APPLICATIONS OF NATURAL LANGUAGE PROCESSING TO

MEASURE AND ASSESS RISK IN THE U.S. FINANCIAL SYSTEM

BY

SANDRA J.H. ROLNICKI

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Management Science
in the Graduate College of the
Illinois Institute of Technology

Approved Yiwu Fang
Adviser

Chicago, Illinois
July 2018

ACKNOWLEDGEMENT

I am grateful to a number of people for their roles in supporting my Ph.D. journey and research. First and foremost, I thank my immediate family, including my husband, my mother, my brother, my godparents, my in-laws, and my extended family for their unwavering belief in me that inspired me to forge ahead.

I also extend gratitude to my management and colleagues at the Federal Reserve Bank of Chicago for seeing the value of this work and supporting my pursuit of this audacious goal.

To my Dissertation Committee, IIT Stuart professors and classmates, I appreciate the tools and wisdom you have shared with me.

Special thanks to my husband Andrew who constantly inspires me to continually improve myself, while simultaneously demonstrating how much he loves me just the way I am.

I also gratefully acknowledge the memory of my father who, along with my mother, raised me in a home where the first, last and only principle of parenting could be summarized in two words "Unconditional Love." I knew that no matter what I did or who I became that they would love me. Their reassurance has made all the difference.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure | Page

LIST OF ABBREVIATIONS

Abbreviation	Term
CAARs	Cumulative average abnormal returns
CFA	Confirmatory factor analysis
CIK	Central Index Key (unique firm identifier used by the SEC)
CRSP	Center for Research in Security Prices
CUSIP	Committee on Uniform Security Information Procedures (unique securities identifier)
EA	Enforcement action (issued by a Federal Banking Agency)
EDGAR	Electronic Data Gathering, Analysis and Retrieval (provided by the SEC for 10-K reports and other company filings)
FBA	Federal Banking Agency, e.g., Federal Reserve, OCC, FDIC
FDIC	Federal Deposit Insurance Corporation
FRB-NY	Federal Reserve Bank of New York
FRY-9C	Federal Reserve Consolidated Financial Statements for Bank Holding Companies
GVKEY	Global Company Key (unique firm identifier used by Standard & Poor in the Compustat database)
NLP	Natural language processing
OCC	Office of the Comptroller of the Currency
PSM	Propensity score matching
RSSD ID	Replication Server System Database ID (unique firm identifier used by the Federal Reserve for regulatory reporting)
SEC	Securities and Exchange Commission
STM	Structured topic modeling
WRDS	Wharton Research Data Services

ABSTRACT

Financial institutions play a critical role in maintaining the stability of U.S. financial market due to their function as intermediaries and their role as some of the market's largest publicly traded companies. Annually these firms file 10-K reports with the Securities and Exchange Commission (SEC). While there have been recent advances in performing content analysis of 10-K reports to assess the value of the information they contain (Balvers et al., 2015; Bodnaruk et al., 2015; Ertugrul et al., 2017; Lang and Stice-Lawrence, 2014; Loughran and McDonald, 2011, 2014a, 2014b, 2016; McClelland et al., 2010), researchers have been limited in their application of natural language processing (NLP) to establish a quantitative link between the content of 10-Ks and the actual outcomes for the firms, with notable exceptions of Gandhi et al. (2018) in their use of sentiment as a proxy for financial distress in U.S. banks.

To fill this gap in the literature, I have applied NLP techniques, including unsupervised machine learning, to over 11,000 annual reports covering over 1,000 firms between 1994 and 2016. A hand-collected dataset of enforcement actions from Federal Banking Agencies' public websites provided over 2,000 firm-year examples for comparison against banks without enforcement actions.

My results of structured topic modeling and sentiment analysis of 10-K reports yielded unique insights about the power of banks' words to provide indicators of risk, both at the individual bank level and at the collective, systemic level. Portfolio constructions based on sentiment yielded signals that predicted both the Dot-Com Bubble and the Financial Crisis, while significantly out-performing both the Standard & Poor 500 and a Bank Index of stocks. Event studies that grouped banks by sentiment

consistently produced abnormal returns according to sentiment and identified patterns that reflect how the market reacts to and absorbs negative news from banks with enforcement actions.

These applications could complement existing measures of risk in the U.S. financial system. Shareholders and other financial market participants, including financial firms' clients, analysts, employees and regulators will benefit from this work.

CHAPTER 1

INTRODUCTION

Financial institutions, e.g., banks, bank holding companies, thrifts, etc., play a critical role in maintaining the stability of U.S. financial market due to their function as intermediaries and their role as some of the market's largest publicly traded companies, e.g., the top five publicly traded bank holding companies (by assets) held over $9.6 trillion on behalf of their clients and represented nearly $1.3 trillion in market capitalization on the U.S. stock markets as of September 30, 2017 and December 31, 2017 respectively.[1] The financial firms whose shares are traded on the major exchanges are required to file Form 10-K ("10-K report" or "10-K") annually with the Securities and Exchange Commission (SEC). The 10-K is often considered the most comprehensive single source of information about a firm's performance and financial position beyond what is provided in earnings announcements (Griffin, 2003), with the average 10-K containing over 40,000 words (in this sample). Despite their importance, the only portion of a 10-K that requires independent assessment before submission to the SEC is the financial statement section.

1.1 Motivation for Research

While there have been recent advances in performing content analysis of 10-Ks to assess the value of the information they contain (Balvers et al., 2015; Bodnaruk et al.,

[1] *Sources*: Bank rankings by assets: National Information Center (NIC); Market capitalization: Google Finance. Firms are: JPMorgan Chase & Co., Bank of America Corporation, Wells Fargo & Company, Citigroup Inc., Goldman Sachs Group, Inc.

2015; Ertugrul et al., 2017; Lang and Stice-Lawrence, 2014; Loughran and McDonald, 2011, 2014a, 2014b, 2016; McClelland et al., 2010), the literature thus far has not attempted to establish a quantitative link between the content of 10-Ks and the actual outcomes for the firms, with notable exceptions of Gandhi et al. (2018) in their use of sentiment as a proxy for financial distress in U.S. banks.

Establishing this relationship would benefit investors, analysts and regulators because this would provide a more fulsome view of banks' risk profiles beyond what is required to be reported in their financials, information that Gandhi et al. (2018) term "window dressing." Stakeholders could then utilize this information to make sound investment and regulatory decisions regarding that firm. In addition, to date researchers have not exploited the extensive regulatory reporting required by financial firms (such as reporting on risk-weighted assets and loan portfolio performance) to obtain deeper insights as banks' individual and collective risk profiles.

