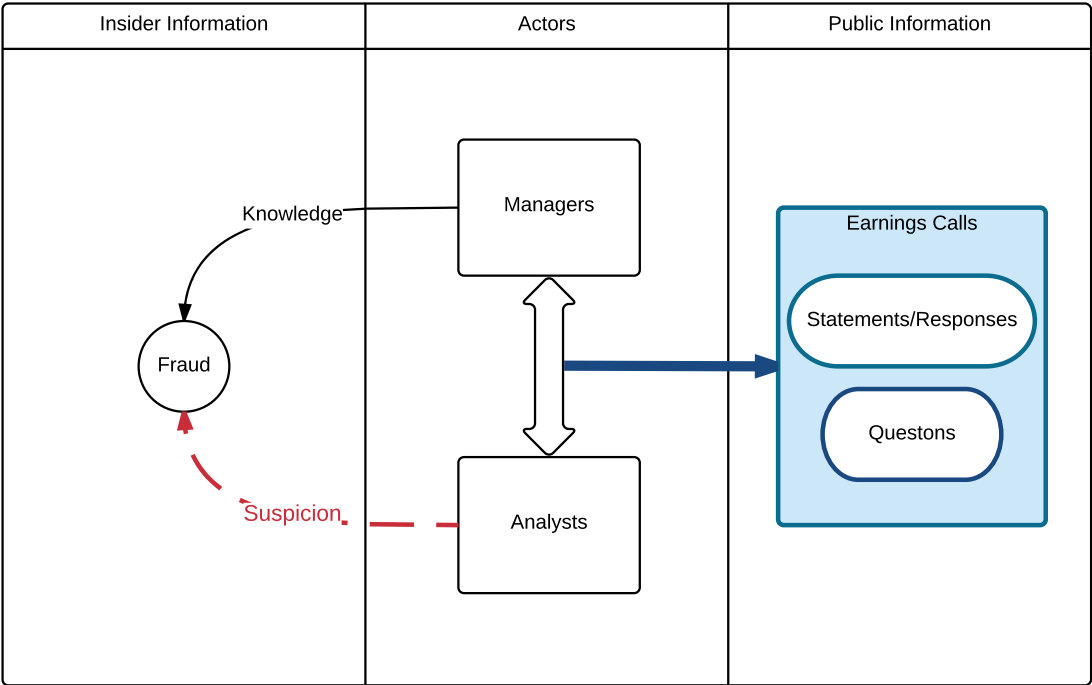
managers to make a credibility judgment. I use a natural experiment with analysts who covered both fraudulent and non-fraudulent firms in the same industry. I measure the rate of questions to statements in analyst turns-at-talk, with abnormal inquisitiveness as a proxy for suspicion. I also use Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003) to assign a topic to each analyst turn-at-talk. LDA is a probabilistic algorithm that uses a corpus of documents to identify semantically distinct topics. After identifying topics, LDA can then report the probability distribution of the topics over each document. Such a technique can reveal the topics of analyst turns-at-talk under varying firm performance conditions, including fraud.

This paper makes several contributions. First, it demonstrates the usefulness of analyst

questions to identify potential misstatements. It is also one of the first studies interactive deception in a high-stakes, real-world setting. This research also contributes to a growing literature on latent topic modeling over financial texts. This paper also shows the potential of quasi-experimental designs with analyst interactions. It also provides preliminary evidence that analysts tend to spend more time discussing topics that firms under report during fraud.

4.2   Literature Review

4.2.1   Interpersonal deception

Most deception research focuses on the behavior of the message creator, with little attention paid to message receivers beyond their judgments of message veracity. This is a reasonable approach, given the findings from Bond and DePaulo (2006) that individuals have an average deception detection rate of 54%. Bond and DePaulo go on to note that experiments on deception detection typically require participants to hear a message and judge its veracity in the same session. In financial fraud detection, judgments are not immediately required, and analysts and investors have financial reports, colleagues, and repeated, direct interactions with managers that may be concealing a fraud. This environment places analysts in a position to identify areas of insufficient or suspicious information in reports and dig deeper through interviews. Most people who discover lies rely on third party information and physical evidence (Park, Levine, McCornack, Morrison, & Ferrara, 2002).

In addition to a lack of corroborative information in most experimental settings, judgment accuracy is impeded because individuals regard communications as truthful, often termed

"truth-bias". However, a conclusion of truth does not preclude suspicion. Hancock, Curry, Goorha, and Woodworth (2007) found that when a sender was lying to a conversation partner on a topic, the partners asked more questions and used more words than when the sender was truthfully discussing a topic. While the conversation partners showed increased suspicion in the presence of lies, they were still much poorer judges of sender deception than they were of sender truth (Hancock, Woodworth, & Goorha, 2010).

Identifying suspicion might be a better way to measure deception than straight judgments. In experimental research, study design usually induces suspicion in the receiver, though it also occurs without external (researcher) induction (Hancock et al., 2007). Recent evidence suggests that humans can detect lies at an unconscious level, even though their conscious judgments are poor (ten Brinke, Stimson, & Carney, 2014).

To detect fraudulent financial reporting, financial analysts have access external information sources to uncover potential misstatements, and they also have a venue (earnings calls) to probe managers. Assuming that analysts desire to know the true state of the firm they are analyzing, they should be motivated to investigate any information aberrations that may result from misreporting.

To understand why analyst behavior may change when managers are presenting a fraudulent story, Interpersonal Deception Theory (IDT; Buller & Burgoon, 1996) outlines expected behaviors of deceptive senders and suspicious receivers during an interaction. Specifically, IDT posits that a receiver's behavior will be influenced by his or her suspicions. Given that investment analysts are experts about a company and the industry in which that company operates, analyst behavior may change when interacting with fraudulent managers.

IDT argues that accuracy will increase with receiver informational and behavioral famil-

iarity (high in this context), their decoding abilities, and deviation of sender communication from expected patterns. IDT also predicts that suspicion will manifest through a combination of strategic and non-strategic behaviors. Questioning is largely strategic. One strategy that analysts can use to make more confident judgments is to select questions they believe are most important. We can observe question topic choices in this context. Early empirical work showed that receiver behavior changes, and a similar study showed that outside observers were more accurate judges than those who interacted (Buller, Strzyzewski, & Comstock, 1991; Buller, Strzyzewski, & Hunsaker, 1991).

There is exploratory empirical evidence that actively suspicious individuals may increase their use of probes (i.e. actions designed to obtain more information) possibly as an attempt to identify a liar (Burgoon, Wilson, Hass, & Schuetzler, 2016).

The use of more questions by analysts may indicate greater suspicion. This leads to first hypothesis:

H1: Analysts will ask more questions when interacting with fraudulent managers than when interacting with nonfraudulent managers.

Another indicator of suspicion and probing behavior could manifest in the topics of questions. Some aspects of narrative reports may be more likely to draw attention from analysts during the duration of a fraud. The novelty of latent topic modeling over financial documents makes it difficult to specify *ex ante* hypotheses on which topic areas are likely to draw suspicion. However, there several empirical results that can provide some comparison. Dechow et al. (2011) reported that 54% of firms identified in an AAER had misstated revenue, 27.2% capitalized costs as assets, and 25.1% misstated other expenses or shareholder equity accounts. Hoberg and Lewis (2017) found that fraudulent firms spent more time discussing

acquisitions, derivative financial instruments, growth, adverse results, incurred costs, and interest rates; and less discussion of management, gains on asset sales, legal proceedings, offsets, marketing expenses, and research & development. I leave the investigation of topics as an exploratory research question:

RQ1: Do analysts modify question topics interacting with fraudulent firms? That is, do analysts ask greater or fewer questions for some topics when interacting with fraudulent firms versus interactions with similar, nonfraudulent firms?

4.2.2   Time effects on suspicion

Because earnings calls are interactive and typically have many turns-at-talk, the time when an individual turn-at-talk occurs may be important.

If analysts grow more suspicious or uncertain as a call progresses, they may increase their use of probes. However, if analysts are satisfied with what they have heard or want to appear more favorable to managers, they may reduce the amount of questions they ask as a call progresses.

The interactive nature of deception and suspicion presents the possibility of time effects on behavior. Increased suspicion may increase strategic behavior of deceivers (Buller & Burgoon, 1996; Buller, Strzyzewski, & Hunsaker, 1991); that is, more probes lead to greater concealment on the part of deceivers. The effect of time on suspicion may reveal more information about questioning strategies. Within a call, a suspicious analyst may attempt to conceal suspicion by using fewer questions and more statements and phatic communication; they may also ask more questions to resolve uncertainties. While the direction is unclear, the revelation of new information during the call should lead to a change in questioning

behavior if suspicion is present.