1.2 Research Questions

The primary research questions to be answered focus on whether and to what extent the sentiment and content used by banks in their 10-K report can be used as an indicator of systemic risk in the U.S. financial system:

- Do the words banks' use in 10-K reports and how they use them (content and sentiment) provide insights into their risk profiles?
- To what extent do the content and sentiment of banks' reports differ when they experience adverse outcomes during the reporting period?
- (How) Does the market perceive the sentiment of banks' 10-K reports?

1.3 Contributions

In answering these questions, this research will have multiple contributions to this growing body of literature. I will apply transparent and reproducible natural language processing (NLP) techniques, including unsupervised machine learning, to assess the content and sentiment of banks' 10-K reports. I will leverage regulatory reporting required by financial firms to establish risk measures to compare banks' profiles against NLP results. I will create a collection of bank enforcement actions (EA) to represent adverse events for bank sample population (>2,000 firm-year observations). I will construct sentiment-based portfolios of banks' stocks to measure and compare market performance. Lastly, I will perform event studies around 10-K disclosure dates to assess the impact of 10-K sentiment and adverse events. The sum of this work will provide unique insights to measure and assess risk in the U.S. Financial System. Shareholders and other financial market participants, including financial firms' clients, analysts, employees and regulators will benefit from this work.

The remainder of this paper will proceed as follows: Chapter 2 offers background on 10-K reports, and provides rationale as to why a focus on financial firms will benefit this body of knowledge. This chapter also introduces streams of past research, as well as the theories leveraged and key findings from past empirical work. This chapter concludes with the direction of future research by describing the gaps in the current literature and trends toward which the research is headed. Chapter 3 outlines the hypotheses based on the research questions to be answered. Chapter 4 presents the data utilized and methods drawn from the literature. Chapter 5 discusses the results and implications thereof. Chapter 6 discusses both challenges faced with this research and opportunities to further these studies. Chapter 7 concludes.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter provides an introduction to 10-K annual reports and risk disclosures; discusses streams of prior research, theories leveraged and key findings; identifies gaps in past research and trends observed that are likely to continue as more work is done in this area. The chapter closes with my contributions to the growing body of literature.

2.1 Introduction to 10-K Reports

Annually per the Securities Exchange Act of 1934, all publicly traded firms must file Form 10-K with the SEC for the primary benefit of their shareholders as well as analysts, clients, employees and regulators. Contents of the 10-K report include the following: year-end financial statements, management discussion and analysis, business line descriptions and performance, and risk (factor) disclosures.

Given the volume of information they contain, as noted by Loughran and McDonald (2014a), the average 10-K report contains more than 38,000 words. However, within these extensive contents, only the financial statements are required to be audited by an independent (external) auditor prior to the submission to the SEC. While the market expects management to present accurate and bias-free commentary in their firms' 10-K, the firm may also view the annual report as a mechanism by which to convince investors to buy stock or to reassure current shareholders of their investment decision (Bettman and Weitz, 1983). Given the discretion that management has with the 10-K outside of the financials, the firm may choose to increase their reporting of positive information while lessening the negative information presented (Dobler, 2008).

2.2 Risk Disclosures

Since one objective of the 10-K is to provide information as to the "amount, timing, and uncertainty of future net cash inflows" (SEC, 2010), there are several sections within the report that require or encourage firms to make risk disclosures that will inform their shareholders as to the nature and extent of these uncertainties.

For example in 2005, the SEC began requiring listed firms to include a Risk Factor section in their 10-K (Item 1A) to discuss ''the most significant factors that make the company speculative or risky'' (SEC, 2010). In addition Item 7A, "Quantitative and Qualitative Disclosures about Market Risk," requires firms to provide information about their exposure to risks such as those related to interest rates and foreign currency exchanges (SEC, 2010).

There has been significant work to examine the information within risk disclosures and develop quantitative measures of the disclosure contents (Abraham and Cox, 2007; Berger, 2011; Beretta and Bozzolan, 2004; Dobler, 2008; Dobler et al., 2011; Lajili and Zéghal, 2005; Linsley and Shrives, 2006; Roulstone, 1999; Solomon et al., 2000). In spite of this, the methods by which risk disclosures are assessed remain under development, resulting in discord about how best to extract meaningful information (Miihkinen, 2012).

2.3 Streams of Past Research on Firm Disclosures

Several streams of past research were identified throughout the literature review. Below are examples from work performed on content analysis of firm disclosures, event studies based on timing of the release of firm disclosures, as well as work that links risk disclosures with financial and non-financial measures.

Content analysis of firm disclosures. Textual risk disclosures are often quantified by counts of specific "key" words, appearing in specially constructed word lists and dictionaries (Campbell et al., 2014; Loughran and McDonald, 2011, 2014a, 2014b; Kravet and Muslu, 2013). For example, recent research by Ertugrul et al. (2017) used word counts and keyword dictionaries developed by Loughran and McDonald to conclude that firms that are more ambiguous in their annual reports pay higher costs when taking out loans.

With the exception of Ertugrul et al. (2017) and Gandhi et al. (2017), a large portion of the literature has focused on analyzing the content of firm disclosures without linking those disclosures to firm performance, i.e., stock prices or firm financial measures. Firm disclosures analyzed in these studies have included the full 10-K reports, while others have focused on a specific section within the annual report, such as the Management Discussion and Analysis (MD&A). Additional firm disclosures considered in the literature include Initial Public Offering (IPO) prospectuses and SEC Form 424 (Amendments to IPO filings).

Loughran and McDonald have contributed significantly to the body of literature around content analysis of 10-K reports (Loughran et al., 2009; Loughran and McDonald, 2011, 2014a, 2014b, 2016). Focused on readability and links with non-financial measures, their work has contributed word dictionaries on topics such as ethics and litigation to assist current and future researchers who want to build on this practice. They have demonstrated why previously accepted measures for readability such as the Fog

Index are not effective in studying 10-K reports. They have also created a database of 10-K headings that includes firm information on each 10-K and 10-Q filing from 1994 to 2016 such as word counts related to positive and negative sentiment.

Miihkinen (2012) compared qualitative and quantitative risk disclosure levels based on firm characteristics such as profitability, size and foreign listing status with the main findings that larger firms and those required by regulation will disclose more quantitative risk information.

Lajili and Zéghal (2005) analyze specific sections of Canadian firms' annual reports (MD&A and "Notes to the Financial Statements") to glean information about financial and non-financial risks such as operational, regulatory and environmental risks. They measure the volume and "intensity" of risk disclosures (frequency with which a specific type of risk disclosure appears) while highlighting the most frequently disclosed combination of risks. A valuable contribution to this body of literature is the authors' development of a framework that identifies firms' disclosed Risk Sources and related Risk Management techniques, as well as an extension to Likelihood and Consequence of these risks occurring (see Tables 5 and 6 in Lajili and Zéghal, 2005). Their paper supports the implementation of certain mandatory risk disclosures and the development of a risk disclosure index.

Event studies based on timing of the release of firm disclosures. Some of the literature on firm risk disclosures focuses on stock market and analyst activity around the timing of the disclosure released, e.g., 10-K and its quarterly equivalent, Form 10-Q. Although these researchers do not perform content analysis, they judge the information value of these risk disclosures by subsequent stock price movement and analyst forecast changes.

In an early event study that focused on the timing of firm disclosures, Beaver (1968) examines investor reaction to earnings announcements. Looking at 143 firms from 1961-1965, he found that both market price and volume were impacted on the days surrounding the announcements thereby supporting “the contention that earnings reports possess information content.” Griffin (2003) also performed an event study tied with the release of 10-Ks and 10-Qs, showing that companies experience greater excess returns (in absolute value) on the day of and two days immediately following the filing. Similarly Qi et al. (2000) do not analyze the content of 10-K reports; instead they use analysts’ ratings as a method by which to measure the “informativeness” of the 10-Ks. This study focuses on the advent of electronic 10-K filings and found, based on abnormal market returns, that the market extracted more information in a timelier manner from 10-Ks once electronic filings became mandatory.

Linked with financial measures. While some of the literature has included links with firms’ stock market performance or other key financial metrics that are reported in their public filing, there is less research that uses non-public and/or non-financial measures.

Brown and Tucker (2011) focus on year-to-year changes in the MD&A section of the 10-K and relate their findings to firms’ stock market price changes and analyst earning forecast revisions. Similarly, You and Zhang (2009) use a simple proxy (total word count) for assessing 10-K complexity and observe stock returns following the 10-K filing date. They find that investors “underreact” to more complex annual reports.

Kravet and Muslu (2013) use 22 key words and sentence counts within 10-K reports to assess year-on-year changes to the full text of 10-K reports. They compare this figure to changes in stock market activity and analyst activity around the time the 10-Ks are filed. They find that while textual risk disclosures increase investors' risk perceptions of a firm, the firms' annual reports are likely to contain "boilerplate" language that is common within their given industry.

Campbell et al. (2014) focus on the Risk Factors with the 10-K (Section 1A) and find that "firms facing greater risk disclose more risk factors." They are able to relate the information from the risk factor disclosures to the firm's level of systemic and idiosyncratic risk. Their results suggest that while the market equates increased risk factor disclosure with an increase in the firm's risk, "the public availability of risk factor disclosure decreases information asymmetry among that same firm's shareholders." Campbell et al. categorize risk factor disclosures into five groups: (1) financial, (2) tax, (3) legal, (4) other-systematic, and (5) other-idiosyncratic; however, their "pre-disclosure proxies for risk" do not link directly to any of these risk factor categories, e.g., although there is a legal risk category, there is not a proxy for legal or litigation risk.

Rutherford (2003) attempts to link poor firm performance with annual report textual complexity although his hypothesis is ultimately not supported (he does not find that poorly performing firms obfuscate their reports).

Bodnaruk et al. (2015) furthers the work of Loughran and McDonald (2011, 2014a, 2014b) by creating a unique lexicon of words that indicate "financial constraint." They apply this word set to 10-Ks and compare their results with the firms' performances in following years. They find that "the frequency of constraining words… predicts subsequent liquidity events" that were not foretold by traditional financial statement measures.

Barakat and Hussainey (2013) study bank operational risk disclosures and found they were occurred more often in firms that exhibit certain governance characteristics such as a higher proportion of outside directors and a more active audit committee. However, the content analysis of European banks' annual reports spanning three years was performed without leveraging any automated tools and was therefore limited to 243 firm-year observations.

Linsley et al. (2006) focus on Canadian banks' annual reports. They count sentences that contain risk and risk management terms (not individual words) in order to measure the volume of risk disclosure. They use book-to-market value of equity as a proxy for risk and measure the correlation to the types of risk disclosures made by the banks (e.g., positive, negative and neutral). However, beyond this association, the authors did not directly link the nature of the risks with any firm financial or performance measures.

Lang and Stice-Lawrence (2014) examine annual report text from over 15,000 non-US firms and use a lagged disclosure variable to show that increases in annual report metrics (such as report length) preceded increases in liquidity, institutional ownership and number of analysts following their firms.

Linked with non-financial measures. As mentioned above, there is less published research that associates textual analysis of annual reports with non-financial measures. Several notable exceptions are summarized here. Loughran and McDonald (2014b) studied 10-Ks, IPO prospectuses and Form 424 between 1994 and 2009 to find that firms with higher corporate governance scores (based on Gompers et al., 2003) file more readable documents.

Nelson and Pritchard (2007) relate disclosures in annual reports to one form of litigation risk by investigating the use of cautionary language by firms that face the risk of being sued for securities fraud. Under the safe harbor provision of the Private Securities Litigation Reform Act of 1995, firms are protected from subsequent liability if they provide cautionary language in annual reports that informs their investors of the uncertainty of forward-looking statements such as earnings and revenue forecasts. The authors find firms that are more at risk of securities litigation do include more cautionary language, revise their cautionary language to a greater extent than other firms from year-to-year and use more plain English language, i.e., that scores more favorably on readability measures.

Balvers et al. (2015) relate the frequency of “customer satisfaction” phrases in firms’ 10-K reports to the firms’ scores published by the American Customer Satisfaction Index (ASCI). They find a significant relationship between the frequency of (positive) customer satisfaction language and firms’ ACSI scores. Firms that merely provided commentary that they are addressing negative customer satisfaction issues received lower ACSI scores on average.

Michalisin (2001) tests the validity of firms’ innovativeness assertions within their 10-Ks by comparing frequency of key words and measures of innovation, such as trademark applications, and finds such claims to be valid based on his empirical measures.

Perry and de Fontnouvelle (2005) use operational loss announcements to measure banks’ reputational losses, which they define as the amount lost in stock market price as compared to the amount of operational loss. They find reputational losses occur when internal causes rather than external drivers bring about the operational failures.

2.4 Key Theories Leveraged in Past Research

Several management theories were identified in the literature reviewed. Most of the research leveraged agency theory and/or information asymmetry to some degree. Other theories, such as signaling, were also evidenced and so they are summarized here.

Agency Theory and Information Asymmetry. Ross' theoretical paper (1973) on the principal-agent problem highlighted the significance of information flows. His work influenced Jensen and Meckling's (1976) theory of the firm that develops a model to in part describe managerial behavior under information asymmetry.

In applying these theories, past research studied here noted that when firms disclose more, information asymmetry is reduced among both current and prospective shareholders, and that firms with greater disclosure have more liquid securities (Campbell et al., 2014; Easley and O'Hara, 2004, Kothari et al. 2009).

Abrahamson and Park (1994) provide an earlier work that relates disclosure levels with information asymmetry. By analyzing negative word counts in over 1,000 letters to shareholders and relating these to firms' return on assets (as a measure of performance), they found that low disclosure precedes stock sales by top level management and outside directors, "supporting the claim that concealment by officers and its toleration by directors may be intentional."