H2: The presence of fraud moderates the relationship between the ratio of questions to statements as an earnings call progresses

It is also possible that the topics of discussion change when fraud is present. Analysts may prioritize certain areas early in a call, or they may defer until later. This leads to my second research question:

RQ2: Do certain topics become more or less prevalent during the progression of an earnings call when fraud is present, versus when fraud is not present?

4.3 Data

See Chapter 3, section 3.3 for the discussion on identifying fraud, obtaining earnings calls, and including financial data. The resulting sample includes 64 companies, with 32 fraudulent companies and 33 non-fraudulent companies. This is slightly larger than the sample in Chapter 3 because the requirement of a matching financial statement no longer applies. There are 120 earnings calls from fraudulent companies and 119 calls from non-fraudulent companies. The number of turns-at-talk, aggregated by fraud group and speaker type, are shown in Table 4.1. There were a total of 7704 analyst turns at talk in this dataset.

Table 4.1: Sample Data

Fraud Group	N utterances	N calls	N companies
Fraud	3845	120	32
No Fraud	3607	119	32

There were 3963 analyst turns-at-talk in fraudulent earnings calls, and 3741 turns-at-talk in non-fraudulent earnings calls.

4.3.1 Analyst-Company Overlap

Within this dataset, there are 82 analysts who cover multiple companies. There are 2823 utterances from these analysts. These utterances come from 154 earnings calls from 42 companies. 22 of these companies were fraudulent, and 20 were nonfraudulent.

I used Institutional Investor (II) analyst rankings as a measure of analyst skill. These rankings are good indicators of analysts with superior ability (Leone & Wu, 2007). I matched name variations (e.g "Timothy" and "Tim"). There are many analysts in the II database with no score, and those analysts that appeared in the calls but not in the II database also were marked with no score. There were 886 analyst-years in the sub-sample, with 494 having a ranking in Institutional Investor (see Table 4.2).

Table 4.2: Analyst-years Ranked by Institutional Investor

II	Count
Unranked	392
Ranked	494

4.4 Method

Each analyst turn-at-talk contains valuable linguistic and contextual information. The subsequent paragraphs describe the methods for extracting the number of sentences and questions in each analyst turn-at-talk and the creation and evaluation of latent topic models.

4.4.1 Identifying analyst questions

To measure information about style and syntax in each utterance, I used SPLICE (Moffitt & Giboney, 2011). This tool parses sentences and measures linguistic constructs. I retrieved

the total number of sentences in each utterance and the number of questions in each sentence.

The question ratio is the number of questions divided by the total number of sentences.

Table 4.3 shows the mean question ratio for analysts who covered multiple firms. The

question ratio measures the number of sentences an analyst chooses to ask questions, rather

than using greetings or discussing contexts.

Table 4.3: Question ratios for multi-firm analysts

Mean question ratio for analysts utterances in fraudulent and nonfraudulent calls N = 2820
utterances, 82 analysts. The difference in ratios is significant, $p < 0.001$

Fraud	Sentences	Questions	Question Ratio	N (utterances)
No Fraud	2.5519	0.9627	0.4285	1339.0000
Fraud	2.6023	1.0864	0.4774	1481.0000

4.4.2 Topic model creation

I use latent Dirichlet Allocation (Blei et al., 2003) to assign a topic to each turn at talk in

an earnings call. One key advantage of LDA is its ability to accurately assign topics to short

passages of text. The turns-at-talk in an earnings call can occasionally be brief, and LDA

can handle this.

The use of LDA has shown promise in identifying fraudulent firms by comparing the

topics they use in a 10-K MD&A form (Hoberg & Lewis, 2017). LDA takes a user-defined

number of topics k and a list of n documents. The algorithm then fits a model of latent

topics, and for each document estimates the probability distribution over the k topics.

This outputs top h words for k topics. Each utterance has a probability distribution

for each topic. For example, the most probable words (lowercase and stemmed) for a topic

may look like `share stock buy count buyback repurchas dividend sharehold`. For

an example turn-at-talk, "So, should you think this puts the stock buyback option off the table now, given this facility?" this turn-at-talk has a 58% probability of being related to the aforementioned topic. It is important to note that topics are not mutually exclusive in turns-at-talk. The example above may also be related to topics discussing assets, and also sense-making ("should", "think").

### 4.4.3 Topic Model Creation & Evaluation

To implement LDA I used MALLET (McCallum, 2002), an open source topic modeling tool. First, I cleaned the text by converting each utterance to lowercase, removed numbers and symbols, stemmed the words (using the Porter stemming algorithm), and removed stop words (function words and words with little semantic value due to frequent use). I used the standard English stop word list in MALLET. Standard practice in LDA modeling is to remove stop words, which will otherwise interfere with the quality and interpretability of the model. In addition to the standard word list, I removed domain-specific stopwords that occurred in more than 5% all utterances (e.g. "business", "finance").

The number of topics in LDA is user-defined. This requires some evaluation to determine a good number of topics. Perplexity is a standard method of measuring how well a language model fits out-of-sample data. The goal of this evaluation is to find a point where an increase in topic count yields only a trivial reduction in perplexity. I created five disjoint subsets with 20% of the utterances randomly assigned to each. At each topic level k, five models were created, with each using four subsets combined into a single training set and the fifth subset to test, such that each subset is used in testing exactly once. The number of topics ranged $k = 2$ to $k = 60$ in increments of two. Perplexity reduction becomes trivial around

25 topics, and reaches a minimum near 35. I use levels of k between 15 and 45 to reduce the likelihood that results are caused by a specific choice of k, and to assess topic stability across models.

Figure 4.2: LDA Fit

This figure shows perplexity scores for topic counts between 2 and 60. I used 5-fold cross-validation for each topic count. The perplexity is lowest between 25 and 40 topics.
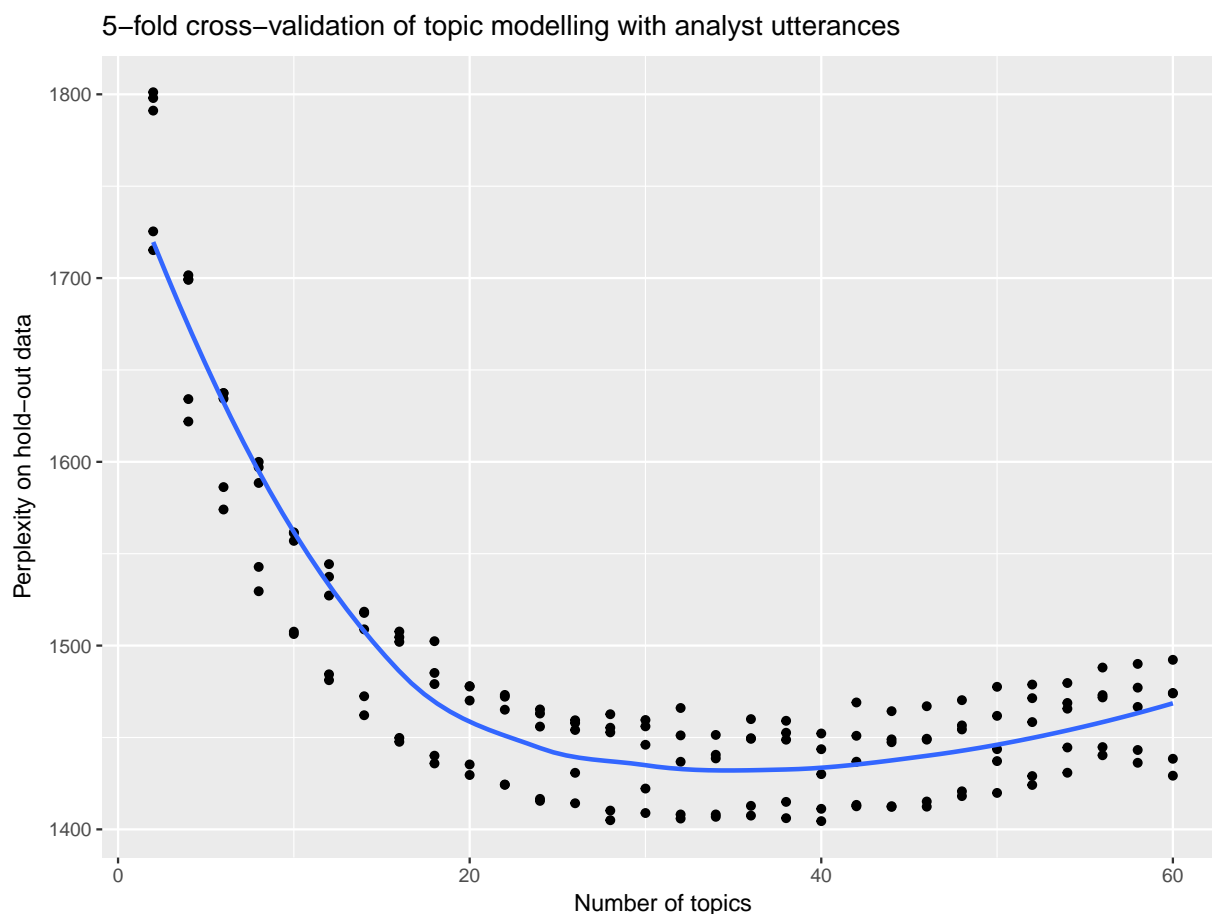
Based on perplexity scores, the optimal range for k lies between 25 and 45 topics. I chose to fit models with $k \in \{15, 25, 35, 45\}$. The level of $k = 15$ was included to assess the trade-off between a higher perplexity and fewer topics to interpret. Levels above 45 increase perplexity, produce more topics to interpret, and increase the risk of overfitting. The training corpus included all analysts utterances in the dataset. I fit the model with

5,000 Gibbs sampling iterations. These topic counts are similar to the 31 topics used by N. C. Brown, Crowley, and Elliott (2018), and somewhat less than Hoberg and Lewis (2017), who used 75 topics.

4.4.4 Contextual Variables

In addition to data extracted from *what* gets said, I also capture information about the within-call context surrounding each turn-at-talk. One measure is the location in the Q&A sequence (total number of analyst turns at talk), and the sequence within an analyst's time on the call (e.g. from 1 to 3 for an analyst who had three turns-at-talk).

4.5 Analysis

After extracting linguistic measures and matching them to the corresponding financial and analyst variables, I constructed regressions to test H1 and H2. Similar regressions are used to test for topic differences between fraudulent and non-fraudulent firms. All regressions use the subsample of analysts who cover at least one fraudulent and at least one nonfraudulent firm in my sample.

I construct two base models (Model 1 and Model 2 in Table 4.4) to test the effects of fraud and position in the Q&A before controlling for financial variables. These models show a positive, significant main effect for fraud. The question ratio steadily declines as calls progress; however, the presence of fraud mitigates much of this effect.

Table 4.4 shows the results of the regressions testing differences in question ratio in each utterance. Not all periods had full financial data, so 485 turns-at-talk were omitted from Model 4.

The regressions show a strong main effect of fraud on the ratio of questions per utterance. The final model shows an increase of 10.90% in the question ratio per analyst turn-at-talk between fraudulent and non-fraudulent firms ($p = 0.0376$), supporting H1.

An analyst turn at talk contains more questions when interacting with managers later found guilty of fraud than when interacting non-fraudulent industry peers.

4.5.1 Analysis of time effects

This model also tests the second hypothesis on time effects of questions. The coefficient of the interaction between fraud and position in the Q&A is positive and and significant, supporting H2. In Model 4, the interaction with fraud negates the effect of time. When discussing with deceptive managers, analysts persist in asking questions for the duration of the call. Though the coefficient is small, fraudulent calls averaged 32.0 analyst turns-at-talk, and nonfraudulent calls averaged 30.3 analyst turns-at-talk. Turns-at-talk with nonfraudulent managers had approximately 9.9% fewer questions at the end of the average-length call, but approximately 1.5% *more* at the end of the average call involving fraudulent managers. The ratio of questions to the total number of sentences in an utterances decreases as a call progresses, but the presence of fraud negates much of this decrease. This suggests that analysts may be more willing to risk social capital by using more probing than normal.

In addition to the time effect for all participants, there was a significant three-way interaction between fraud, call sequence, and II ranking. Analysts in the II rankings asked significantly fewer questions as calls progressed.

Table 4.4: Question Ratio Regressions

Regressions testing the effect of fraud on the ratio of questions per analyst turn-at-talk. Dependent variable is question ratio in each model. Model 1 includes just the effect of fraud. Model 2 adds the sequence of the call when a turn-at-talk occurred Model 3 includes analyst fixed effects and Institutional Investor rankings, and Model 4 adds financial controls, industry fixed effects, and fiscal period fixed effects.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	**0.4249*****	**0.4271*****	**0.5677*****	**1.0744*****
	(0.0099)	(0.0099)	(0.0754)	(0.3114)
fraud	**0.0497*****	**0.0474*****	**0.0577***	**0.1090****
	(0.0137)	(0.0137)	(0.0249)	(0.0392)
qa.seq		**−0.0023*****	**−0.0044*****	**−0.0035****
		(0.0005)	(0.0011)	(0.0013)
fraud:qa.seq		**0.0020****	**0.0034***	**0.0035***
		(0.0008)	(0.0013)	(0.0016)
ii			−0.0691	−0.1005
			(0.0591)	(0.0651)
fraud:ii			−0.0163	−0.0111
			(0.0310)	(0.0394)
qa.seq:ii			**0.0024†**	**0.0026†**
			(0.0013)	(0.0015)
fraud:qa.seq:ii			**−0.0032†**	**−0.0048***
			(0.0017)	(0.0020)
a.seq				−0.0060
				(0.0046)
loss				0.0321
				(0.0274)
fraud:a.seq				0.0034
				(0.0065)
Analyst FE			Included	Included
Financials				Included
R^2	0.0047	0.0114	0.0700	0.1109
Adj. R^2	0.0043	0.0103	0.0399	0.0739
Num. obs.	2814	2814	2808	2228
RMSE	0.3623	0.3613	0.3557	0.3495

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^\dagger p < 0.1$

Figure 4.3: Three-way interaction between fraud, time, and II ranking

Regression lines on the residualized question ratio. Question ratio was regressed on analyst fixed effects and financial control variables. Analysts who were not in the II rankings asked more questions of fraudulent managers as calls progressed, and fewer questions of non-fraudulent managers as calls progressed. Analysts who were ranked by II actually asked fewer questions of fraudulent managers as calls progressed.
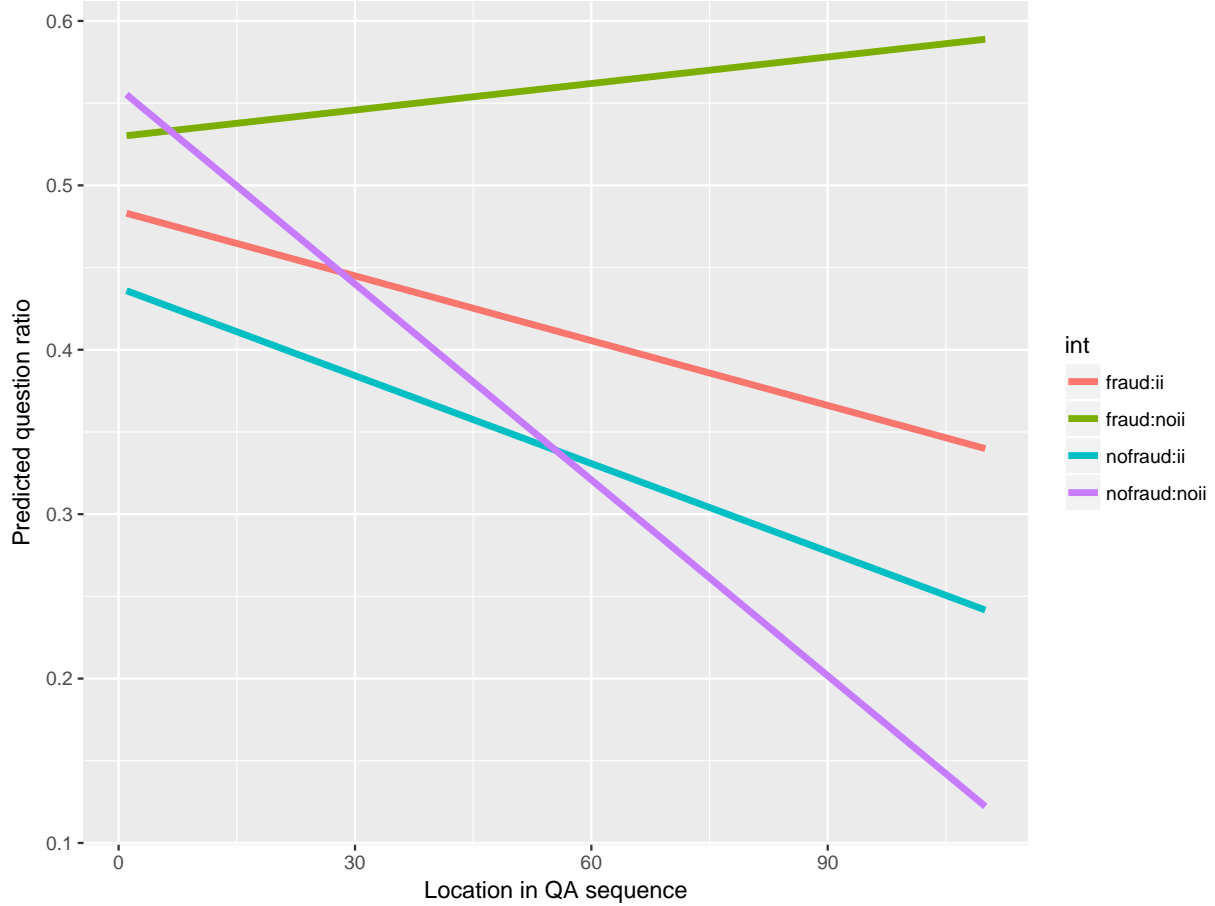
4.5.2 Changes in topic composition