Dobler (2008) applies the "cheap talk model," wherein management disclosures cannot be verified by the market, to support the view that firms weigh the cost of disclosure of non-verifiable information by considering the impact of such information on their cost of capital. In similar research, Balvers et al. (2015) identified firms' cheap talk on customer satisfaction by noting lower ASCI scores at those firms.

Dobler et al. (2011) note that management decides whether and to what extent risks should be disclosed based on risk disclosure incentives. When regulatory requirements around disclosure are present, the risk disclosure quantity increases.

Helbok and Wagner (2006) demonstrate that banks with lower capital and/or profitability ratios have greater incentive to assure the market that operational risk at their firms is well managed and therefore they see greater levels of disclosure of risks by these firms.

Signaling and other management theories. Early work by Akerlof (1970) on information asymmetry and signaling provided inspiration for Rothschild and Stiglitz (1978) to conclude that when signaling is present in a market with asymmetric information, the resulting equilibrium looks significantly different than a market that lacks the transfer of information.

As demonstrated by Balvers et al. (2015), in which firms that more frequently discussed customer satisfaction in a meaningful way had higher ACSI scores, firms can use their annual reports as a signaling mechanism. Similarly, Michalisin (2001) was able to empirically validate that firms who make more claims about their innovation in their annual reports do engage in more innovative behavior, such as filing more trademark applications than their peers. These examples reviewed in concert with those related to information asymmetry would suggest that when management is exhibiting positive behaviors that bring about opportunities for their firms (opportunities being the converse of risks), there is a greater likelihood of disclosure.

McClelland et al. (2010) explore the phenomenon of CEO's commitment to status quo (CSQ) by reviewing letters to shareholders from 129 firms for phrases indicative of CEO CSQ. These metrics were analyzed relative to other CEO attributes such as age and length of time in their role to identify that in high discretion industries (e.g., computer equipment), firms whose CEOs had higher CSQ subsequently experienced worse financial and market performance. Similarly Bettman and Weitz (1983) studied letters to shareholders to identify patterns of self-serving attributions and find that negative outcomes were attributed to external causes while positive firm outcomes were attributed to management prowess.

2.5 Key Findings in Past Empirical Research

Campbell et al. (2014) note that theory would suggest greater disclosure would result in lower costs of capital. In line with this, empirical results from Ertugrul et al. (2017), that attributed the obfuscation in annual reports to "managerial information hoarding" (p.811), determined that this behavior resulted in higher costs of borrowing. However, Campbell et al. also note that some empirical work has not supported this lower-cost-of-capital-with-greater-disclosure expectation. One reason for these mixed results mentioned by Kothari et al. (2009) is because tone of the management disclosures cannot easily be measured empirically, though I would suggest that this research does identify an empirical measure for tine, i.e., sentiment. Also Brown and Tucker (2011) found lower stock price reactions to MD&A disclosures suggesting that investors are finding this information less useful.

Brockman and Cicon (2013) examine the contents of earnings announcements and confirm a positive relationship between the "surprise component" and abnormal returns, indicating that investors find announcements informative. Tetlock et al. (2008) take a slightly different approach by analyzing externally written news content rather than internally written firm disclosures. They find that "linguistic media content" discloses information about the firms that investors assimilate more easily than financial reports, as evidenced by the timing of stock price reactions relative to the release of these differently sourced contents.

2.6 Gaps in Past Research

As discussed, there is a lack of literature that relates the level of risk disclosures to the level of risk behaviors undertaken by firm management. Although examples were provided in the previous section as to the extensive analysis of the text and content performed on 10-K and other firm disclosures and the work that has been done to correlate disclosure measures with financial and non-financial data, there is no direct link between the risks disclosed and the actions taken by firm management that contribute to actual risk undertaken by the firm. Clearly, there are empirical challenges in creating these measures; however, investors and regulators would benefit from this additional transparency.

2.7 Focus on Financial Firms

As previously noted banks, bank holding companies and thrifts play a critical role in the functioning of the U.S. financial system both due to their role as financial intermediaries for individuals and businesses and because of their aggregate market capitalization.

Prior research in 10-K disclosures has been cross-industry in nature, including financial and non-financial firms. Therefore those research efforts have only been able to exploit information that is reported by all types of companies. By focusing on financial firms, data from regulatory reporting (such as the quantification of risk weighted assets) will be assessed in order to provide a more complete measurement of the risks being disclosed by banks and the risks undertaken by management. Also, oftentimes financial firms are excluded from cross-industry studies because of significant regulatory oversight that could create anomalies in cross-industry data.

2.8 Research Trends Observed

The two main trends observed in the literature are the automation of content / textual / sentiment analysis and the move toward interdisciplinary studies. These are both explored below.

Automation of content / textual / sentiment analysis. Content analysis is becoming increasingly automated through the use of supervised and unsupervised machine learning techniques. The benefits of this approach include a reduction in human error (due to fatigue or inconsistent training) and the ability to scan volumes of information that could not previously be analyzed directly by researchers and their assistants. For example, Li (2010) applied a naïve Bayesian algorithm to the MD&A section of 140,000 10-K and 10-Q filings (from 1994-2007) to examine the information content in the firms' forward-looking statements. Similarly, Huang et al. (2014) use a naïve Bayesian algorithm to examine nearly 364,000 analyst reports. Their work produces a model to predict future earnings growth.

Groth and Munterman (2011) use four types of machine learning: naïve Bayes, k-nearest neighbor, neural network and support vector machine, to detect patterns in textual (news) data posted on investment social media websites and their impact on banks' stock prices. They were able to identify the types of disclosures that resulted in the greatest abnormal stock price volatility for the disclosing firms.

However, these advancements in technology may (counter-intuitively) contribute to a reduction in reproducibility of research as programmers may be less inclined to share their code or a propriety tool that they have developed for this analysis. For example, the quantitative stock prediction system based on financial news (AZFinText), which was developed by Schumaker and Chen (2009), is cited in several journal articles as an example of how content analysis can yield measurable improvements in stock performance prediction, yet their published description of AZFinText does not include reproducible methods, thereby requiring other researchers to buy a license to their software in order to benefit from their work.

Move toward interdisciplinary work. Older literature reviewed was divided in that content and textual analysis appeared in social science journals while event studies and portfolio construction were in accounting and finance journals. More recent literature incorporates both disciplines, while considering new methods to ascertain information from various risk disclosures and risk measures, such as propensity score matching, which originated in the medical discipline.

2.9 How This Research Will Contribute

I will focus on financial firm and leverage the rich dataset of banks' regulatory reporting. I will apply NLP techniques, including unsupervised machine learning, in a transparent, reproducible manner to perform interdisciplinary work that will draw on methods used in finance, social sciences, linguistics and the medical field.