RQ1 proposed a test for different topic uses between fraud and non-fraud companies, I use a method similar to Hoberg and Lewis (2017). This method treats topic usage as a function of the presence of fraud and financial variables. I estimate regressions for each topic h within each level of k, with the topic probability as the dependent variable, and for each company i at time t a fraud indicator variable with other controls as the predictors. This will show which topics are over- or under-represented in analyst questions during the presence of fraud.

If the fraud variable is significant, this shows that for a given topic, there are differences in the number of times that this topic is discussed between fraudulent and non-fraudulent companies.

For each topic, I fit a linear regression with analyst fixed effects, II rankings, and financial control variables for the firm in each period. The regressions are the same as Model 4 of the question ratio model, though the dependent variable changes.

Table 4.5 shows the the number of topics that were mentioned significantly more often (M+) and significantly less often (M-) in analyst interactions with fraudulent managers. The number of topics that differed in the presence of fraud was significantly higher than random chance, with approximately 20% of topics either more or less frequent.

Table 4.5: Significant topic differences versus expected (main effects)

Count of topics that with significant differences in frequency between fraudulent and nonfraudulent firms. M+ (M-) indicates the count of topics that had a significantly higher (lower) main effect of fraud. Main effects were included for $p < 0.05$.

k-level	Sig, M+	Sig, M-	Expected	Actual	Percent
15	2	1	0.75	3	**20.00%**
25	3	2	1.25	5	**20.00%**
35	3	5	1.75	8	**22.86%**
45	4	5	2.25	9	**20.00%**

There were also a number of topics that showed interaction effects between fraud and call sequence (Table 4.6). RQ2 proposed an exploration of changes in topic frequency over the course of a call, specifically looking for different rates of change between fraudulent and nonfraudulent firms. Like the main effect table, there are consistently more interaction effects than chance. This suggests that analysts varied the locations on some topics depending whether or not fraud is present.

Table 4.6: Significant topic differences versus expected (interactions)

Count of topics that with significant changes as a call progresses, interacted with fraud. I+ (I-) indicates the count of topics that increased (decreased) in frequency as a call progressed when fraud was present. Interaction effects were included for $p < 0.05$.

k-level	Sig, I+	Sig, I-	Expected	Actual	Percent
15	1	1	0.75	2	**13.33%**
25	2	1	1.25	3	**12.00%**
35	3	2	1.75	5	**14.29%**
45	2	4	2.25	6	**13.33%**

Table 4.7 highlights the topics that had significant differences in the frequency of discussion between fraudulent and non-fraudulent calls. It also contains the top words for each topic. To verify that a topic was the same between multiple levels of k, I compared the word probability vectors for each topic in the larger model to each topic in the smaller model (e.g. 35 vs. 25) using cosine similarity. I use this to identify the topics that have the highest similarity to each other across the models. For example, topic 16 in 4.7, the similarity score between the topic when $k = 45$ and when $k = 35$ is 0.95, $k = 35$ vs. $k = 25$ is 0.93, and $k = 25$ vs. $k = 15$ is 0.92. The mean similarity for all pairs is just 0.06 when comparing $k = 45$ to $k = 35$, 0.07 when comparing $k = 35$ to $k = 25$, and 0.11 when comparing $k = 25$ to $k = 15$, which provides additional validity to this measure by showing that random topic comparisons have very little similarity, while some topics show markedly higher similarity between different models.

4.6 Conclusion

Analysts provide information that market participants can use to make decisions. Considering their substantive interactions, analysts may play a role in uncovering (or prolonging) a fraud.

Analysts may provide a signal for fraud detection. This research also presents a method for analyzing the analysts' questions in an earnings call, which is an element of narrative disclosures that has received little attention in the literature.

This work builds on the growing body of literature supporting analysts' abilities to reduce information asymmetry between companies and investors. There remains a considerable amount of work to understand why some analysts can extract this hidden information. Analyst skill is often measured by forecasting ability. Other dimensions of skill are report readability (Franco, Hope, Vyas, & Zhou, 2013), objectivity, and rigor (Salzedo, Young, & El-Haj, 2014).

This study shows evidence of suspicion in analyst turns-at-talk. When interacting with companies later found guilty of fraud, analysts ask more questions per turn-at-talk, and maintain this orientation toward questioning through the progression of the call. Experimental research in this area (Buller, Strzyzewski, & Hunsaker, 1991; Hancock et al., 2007) showed that suspicious receivers in an interpersonal interaction would make greater use of probes. This study is perhaps the first to demonstrate this phenomenon in a real-world setting. Another contribution of this study comes from the fact that analysts have knowledge and resources at their disposal, and repeated interactions with deceivers.

When fraud is present, analyst turns-at-talk are more likely to relate to reserves, credit, and charge-offs, settlements, litigation, and net income. They are less likely to relate to balance sheets, yields, mortgages, leases and renting, and stock repurchases. I fit topic models with the number of topics ranging from 15 to 45, and showed that the topics that changed were relatively stable and robust to varying topic counts, and that the results are not likely to be caused from a specific choice of k.

There are many avenues for future research in this area. The content and style of analyst turns-at-talk may also be predictive of events other than fraud, like profitability, share price performance, and uncertainty. Latent topic models open the possibility of more fine-grained, automated analysis of text at more granular level than full-period analyses. A single turn-at-talk from a manager or a single paragraph in an MD&A could be assigned a topic and through the topic label be associated to a variable of interest. Instead of spending effort toward manually coding each turn-at-talk or paragraph, a researcher could simply associate a generated topic to a variable of interest (the topic Q in this study could easily be linked to stock issuance and buybacks).

In addition, suspicion may affect truthful communications because those who have suspicion view senders as less trustworthy and less competent (Hubbell, Mitchell, & Gee, 2001).

Table 4.7: Top words for topics with significant differences

Top words for topics that were significantly different for at least one level of k. Topics across models were matched based on the similarity of the top 10 words. M+ (M-) indicates a topic that appeared significantly more (less) often in analyst turns-at-talk with fraudulent firms. Several topics in this table did not have a significant main effect, but did show an interaction between time and fraud.

Topic	Top Words	k = 15	k = 25	k = 35	k = 45	Count
Topic 1	advertis financi categori ad site promot sell traffic network vertic			M-		1
Topic 2	asp fiber provid auto mobil laserscop laser card ross direct				M-	1
Topic 3	balanc yield secur sheet portfolio basi liquid accret deposit asset		M-	M-	M-	3
Topic 4	capit asset basel ratio iii phase bank target debt action	M+				1
Topic 5	consult financi corpor legal work vlociti restructur job final big				M+	1
Topic 6	contract project branch mani leas develop place state increas rent					0
Topic 7	credit portfolio commerci real construct estat loss home card due			M+		1
Topic 8	europ north engin america truck fleet trend plant american countri				M+	1
Topic 9	inventori order retail product end level cancel channel backlog plant					0
Topic 10	leas rent develop place occup properti capex tenant portfolio bear			M-	M-	2
Topic 11	morn color follow comment provid sound answer detail made miss					0
Topic 12	mortgag origin servic gain bank volum side industri hedg pipelin		M-	M-	M-	3
Topic 13	net incom interest line nii trade level earn dollar normal		M+		M+	2
Topic 14	project contract state backlog construct start pipelin review specif delay					0
Topic 15	reserv credit portfolio loss commerci real charg construct provis chargeoff	M+	M+	M+		3
Topic 16	reserv settlement claim issu litig label privat exposur risk agreement		M+	M+	M+	3
Topic 17	share card repurchas credit stock count issu gain buyback book	M-			M-	2
Topic 18	valu spread mortgag hedg gain basi book secur asset today			M-		1
Total		3	5	8	9	

Chapter 5

LANGUAGE STYLE MATCHING

5.1 Motivation

This research explores the linguistic characteristics of dyadic interactions between analysts and managers in earnings calls to identify signs of dominance and language style matching. Several lab experiments have shown that dyadic discourse involving a deceptive individual results in language that is different in style and content from similar, truthful dyads. This paper covers the literature in this area, followed by descriptions of the data and identification of fraud. I then describe the analysis methods and results. The earnings calls of fraudulent firms tend to contain more negativity and less dominance than calls from nonfraudulent firms. These effects are likely driven by analysts, since managers show the opposite trend: less negativity and more dominance. The paper concludes with a discussion of next steps and future research.