CHAPTER 3

HYPOTHESES

The following hypotheses were developed based on the literature review and data analyzed. These will be tested, refined and validated on a population of over 11,000 firm-year observations (10-K reports) that span the filing years 1994 through 2016 and cover over 1,000 unique financial institutions.

With regard to what banks say and how they say it following an adverse event such as receiving an enforcement action (EA), Barakat and Hussainey (2013) concluded operational risk disclosures were more evident in firms with stronger governance. Firms that have experienced a negative event, such as formal action from a FBA or civil money penalties for bad behavior, will differ in what and how they disclose information to their stakeholders. In terms of content (the "what"), Hargie et al. (2010) analyzed public testimony from four bank executives to the Banking Crisis Inquiry of the Treasury Committee of the UK House of Commons. They found that the executives employed several tactics to separate, or dissociate, their firms and themselves from the financial crisis. Rather than discuss the technical nature of the risks they undertook or their business models, they spoke to the character of their leadership and organizations. Similarly, I expect banks that receive an EA would be less likely to discuss the technical aspects of their business in their subsequent-year 10-K and revert to more of a character profile of their firms.

H.1 Technical topics will be less prevalent in the content of banks' 10-Ks that received one or more EAs in the prior fiscal year, compared to peer banks that did not receive EAs that year.

H.2 Technical topics will be less prevalent in the content of banks' 10-Ks that have one or more EAs in effect during the sample period, compared to peer banks that did not receive EAs in the sample period.

As to how banks with EAs disclose information to their stakeholders, I would again leverage the work of Hargie et al. (2010) in their finding that the bank executives peppered their testimonies with words that have been characterized by this research as "high" sentiment, e.g., "we are profoundly… unreserved sorry." Therefore, I would expect banks in receipt of an EA to utilize high sentiment in their 10-Ks to a greater degree than those banks that have not.

H.3 For banks that receive one or more EAs in a given fiscal year, the sentiment expressed in their 10-K will be higher than peer banks that did not receive EAs that year.

H.4 For banks that have one or more EAs in effect during the sample period, the sentiment expressed in their 10-K will be higher than peer banks that did not receive EAs in the sample period.

Gandhi et al. (2018) found that more negative sentiment in banks' 10-K reports was associated with greater likelihood for adverse outcomes, such as delisting of stock and lower subsequent return on assets. As a higher level of (any type of) sentiment is characterized in this research as "high" sentiment, I would expect banks that employ higher sentiment compared to their peers to experience decreased market performance. I would even expect this to be maintained when comparing banks with EAs in effect.

H.5 Based on the sentiment expressed in banks' 10-Ks, portfolios constructed on high (*low*) sentiment will have lower (*higher*) abnormal returns than their peers with low (*high*) sentiment.

H.6 Even when comparing only those banks with EAs initiated in the prior fiscal year, based on the sentiment expressed in banks' 10-Ks, portfolios constructed on high (*low*) sentiment will have lower (*higher*) abnormal returns than their peers with low (*high*) sentiment.

H.7 Pre- and post-10-K disclosure cumulative average abnormal returns (CAARs) will be lower (*higher*) for firms that employ high (*low*) sentiment in their 10-K report.

H.8 Even when comparing only those banks with EAs initiated in the prior fiscal year, pre- and post-10-K disclosure cumulative average abnormal returns (CAARs) will be lower (*higher*) for firms that employ high (*low*) sentiment in their 10-K report.

After detailing the results of my hypothesis testing in Chapter 5, I will discuss the informative (and even predictive) power of the NLP methods employed here in assessing banks' individual and collective levels of risk relative to the U.S. Financial System.

CHAPTER 4

DATA AND METHODS

With an understanding of the body of literature, the future direction of NLP research, my research contributions and the hypotheses to be tested, this section will describe the data and methods to be used for these tests.

4.1 Sources of Data

As described in prior sections, there will be several sources of data used in this research. All are either publicly available or obtained through the University's subscription to Wharton Research Data Services (WRDS).

Step 1: Constructing the bank population. A dataset constructed by the Federal Reserve Bank of New York (FRB-NY) [2] joins each bank's unique identifier for regulatory reporting (RSSD ID) with the a PermNo, which is assigned by the Center for Research in Securities Prices (CRSP). From this, I created a dataset of matched bank identifiers to include each bank's RSSD ID, PermNo, GVKEY (used by Standard & Poor), PermCo (also used by CRSP), stock ticker symbol and CUSIP (used for issuance of debt and securities) by joining the FRB-NY dataset with CRSP and Compustat tables, via WRDS. I then used the link dates from the FRB-NY dataset and CRSP / Compustat to determine the identifiers of those banks operating within the sample period of 1994-2016. I followed a process similar to Gandhi et al. (2018) to link banks identifiers with their

---

[2] Matched using the *Banking Research Dataset* by the FRB-NY, 2017. *https://www.newyorkfed.org/research/banking_research/datasets.html*.

respective CIK identifiers in order to retrieve the 10-K reports. Appendix B details the steps described here.

Step 2: Retrieving the 10-K reports. I used the CIK identifiers to retrieve the banks' 10-K reports from the SEC EDGAR website for the sample period. I included all variations of 10-K reports, e.g., amended 10-K reports designated as 10-K-A reports. I retained only one 10-K per bank per year, i.e., if an amended report was filed, only the final 10-K was retained for analysis. The work of Bill McDonald, PhD of the University of Notre Dame and his team was leveraged using their Stage One 10-X Parse Data [3] and 10-K Summaries [4] to confirm completeness of my sample and obtain word count statistics. Table 1 provides summary statistics for the 10-K reports retrieved.

Table 1

Summary Statistics of 10-K Reports Retrieved

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Words per 10-K	11,123	40,809	40,809	2,774	449,322
Unique Words per 10-K	11,123	2,678	2,678	571	6,311

Note: The number of unique banks represented is 1,037.

Figure 1 provides a graphical interpretation of the reports over the sample period of 1994 through 2016. The top half of the figure provides insights as to the average number of words (thin green line) and the average number of unique words (thick pink

[3] See University of Notre Dame's Software Repository for Accounting and Finance (SRAF) for Stage One Parse Data. *http://sraf.nd.edu/data/stage-one-10-x-parse-data/*

[4] See University of Notre Dame's SRAF for 10-K summaries. *http://sraf.nd.edu/textual-analysis/resources/#LM_10X_Summaries*

line) in the sample 10-K population. The scale of the increase in the number of words closely tracked the number of unique words as shown by the co-movement of these two lines. The number of words per 10-K increased 91.3% between 2002 and 2016, with a year-over-year increase of 23.6% between 2008 and 2009 in reaction to the Financial Crisis. The grey bars in the lower half of Figure 1 track the number of banks filing 10-K reports throughout the sample period. There was an increase in the number of banks filing 10-Ks of 405% between 1994 and 1997, while there was a decrease in the number of banks filing 10-Ks between 2004 and 2016, reflecting industry consolidation and the aftermath of the Financial Crisis.