5.2 Literature Review

The literature on deception in dyadic discourse is sparse, especially when using real-world data. The question-and-answer portion of earnings calls results in dyadic communication between analysts and managers. These two groups often have competing goals. Managers want to portray the company in a favorable light (or as legitimate in the case of fraud), and analysts want to discover the true state of the firm. Dyadic power theory (DPT; Dunbar,

2004) provides a structure and predictions to evaluate manager-analyst interactions. While this theory was constructed in the context of marriages and familial relationships, it refers to "interactants have established history, are dependent on one another for outcomes, and expect to have a continued interaction in the future." (Dunbar, 2004, p. 243) There are many parallels between contexts: familiarity between parties, normative behaviors, observable communications, and situations with incongruent goals.

Managers can use power processes to achieve their goals of successfully deceiving analysts and the market. Powerful language is viewed more favorable (Holtgraves & Lasky, 1999). Managers might also seek to equate their dominance displays to analysts to improve relationship satisfaction (Dunbar, 2004).

Power bases form the foundation for control over others. Managers have the power to reward or punish analysts by controlling their ability to participate in calls (Cohen, Lou, & Malloy, 2013; Mayew, 2008). Managers also have informational power, since the information they hold is of great value to analysts. Analysts have expert power (knowledge of the industry) and referent power (investors follow and take actions based on analyst reports). From an outside perspective, the relative power of each party appears to be nearly equal, perhaps with a slight edge for managers. DPT predicts a curvilinear relationship between perceived relative power and control attempts. The more equal partners perceive their relative power, the more control attempts a partner will make.

When discussing fraud, managers may cede power, especially early in a call, to reduce perceived threat and increase engagement and persuasiveness. Zhou, Burgoon, Zhang, and Nunamaker (2004) show that liars assume a submissive role early in conversations and increase dominance behavior as an interaction progresses. Individuals may also use more

certainty when lying about beliefs (Mihalcea & Strapparava, 2009). I propose to examine these mechanisms in the context of financial fraud. I will answer two questions. First,

Do fraudulent managers use more dominant language than nonfraudulent managers, and does the use of dominant language change over time?

The second question relates to language style matching. Niederhoffer and Pennebaker (2002) hypothesize language style matching is a strategic behavior to increase engagement with a conversation partner. Language style matching between liars and their conversation partner is highest during a lie and when the deceiver is highly motivated (Hancock et al., 2007). Hancock et al. (2007) go on to say "When motivated liars were lying, liars and partners matched their rates of generation for words, second- and third-person pronouns, negative affect, and negations ... [t]he data in this study suggest that motivated liars may have used LSM as a strategy to appear more credible to their partner."

Do fraudulent managers match the linguistic style of analysts more than nonfraudulent managers?

5.3 Data

The data from this study is similar to that of Chapter 4, but it also includes manager turns-at-talk, in addition to those from analysts. I use data from 239 earnings calls (Table 5.1). There were 120 calls from companies that were committing fraud, and 119 calls from similar companies that were not committing fraud.

I use the question-and-answer (Q&A) portion of each call. There are a total of 16,626 turns-at-talk from both analysts and managers (Table 5.2). The Q&A portion of calls was nearly equal for both fraudulent and non-fraudulent firms (Table 5.3).

Table 5.1: Number of calls by group

fraud	Count
Fraud	120
No Fraud	119

Table 5.2: Number of Utterances by Group and Speaker

Fraud Group	Speaker Role	Turns-at-Talk
fraud	Analyst	3963
nofraud	Analyst	3741
fraud	Executive	4908
nofraud	Executive	4886
Totals		17498

5.4   Method

I quantify the earnings call transcripts by extracting linguistic features, identifying com-
munication dyads, and computing language similarity for each interaction. First, I extract
linguistic measures of dominance using SPLICE, a tool to extract linguistic features from
text (Moffitt & Giboney, 2011). A sentence increases in dominance when it uses inflexible
phrasing ("I will not"), contradictions ("that is false"), knowing ("I am aware"), positive
self-evaluation ("I am smart"), and "I can" phrasing. A sentence decreases in dominance
when seeking guidance ("how can I") or permission ("Can I do"), when using "don't know"
phrasing, and negative self-evaluation.

SPLICE also extracts function words from sentences using NLTK (Bird et al., 2009) and
custom word lists. I use linguistic style markers from Ireland et al. (2011). These features
are articles, prepositions, conjunctions, negations, personal pronouns, auxiliary verbs, and
adverbs.

I then identify analyst-manager dyads from Q&A call segments. I define an interaction

Table 5.3: Mean number of turns-at-talk in the Q&A section

fraud	position
No Fraud	69.66
Fraud	70.53

as an analyst turn-at-talk that is followed by a manager turn-at-talk. If two managers respond to an analyst, this will count as two interactions.

Next, I measure linguistic style matching (LSM) by comparing the style measures of each speaker in each interaction, which gives an interaction-level similarity score (LSM_SCORE). I use the similarity formula from (Ireland et al., 2011). The formula (Equation 5.1) computes the similarity score for each measure, and the final LSM_SCORE is the mean of all style measures. The example in Eq. 5.1 shows the measure for word counts.

$$LSM_{words} = 1 - \frac{|words_A - words_E|}{words_A + words_E} \tag{5.1}$$

The mean similarity is 0.5211 for all interactions. There were 8975 total analyst-manager interactions (Table 5.4)

Table 5.4: Number of turns-at-talk interactions by group

fraud	uniquecall
fraud	4480
nofraud	4495

5.5   Results

I used linear mixed effects regressions to test the relationship between fraud and linguistic measures of dominance. The models included levels for speaker and company (*tic*). The

model includes analyst turns-at-talk to control the influence that analysts may have on manager behavior.

Table 5.5 shows the regression results for dominance behavior. The coefficients in the quantity regression represent the change in the number of words. In the latter four regressions, the coefficients represent the change in the percentage of words used. A positive coefficient cf 0.01 indicates a change in the proportion of the dependent variable by 1% in a turn at talk (e.g. from 3.5% negative words to 4.5% negative words).

There was no effect of fraud on quantity. Negativity was higher for all participants when fraud was present. Managers of fraudulent firms used less negativity in their responses to analyst turns-at-talk. There were no significant differences for positivity or hedging. There were significant main effects for fraud on negativity and dominance. All participants on a call used more negative language and less dominant language.

Table 5.6 shows the regression results of LSM on financial fraud. Managers of fraudulent firm showed no significant differences with managers of non-fraudulent firms in matching linguistic style of analysts during earnings calls. The managers of fraudulent firms had a slight (marginally significant, $p = 0.0688$) decrease in LSM as a call progressed. The calls in my sample had a mean of 37.5 analyst-manager interactions, implying that fraudulent managers' language similarity scores decreased by 1.8% by the end of a call.

5.6   Conclusion & Future Work

The purpose of this research was to investigate the use of dominance and linguistic style matching from managers of fraudulent and non-fraudulent earnings calls. Research in this domain have shown that managers and analysts both have power in earnings calls, and this

affects call participation (Cohen et al., 2013; Mayew, 2008). Research on the behavior of deceptive parties show that dominance displays may manifest in verbal behavior; submissive early, and increasing in dominance as an interaction progresses (Zhou et al., 2004).

I find that calls involving fraudulent firms use more negative language, yet the managers of fraudulent firms use significantly less negativity than their non-fraudulent peers. The calls of fraudulent firms also have significantly less dominance on the whole; yet managers of fraudulent firms tend to use overall more dominance, though this was only marginally significant ($p = 0.065$).

This study does not find meaningful differences in linguistic style matching between deceivers and truthtellers. Additional work should validate the measures of linguistic style by comparing an analyst's language in calls with other companies. It should also compare manager style to the prepared remarks in the opening portion of the calls.