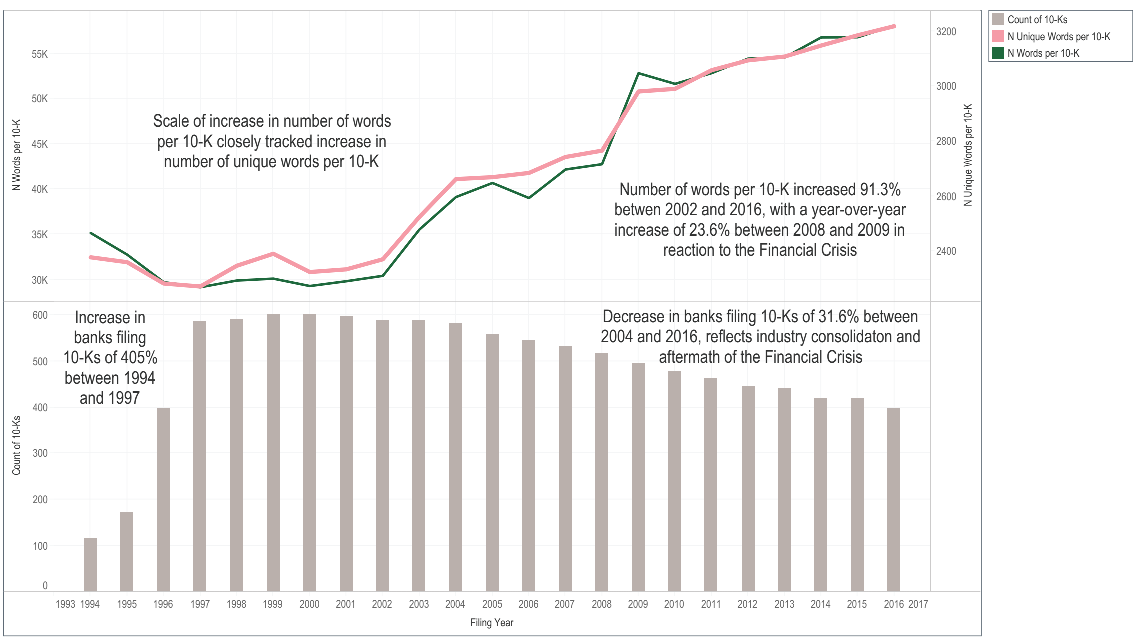

Figure 1. Graph tracking the number of words and number of unique words per 10-K (upper half) and number of 10-K reports filed per year (lower half) from 1994 through 2016.

Step 3: Retrieving bank enforcement actions. An enforcement action (EA) from a Federal Banking Agency (FBA), e.g., Federal Reserve, FDIC, Office of the Comptroller of the Currency (OCC), represents an adverse event for a bank (Jordan et al., 2000). EAs vary widely in their scope and severity but a commonality is that they are all published publicly on the issuing FBA's website in a document format, which was scanned or converted from a letter sent to the recipient bank. I retrieved the EAs from FBA sites for the sample period and performed optical character recognition (OCR) on the files to determine bank name, type of EA, type of communication (initiating an EA, terminating an EA, etc.), and date of EA action.

Step 4: Linking enforcement actions to sample banks. Following the methodology described by Nguyen et al. (2015), I retained the "severe EAs," including 1) Formal agreements, 2) Cease and desist orders; and 3) Prompt corrective actions. I matched EAs to banks by name via the FRB-NY dataset to obtain the bank's CIK. A dummy variable *EA_Initiated* was created in the dataset to reflect in a given 10-K filing year whether an EA was initiated in the prior fiscal year. For example, if an EA was initiated by an FBA on October 1, 2014 and the bank's fiscal year ended December 31 of that year, *EA_Initiated* is set equal to 1 for the filing year 2015 (t=0). Leading variables for t+1 and t+2 were also created.

In addition, a dummy variable *EA_Term* was created in the dataset to reflect in a given 10-K filing year whether an EA was terminated at any point in the prior fiscal year. For EAs terminated in the sample period, I looked up the date on which the EA was initiated and added that to the dataset. This would allow me to see if a bank had an EA in effect during any year of the sample period, even if it was initiated before the start of the sample period. Using the *EA_Initiate* and *EA_Term* dates, a dummy variable *EA_Period* was created in the dataset to reflect in a given 10-K filing year whether an EA was in effect at any point during the sample period. Table 2 below provides summary statistics of the EAs retrieved.

Table 2

Summary of Enforcement Actions (EAs) Retrieved

EA Measure	*Count*
EAs initiated during the Sample Period	344
EAs terminated during the Sample Period	298
Firm-Year Observations with EAs in Effect	2,902
Firm-Year Observations without EAs in Effect	8,221

Note: The number of unique banks represented is 203.

Figure 2 provides a graphical interpretation of the EAs over the sample period. The top half of the graph depicts the number of total EAs in effect for the sample population of banks (pink line) along with the average number of days that EAs terminated in that year had been open (brown bars). Bars on the left had side of the graph likely represent EAs from the Savings and Loan Crisis of the 1980s, while the bars on the right side represent significant findings from the Financial Crisis. The lower half of the graph shows the number of EAs initiated (thick orange line) and the number of EAs terminated (thin blue line) each year. The spike in the range line represents an increase of 636% in the number of EAs opened by FBAs following the Financial Crisis.