Future research in this study should include more variables related to analyst-manger interactions, including analyst ratings and recommendations, and measures of analyst skill. It should also compare alternative methods of linguistic style matching, such as those from (Hancock et al., 2007). I also intend to measure semantic similarity. Managers, when faced with unexpected scrutiny from analysts, tend to revert to scripted responses to analysts' questions (Lee, 2016). Semantic similarity can reveal off-topic or avoidant responses.

Table 5.5: Regression results for language dominance

Regressions testing the effect of fraud on five linguistic measures of dominance. Each model contains the the sequence of the call when a turn-at-talk occurred, financial controls, industry fixed effects, and fiscal period fixed effects.

	Quantity	Negativity	Positivity	Hedging	Dominance
(Intercept)	3.8307	**0.0293*****	**0.0499*****	**0.0508*****	**0.1368*****
	(11.1487)	(0.0042)	(0.0096)	(0.0067)	(0.0276)
fraud	0.8848	**0.0027***	0.0017	−0.0003	**−0.0213***
	(3.4244)	(0.0013)	(0.0027)	(0.0021)	(0.0086)
qa.seq	**−0.0633**†	0.0000	0.0000	**−0.0001*****	**−0.0006*****
	(0.0355)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
manager	**38.7380*****	**0.0025***	**0.0058**†	−0.0017	**−0.0857*****
	(2.7246)	(0.0012)	(0.0033)	(0.0015)	(0.0073)
a.seq	**−0.4600*****	0.0000	−0.0002	**−0.0002****	**−0.0011****
	(0.1344)	(0.0001)	(0.0001)	(0.0001)	(0.0004)
fraud:qa.seq	−0.0019	−0.0000	0.0000	0.0000	0.0000
	(0.0505)	(0.0000)	(0.0001)	(0.0000)	(0.0002)
fraud:manager	−3.5224	**−0.0033***	−0.0023	0.0017	**0.0173**†
	(3.6786)	(0.0015)	(0.0045)	(0.0019)	(0.0094)
qa.seq:manager	−0.0889	−0.0000	0.0000	**0.0001*****	**0.0009*****
	(0.0549)	(0.0000)	(0.0001)	(0.0000)	(0.0002)
fraud:qa.seq:manager	0.0726	−0.0000	−0.0001	−0.0000	−0.0002
	(0.0655)	(0.0000)	(0.0001)	(0.0000)	(0.0002)
Num. obs.	15083	15083	15083	15083	15083
Num. groups: speaker	932	932	932	932	932
Num. groups: tic	60	60	60	60	60
R^2	0.1659	0.0176	0.1529	0.0502	0.0735

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^\dagger p < 0.1$

Table 5.6: Language similarity regression results

Regressions testing the effect of fraud on LSM as measured in Ireland et al. (2011). Each model contains the sequence of the call when a turn-at-talk occurred, financial controls, industry fixed effects, and fiscal period fixed effects.

	LSM_SCORE
(Intercept)	**0.5613*****
	(0.0398)
fraud	−0.0007
	(0.0130)
interactionIndex	−0.0000
	(0.0002)
fraud:interactionIndex	**−0.0005**†
	(0.0003)
Num. obs.	8078
Num. groups: speaker_A	660
Num. groups: tic	60
R^2	0.0435

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, $^{\dagger}p < 0.1$

Chapter 6

CONCLUSION

The purpose of this dissertation was to identify linguistic behaviors from managers and analysts that differed when fraud was present versus when fraud was absent. Research in this area has largely relied upon text analysis methods that have little theoretical significance (bag-of-words), or are very far from the phenomenon of interest. A nuanced approach that explicitly aligns the level of theory to the level analysis can overcome some of these limitations. The literature review chapter in this dissertation outlines the structure of narrative financial disclosures, and the subsequent studies use this information to design and test for signals of fraud in linguistic content and style.

Chapter 3 contains a study of incremental information between earnings calls and the corresponding financial statements. In this study I find that the CFOs of fraudulent companies tend to re-use more language between the prepared remarks of an earnings call and the MD&A than the CFOs of similar, nonfraudulent companies. Chapter 4 uses Interpersonal Deception Theory as a foundation to evaluate analyst scrutiny when interacting with managers of both fraudulent and nonfraudulent firms, finding that analysts use more probing behavior when interacting with managers of fraudulent firms. I also develop and evaluate latent topic models to understand the areas that analysts are more or less likely to address when interacting with fraudulent firms. A quasi-experimental research design in this study should find use in future work involving financial analysts and their interactions

with managers. The third study, in Chapter 5, considers the language of both analysts and managers, and found evidence that fraudulent managers use less negativity and more dominance language than their nonfraudulent peers. There were no significant differences in the level of linguistic style matching managers use when interacting with analysts.

6.1 Limitations and Future Research

There are several important limitations to consider in this work. First, the managerial behaviors associated with fraud may not be caused by lying, but rather by some unobserved traits that predispose managers to committing fraud. The use of analysts and a quasi-experimental design reduces some of this risk, and is worth considering in future work. Another limitation is the low prevalence of detected fraud. This not only limits the number of firms labeled fraudulent, but also leaves the possibility of fraudulent firms avoiding detection and appearing as a control firm in this sample. A third important limitation comes from the small number of analysts participating in earnings calls that also have matching forecasts and recommendations in IBES. Not all analysts report forecasts to IBES, and many who do lack enough information to link with earnings call participation. This may become less of an issue in research concerned with areas other than fraud, since sample size has fewer restrictions.

The literature review and empirical studies in this dissertation helped identify many other interesting and important avenues for future research. Latent topic modeling can facilitate automated annotation of text passages related to specific financial measures. For example, on could identify MD&A paragraphs dedicated to discussing return on assets and use tone or uncertainty in the paragraph to predict future performance on this measure.

This dissertation and Davis and Tama-Sweet (2012) compare studies across multiple disclosures. Researchers have yet to explore the relationships between other firm-produced narratives. This style of inquiry could yield valuable insights on how firms modify reports based on outside impressions.

The earnings call environment provides a standardized, multimodal, and interactive communication channel. With advanced transcription services, like IBM Watson Speech-to-text, it may become possible to identify turns-at-talk in the audio from an entire earnings call. The high costs of doing this manually limit current research to a few thousand observations that may only cover a few minutes of discourse (Mayew & Venkatachalam, 2012).

A recent innovation in this area is the video-based earnings call. Netflix conducts the analyst Q&A using video chat, and makes the videos publicly available. Other companies will post brief interviews with top managers regarding earnings. If the video-based earnings discussion becomes a common occurrence, facial and eye movements may be a valuable line of inquiry on human behavior.

Another avenue for future research concerns countermeasures. As research on narrative reporting grows in visibility and sophistication, firms may use techniques to obfuscate or exaggerate linguistic or behavioral cues that regulators, analysts, and investors use to make decisions. A study that replicates the methods in published findings could determine the extent to which managers adjust their disclosure style to meet presentation goals. With most published studies, there is a gap between the last year used for analysis (e.g. 2008) and publication (e.g. 2012) that would enable validation of the original study's results. Data after the publication date (e.g. 2013-present) would allow for countermeasure identification.

Appendix A

Analyst Data Processing

A.1 Data Challenges

Analyst data comes from several sources that are not linked together.

These data sources are:

- Seeking Alpha (Transcripts)

- I/B/E/S (Forecasts and recommendations)

- Institutional Investor (Rankings)

In order to control for analyst characteristics in my models, I must link these data sources. The only way to link these sources is through text-based matching. Each data source formats names differently.

The Seeking Alpha (SA) data contains multiple issues. First, the names in the transcripts are occasionally misspelled (e.g. Sims vs. Simms). In other cases, analysts use abbreviated name forms (e.g. Charles vs. Chuck). I resolve these issues by identifying each unique name in the earnings calls data, and then manually match variations of name based on last name, derivations of first name, employer, and the company hosting the call where the analyst appears.

Data from IBES contains analyst names in the format of [Last Name][First Initial]. IBES also issues each analyst a unique ID and identifies the firm the analyst represents.

Institutional Investor (II) data contains the first name and last name of the analysts and their employers.

Matching analyst names from these data sources without a tool to assist the process is inefficient at best and infeasible at worst. I developed a program that automates some of the manual tasks undertaken in Cohen et al. (2013) to make this process manageable. I imported the data from each source into a MySQL database tables. Another table stores the "translations", i.e. mappings between the data sources.

The program finds unique names from the SA and II tables and returns the five closest matches from IBES based on Levenshtein distance between the names. Levenshtein distance measures the minimum number of character edits required to change one string to match another. The program displays the base name and affiliation followed by a list of the closest name matches (see A.1). After the user chooses a match (otherwise, the user may choose "0" to report no match), the program will insert the mapping into the translation table.