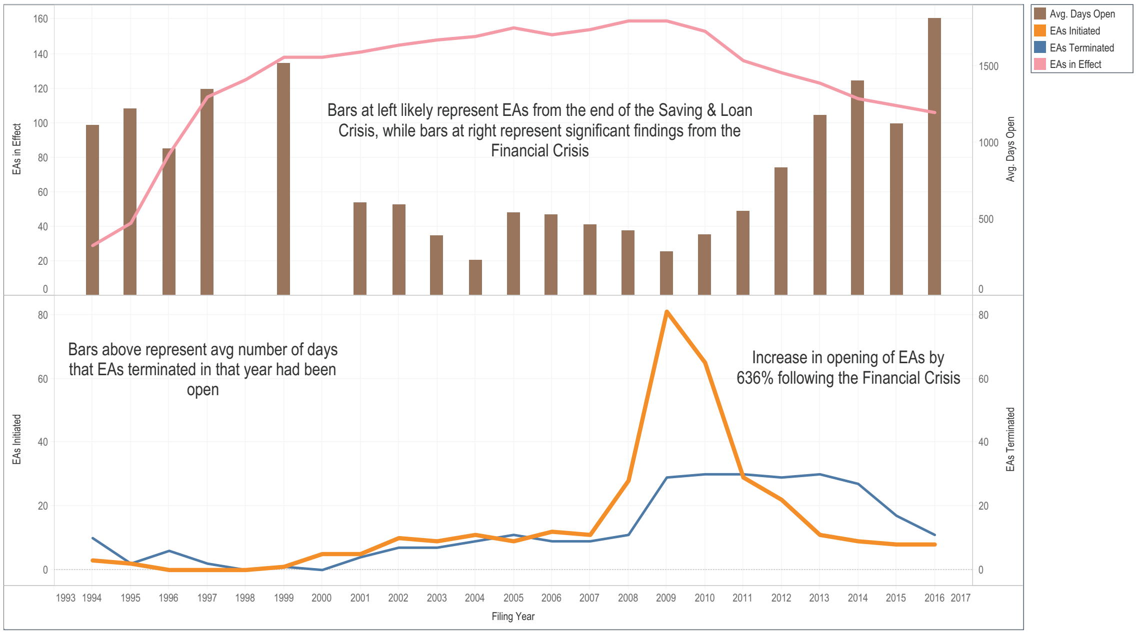

Figure 2. Graph tracking the average number of EAs in effect and the average number of days EAs were open (upper half) as well as the number of EAs initiated and terminated (lower half) per year from 1994 through 2016.

Step 5: Retrieving regulatory reporting data from NIC. Using the bank identifier for regulatory reporting, the RSSD ID, I retrieved data reported by the sample banks on Form Y-9C (Annual Bank Holding Company Report) for each year in the sample period. Form Y-9C contains standard financial statements such as Balance Sheet, Income Statement, Statement of Equity, as well as Off-Balance Sheet, Loan Performance and Risk-Based Capital measures. I used data reported at December 31 of each year as the data is reported on a calendar year, regardless of a bank's fiscal year. I linked the regulatory reporting data for a given year with the subsequent calendar year's 10-K report, e.g., the year-end reporting filed in 2014 was linked to the bank's 10-K report filed with the SEC in 2015. Samples were removed that were missing key data fields, reducing the sample size as shown in Table 3.

Table 3

Summary Statistics of Regulatory Reporting Data from NIC (in $000s)

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Total Assets	8,633	16,500,000	121,000,000	93,180	2,570,000,000
Net Income (Loss)	8,633	149,609	1,074,586	(3,094,179)	24,400,000

Note: The number of unique banks represented is 925.

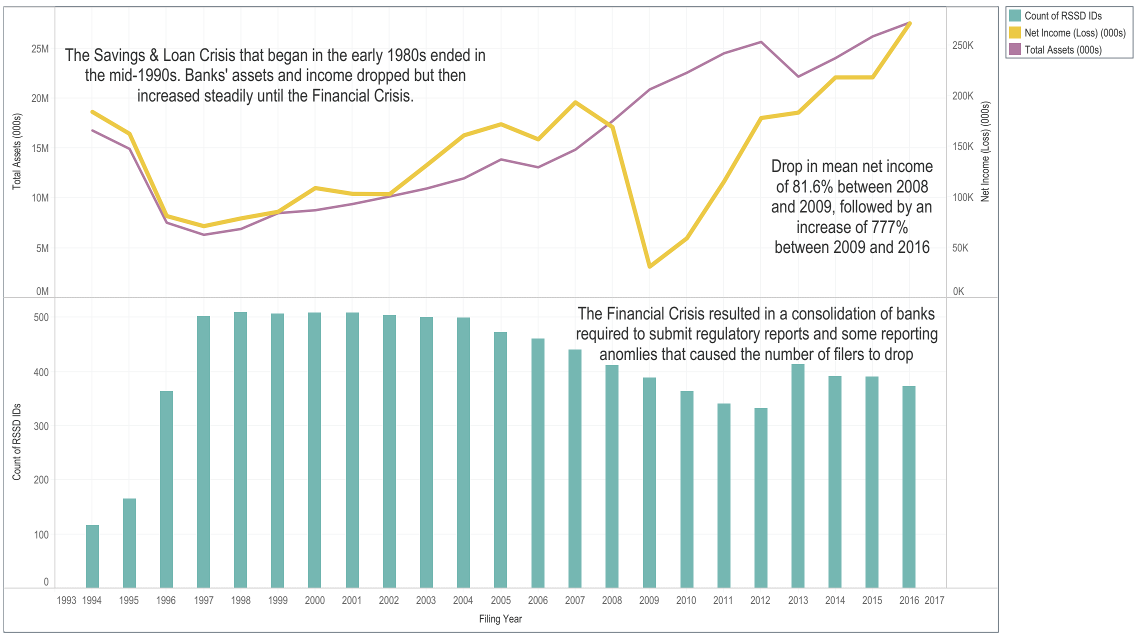

Figure 3. Graph tracking the average total assets and net income or loss (upper half) as well as the number of banks filing regulatory reports (lower half) per year from 1994 through 2016.

Figure 3 provides a graphical interpretation of the regulatory reporting data over the sample period. The upper half tracks the average net income or loss (thick yellow line) and the average total assets (thin purple line) each year. The Savings and Loan Crisis began in the early 1980s and ended in the mid-1990s. Banks' assets and income dropped to a low in the mid-1990s but then both average measures increased steadily until the

Financial Crisis. There was a drop in net income of 81.6% between 2008 and 2009, followed by an increase of 777% between 2009 and 2016. The lower half shows the number of banks submitting regulatory reports (teal bars) each year. The Financial Crisis resulted in a consolidation of banks required to submit regulatory reports. This and some reporting anomalies that precluded banks from inclusion in this sample caused the number of filers to decrease from the peak shown in 2004.

Step 6: Retrieving stock market-related data. I used the banks' PermNo firm identifiers to obtain the stock return and beta data from CRSP (via WRDS) for the all of the sample banks across the sample period. Stock return information was not available for all periods for the full sample, resulting in the number of banks shown in Table 4. I also obtained from WRDS the market returns (R_m) and risk-free rate (R_f) data for each year within the sample period (to be used in capital asset pricing model calculations).

Table 4

Summary Statistics of Bank Stock and Market Information from CRSP

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Annualized Stock Returns	7,514	0.1161	0.3302	(2.1180)	2.7221
Annual Beta Values	7,514	0.4550	0.4746	(1.2152)	3.5475
R_m-R_f	7,514	0.0690	0.1530	(0.2527)	0.2881
R_f	7,514	0.0198	0.0205	-	0.0539

Note: The number of unique banks represented is 797.