This program facilitated the mappings between approximately 780 distinct analyst names from Seeking Alpha to IBES in about three hours, or 14 seconds per distinct analyst name. There were 368 names from Seeking Alpha that matched with a name in IBES.

Figure A.1: Analyst Name Matching Tool

This is a screenshot that shows the console interface to the program that assists the process of matching names from earnings calls, IBES, and Institutional Investor. For a target name, it lists the most similar names from IBES recommendations database.

Appendix B

Control Variables

B.1 Control variables

From Larcker and Zakolyukina (2012, p.31)

- Actual issuance ($capmkt$)

  An indicator variable coded 1 if the firm issues securities or long-term debt ($SSTKQ >$ 0 or $DLTISQ > 0$) and 0 otherwise.

- Seasonal change in cash sales ($sch.cs$)

  $((SALEQ_t - RECTQ_t) - (SALEQ_{t-4} - RECTQ_{t-4}))/(SALEQ_{t-4} - RECTQ_{t-4})$

From Dechow et al. (2011), also in Larcker and Zakolyukina (2012)

- Seasonal change in receivables ($sch.rec$)

  $(RECTQ_t - RECTQ_t - 4)/savgATQ$, where $savgATQ = (ATQ_t + ATQ_{t-4})/2$.

- Seasonal change in inventory ($sch.inv$)

  $(INVTQ_t - INVTQh_{t-4})/savgATQ$, where $savgATQ = (ATQ_t + ATQ_{t-4})/2$.

- Soft assets ($soft.assets$)

  $(ATQ_t - PPENTQ_t - CHEQ_t)/ATQ_t$.

- Seasonal change in ROA ($schroa$)

  $ROAt - ROA_{t-4}$.

- Seasonal days' sales in receivables index ($sdsri$)

  $(RECTQ_t/SALEQ_t)/(RECTQ_{t-4}/SALEQ_{t-4})$.

- Seasonal gross margin index ($sgmi$)

  $((SALEQ_{t-4} - COGSQ_{t-4})/SALEQ_{t-4})/((SALEQt - COGSQ_t)/SALEQ_t)$.

- Seasonal asset quality index ($saqi$)

  $(1 - (ACTQ_t + PPENTQ_t)/ATQ_t)/(1 - (ACTQ_{t-4} + PPENTQ_{t-4})/ATQ_{t-4})$.

- Seasonal sales growth index ($ssgi$)

  $SALEQ_t/SALEQ_{t-4}$.

- Seasonal sales, general, and administrative expenses ($ssgai$)

  $(XSGAQ_t/SALEQ_t)/(XSGAQ_{t-4}/SALEQ_{t-4})$.

- Seasonal leverage index ($slvgi$)

  $((DLTTQ_t + LCTQ_t)/ATQ_t)/((DLTTQ_{t-4} + LCTQ_{t-4})/ATQ_{t-4})$.

Appendix C

Quasi-experimental design

Arguments and rebuttals against the randomness of firms in an analyst's coverage basket:

- The industry the analyst covers may have greater or fewer fraudulent firms

  a. This means it is more likely that the analyst will cover a fraudulent firm, therefore changing the distribution of fraudulent firms in her sample, but this is the result of the industry and not the analyst.

  b. For example, take an experiment with an interviewer: Send participants to this interviewer, some who lie and others who do not. Ex ante, the interviewer has no way of knowing whether a participant is a liar. The interviewer has no impact on liars.

  c. The industry difference means that you would be modifying the base rate, but not the assignment to treatments.

- Fraudulent firms may self-select into being covered by an analyst

  a. Firms likely have little control over who covers them.

  b. Managers can choose to omit analysts who scrutinize management (Cohen et al., 2013; Mayew, 2008). This practice seems likely to occur in this context, since those showing the greatest signs of suspicion may be excluded from participation. That analysts still show signs of increased scrutiny despite the potential consequence of exclusion from participation may strengthen the findings of this study.

Explanation of strengths of current method

- Normal fraud research uses fraud as the treatment, the subject being the fraudulent firm and their managers

- In this approach, the analysts receive the treatment. They are blind to the treatment.

Experimental research with similar designs

- Jensen, Lowry, and Jenkins (2011) had participants view ten videos, five of which contained deception and five that were honest. Participants then made judgments of the interviewee. They did not measure a deterministic judgment on truth or deception, but rather a 7-point scale on how suspicious the participant was of what the interviewee said. The main difference in this study is that it is not interactive, and it uses survey measures rather than behavioral signals from the judges.

- Levine, Park, and McCornack (1999) conduct multiple experiments where people are asked to judge. They manipulate suspicion, use of probes, and base rates.

References

Abbasi, A., Albrecht, C., Vance, A. O., & Hansen, J. V. (2012). Metafraud: A meta-learning framework for detecting financial fraud. *MIS Quarterly*, *36*(4), 1293–1327.

Allee, K. D., & Deangelis, M. D. (2015). The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research*, *53*(2), 241–274. doi:10.1111/1475-679X.12072

Association of Certified Fraud Examiners. (2013). The SEC's newest tool is headed to the front. *Association of Certified Fraud Examiners*. Retrieved December 18, 2014, from http://www.acfe.com/fraud-examiner.aspx?id=4294979300

Association of Certified Fraud Examiners. (2014). *Report to the Nations on Occupational Fraud and Abuse: 2014 Global Fraud Study*. Association of Certified Fraud Examiners.

Barman, S., Pal, U., Sarfaraj, M. A., Biswas, B., Mahata, A., & Mandal, P. (2016). A complete literature review on financial fraud detection applying data mining techniques. *International Journal of Trust Management in Computing and Communications*, *3*(4), 336–359. Retrieved from http://www.inderscienceonline.com/doi/abs/10.1504/IJTMCC.2016.084561

Bauguess, S. W. (2016). *Has big data made us lazy?* Retrieved February 15, 2017, from https://www.sec.gov/news/speech/bauguess-american-accounting-association-102116.html#_edn9

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, *55*(5), 24–36.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *the Journal of machine Learning research*, *3*, 993–1022.

Bloomfield, R. (2012). Discussion of detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*(2), 541–552.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214–234.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "Black Box" of sell-side financial analysts. *Journal of Accounting Research*, *53*(1), 1–47. doi:10.1111/1475-679X.12067

Brown, N. C., Crowley, R., & Elliott, W. B. (2018). *What are You Saying? Using Topic to Detect Financial Misreporting* (SSRN Scholarly Paper No. ID 2803733). Social Science Research Network. Rochester, NY. Retrieved June 18, 2018, from https://papers.ssrn.com/abstract=2803733

Brown, S. V., & Tucker, J. W. (2011). Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research*, *49*(2), 309–346.

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

Buller, D. B., Strzyzewski, K. D., & Comstock, J. (1991). Interpersonal deception: I. Deceivers' reactions to receivers' suspicions and probing. *Communication Monographs*, *58*(1), 1–24. doi:10.1080/03637759109376211

Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs*, *58*(1), 25–40. doi:10.1080/03637759109376212

Burgoon, J. K., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., . . . Spitzley, L. (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.

Burgoon, J. K., Wilson, D., Hass, M., & Schuetzler, R. (2016). Interactive deception in group decision-making: New insights from communication pattern analysis. *Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME*, 37–62. Retrieved February 3, 2016, from http://link.springer.com/protocol/10.1007/978-1-4939-3249-8_2

Cacioppo, J. T., & Petty, R. E. (1989). Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, *10*(1), 3–12. Retrieved April 11, 2017, from http://www.tandfonline.com/doi/abs/10.1207/s15324834basp1001_2

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010a). Detecting management fraud in public companies. *Management Science*, *56*(7), 1146–1160.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010b). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, *50*(1), 164–175.

Cohen, L., Lou, D., & Malloy, C. (2013). Playing favorites: How firms prevent the revelation of bad news. *National Bureau of Economic Research*.

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, *29*(3), 845–868.

Davis, A. K., & Tama-Sweet, I. (2012). Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research*, *29*(3), 804–837.

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

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74.

Dong, W., Liao, S., & Liang, L. (2016). Financial statement fraud detection using text mining: A systemic functional linguistics theory perspective. In *Pacific Asia Conference on Information Systems (PACIS) 2016 Proceedings* (Vol. 188). Retrieved July 28, 2016, from http://aisel.aisnet.org/pacis2016/188

Dunbar, N. E. (2004). Dyadic power theory: Constructing a communication-based theory of relational power. *Journal of Family Communication*, *4*(3-4), 235–248.

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance*, *65*(6), 2213–2253.

Dyck, A., Morse, A., & Zingales, L. (2013). How pervasive is corporate fraud? *Rotman School of Management Working Paper*, (2222608).

Eaglesham, J. (2013). Accounting Fraud Targeted. *Wall Street Journal: Markets*. Retrieved December 18, 2014, from http : / / www . wsj . com / articles / SB10001424127887324125504578509241215284044

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research. *Intelligent Systems in Accounting, Finance and Management*. doi:10.1002/isaf.1386

Franco, G., Hope, O.-K., Vyas, D., & Zhou, Y. (2013). Analyst report readability. *Contemporary Accounting Research*, *Published In Advanced*.

Frank, E., Hall, M. A., & Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (Fourth Edition). Morgan Kaufmann.

Gaganis, C. (2009). Classification techniques for the identification of falsified financial statements: A comparative analysis. *Intelligent Systems in Accounting, Finance and Management*, *16*(3), 207–229.

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems*, *50*(3), 595–601.

Godfrey, J., Mather, P., & Ramsay, A. (2003). Earnings and impression management in financial reports: The case of CEO changes. *Abacus*, *39*(1), 95–123.

Goel, S., & Gangolly, J. (2012). Beyond the numbers: Mining the annual reports for hidden cues indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management*, *19*(2), 75–89.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting*, *7*(1), 25–46.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud - A comparative study of machine learning methods. *Knowledge-Based Systems*. Retrieved May 24, 2017, from http://www.sciencedirect.com/science/article/pii/S0950705117302022

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1–23.

Hancock, J. T., Woodworth, M. T., & Goorha, S. (2010). See no evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation*, *19*(4), 327–343. Retrieved September 8, 2017, from http://www.springerlink.com/index/L02553732537K401.pdf

Hauch, V., Blandon-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4), 307–342. Retrieved September 26, 2016, from http://psr.sagepub.com/content/19/4/307.short

Henning, P. J. (2012). The Mounting Costs of Internal Investigations. Retrieved August 31, 2017, from https://dealbook.nytimes.com/2012/03/05/the-mounting-costs-of-internal-investigations/

Hoberg, G., & Lewis, C. (2014). Do fraudulent firms strategically manage disclosure? *SSRN*, (2298302). Retrieved October 10, 2014, from http://ssrn.com/abstract=2298302

Hoberg, G., & Lewis, C. (2017). Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance*, *43*, 58–85. Retrieved September 1, 2017, from http://www.sciencedirect.com/science/article/pii/S0929119916303637

Hobson, J., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, *50*(2), 349–392.

Holtgraves, T., & Lasky, B. (1999). Linguistic power and persuasion. *Journal of Language and Social Psychology*, *18*(2), 196–205. Retrieved May 24, 2017, from http://jls.sagepub.com/content/18/2/196.short

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

Hubbell, A., Mitchell, M., & Gee, J. (2001). The relative effects of timing of suspicion and outcome involvement on biased message processing. *Communication*. Retrieved from http://nca.tandfonline.com/doi/abs/10.1080/03637750128056

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585–594.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44. Retrieved April 22, 2016, from http://pss.sagepub.com/content/22/1/39.short

Jensen, M. L., Lowry, P. B., & Jenkins, J. L. (2011). Effects of Automated and Participative Decision Support in Computer-Aided Credibility Assessment. *Journal of Management Information Systems*, *28*(1), 201–234.

Karapandza, R. (2016). Stock Returns and Future Tense Language in 10-K Reports. *Journal of Banking & Finance*. Retrieved June 28, 2016, from http://www.sciencedirect.com/science/article/pii/S0378426616300577

Karpoff, J. M., Koester, A., Lee, D. S., & Martin, G. S. (2012). A critical analysis of databases used in financial misconduct research. *Mays Business School Research Paper*, (2012-73), 2012–11.

Kumar, B. S., & Ravi, V. (2016). A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*. Retrieved October 21, 2016, from http://www.sciencedirect.com/science/article/pii/S0950705116303872

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

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*(1), 34.

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.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, *86*(3), 1087–1115. Retrieved February 12, 2016, from http://www.aaajournals.org/doi/abs/10.2308/accr.00000043

Leone, A. J., & Wu, J. S. (2007). *What Does it Take to Become a Superstar? Evidence from Institutional Investor Rankings of Financial Analysts* (SSRN Scholarly Paper No. ID 313594). Social Science Research Network. Rochester, NY. Retrieved August 29, 2017, from https://papers.ssrn.com/abstract=313594

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs, 66*(2), 125–144. doi:10.1080/03637759909376468

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

Lo, K., & Rogo, R. (2014). Earnings management and annual report readability. *Working paper.*

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

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research, 54*(4), 1187–1230. doi:10.1111/1475-679X.12123

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval.* Cambridge, UK: Cambridge University Press.

Matsumoto, D., Pronk, M., & Roelofsen, E. (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. J., Sethuraman, M., & Venkatachalam, M. (2015). MD&A Disclosure and the Firm's Ability to Continue as a Going Concern. *The Accounting Review, 90*(4), 1621–1651. doi:10.2308/accr-50983

Mayew, W. J., Sethuraman, M., & Venkatachalam, M. (2016). "Casting" a doubt: Informational role of analyst participation during earnings conference calls.

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance, 67*(1), 1–44.

McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit. Retrieved September 1, 2017, from http://mallet.cs.umass.edu/

Merkl-Davies, D. M., & Brennan, N. (2007). Discretionary disclosure strategies in corporate narratives: Incremental information or impression management? *Journal of Accounting Literature, 26*, 116–196.

Merkl-Davies, D. M., Brennan, N. M., & McLeay, S. J. (2011). Impression management and retrospective sense-making in corporate narratives: A social psychology perspective. *Accounting, Auditing & Accountability Journal, 24*(3), 315–344.

Mihalcea, R., & Strapparava, C. (2009). The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers* (pp. 309–312). Suntec, Singapore: Association for Computational Linguistics. Retrieved April 23, 2018, from http://www.aclweb.org/anthology/P/P09/P09-2078

Moffitt, K. C., & Giboney, J. (2011). Structured Programming for Linguistic Cue Extraction (SPLICE). Retrieved from http://splice.cmi.arizona.edu/

Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. Retrieved April 22, 2016, from http://jls.sagepub.com/content/21/4/337.short

Park, H. S., Levine, T., McCornack, S., Morrison, K., & Ferrara, M. (2002). How people really detect lies. *Communication Monographs*, *69*(2), 144–157. Retrieved September 13, 2017, from http://www.tandfonline.com/doi/abs/10.1080/714041710

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC2007.

Petty, R. E., & Cacioppo, J. T. (1979). Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of personality and Social Psychology*, *37*(1), 97–109. Retrieved April 11, 2017, from http://www.psy.ohio-state.edu/petty/PDF%20Files/1979-JPSP-Cacioppo,Petty(repetition).pdf

Purda, L., & Skillicorn, D. (2014). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 1–32.

Salzedo, C., Young, S., & El-Haj, M. (2014). *Does equity analyst research lack rigor and objectivity? Evidence from conference call questions and research notes* (SSRN Scholarly Paper No. ID 2433019). Social Science Research Network. Rochester, NY. Retrieved March 1, 2016, from http://papers.ssrn.com/abstract=2433019

ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some Evidence for Unconscious Lie Detection. *Psychological Science*, *25*(5), 1098–1105. doi:10.1177/0956797614524421. pmid: 24659190

Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decision Support Systems*, *74*, 78–87.

U.S. Securities and Exchange Commission. (2008). Topic 9 - Management's Discussion and Analysis of Financial Position and Results of Operations (MD&A). Retrieved March 29, 2018, from https://www.sec.gov/corpfin/cf-manual/topic-9

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, *24*(1), 33–49.

Wilczek, Y. (2014). SEC's 'Robocop' Tool 'in Limbo,' Former Enforcement Official Says. Retrieved February 16, 2017, from https://www.bna.com/secs-robocop-tool-n17179911063/

Worley, T. R., & Samp, J. (2016). Complaint avoidance and complaint-related appraisals in close relationships: A dyadic power theory perspective. *Communication Research*, *43*(3), 391–413. Retrieved May 24, 2017, from http://journals.sagepub.com/doi/abs/10.1177/0093650214538447

Zhou, L., Burgoon, J. K., Zhang, D., & Nunamaker, J. F. (2004). Language dominance in interpersonal deception in computer-mediated communication. *Computers in Human Behavior*, *20*(3), 381–402.

Zhou, L., & Zhang, D. (2008). Following linguistic footprints: Automatic deception detection in online communication. *Communications of the ACM*, *51*(9), 119–122.