For stock performance comparison, I constructed a Bank Index "fund" from bank stocks using the criteria established for the NASDAQ KBW Bank Index. [5] This construction was necessary because the KBW index did not exist for the full sample period. This Bank Index was rebalanced each year according to the criteria. Annualized returns and betas were retrieved from CRSP for each stock in the index. As shown in the upper graph in Figure 4, the Bank Index banks' returns (pink line using the right-hand axis) tracked those of the sample banks' returns (green bars using the left-hand axis). However, in the lower graph the differences in the Bank Index banks' betas (orange line using the right-hand axis) compared to those of sample banks (blue bars using the left-hand axis) were notable and likely due to the smaller asset size and market capitalization of the sample banks. This comparison is also demonstrated in the box and whiskers plot in the lower right side of the figure.

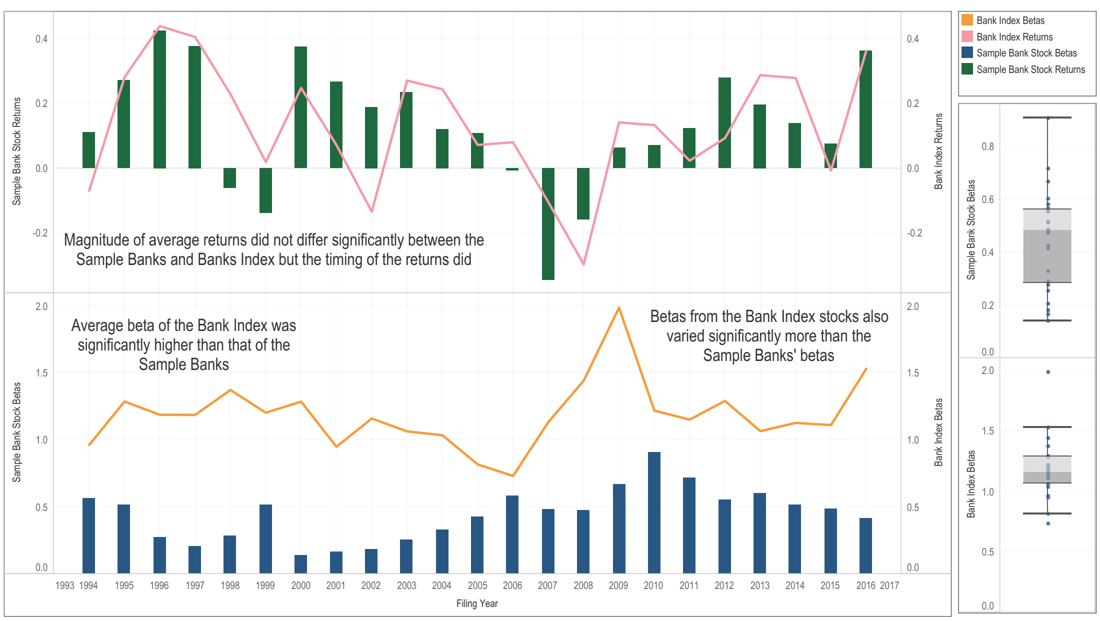

Figure 4. Graph tracking stock returns (upper half) and betas (lower half) of sample banks (LHS) versus those of the Bank Index (RHS) per year from 1994 through 2016.

5 https://indexes.nasdaqomx.com/docs/Methodology_BKX.pdf

4.2 Methods

This section details the methods used in this research, a summary of which appears in Table 5 below to link each method with the hypothesis that the method is employed to confirm.

Table 5

Summary of Methods Employed to Test Each Hypothesis

Method	*H.1*	*H.2*	*H.3*	*H.4*	*H.5*	*H.6*	*H.7*	*H.8*
A: Bank risk measures	X	X	X	X				
B: Systemic risk measures	X	X	X	X				
C: Propensity score matching	X	X	X	X				
D: Structured topic modeling	X	X						
E: Sentiment analysis			X	X	X	X	X	X
F: Confirmatory factor analysis					X	X	X	X
G: Portfolio construction					X	X		
H: Event studies							X	X

Method A: Bank risk measures. Following the methodology of Li et al. (2017), I used data from banks' year-end Form Y-9C to construct the following measures ("Li Risk Measures"). Profit and Loss (P&L) measures are based on Income Statement values and Loan Performance measures. Exposure measures are based on Balance Sheet and Off-Balance Sheet values.

Table 6 presents summary statistics of Total Assets, Net Income (Loss) and the following Li Risk Measures:

- Credit Risk P&L
- Credit Risk Exposure
- Market Risk P&L
- Market Risk Exposure
- Liquidity Risk P&L
- Liquidity Risk Exposure
- Operational Risk P&L
- Operational Risk Exposure

Table 6

Summary Statistics of Li Risk Measures (in $000s)

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Total Assets	8,633	16,500,000	121,000,000	93,180	2,570,000,000
Net Income (Loss)	8,633	149,609	1,074,586	(3,094,179)	24,400,000
Credit Risk P&L	8,633	420,040	2,771,724	(200,160)	62,000,000
Credit Risk Exposure	8,633	12,900,000	99,600,000	40,929	2,500,000,000
Market Risk P&L	8,633	13,033	254,378	(3,642,565)	12,400,000
Market Risk Exposure	8,633	4,415,796	45,500,000	(4,089)	1,770,000,000
Liquidity Risk P&L	8,633	152,248	1,397,582	(58,182)	34,300,000
Liq Risk Exposure	8,633	6,323,215	59,400,000	(4,089)	2,040,000,000
Operational Risk P&L	8,633	(241,669)	1,584,856	(41,400,000)	1,368,739
Op Risk Exposure	8,633	20,100,000	160,000,000	95,100	3,810,000,000

Note: The number of unique banks represented is 925.

Method B: Systemic risk measures. Following the methodology of the Basel Committee (2013) to identify Globally Systemically Important Financial Institutions (G-SIFIs), I used data from banks' year-end Form Y-9C to construct the following components of systemic importance / systemic risk measures. Each of the five components is equally weighted in determining systemic importance, so these were added together to create a *Total_SIFI* variable. Table 7 presents summary statistics of the Systemic Risk (Importance) Measures according to the following components.

- *SIFI_Sub1*: Cross-jurisdictional activity
- *SIFI_Sub2*: Size
- *SIFI_Sub3*: Interconnectedness
- *SIFI_Sub4*: Substitutability / financial institution infrastructure
- *SIFI_Sub5*: Complexity

Table 7

Summary Statistics of Systemic Risk (Importance) Measures (in $000s)

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
SIFI_Sub1	8,633	155,285	1,529,612	-	39,700,000
SIFI_Sub2	8,633	197,543	1,474,380	(28,868)	33,100,000
SIFI_Sub3	8,633	630,534	4,371,332	(172,356)	113,000,000
SIFI_Sub4	8,633	178,572	2,116,440	(26)	57,700,000
SIFI_Sub5	8,633	279,948	2,604,184	-	65,600,000
Total_SIFI	8,633	1,441,882	11,100,000	-	275,000,000

Note: The number of unique banks represented is 925.

Method C: Propensity score matching. In order to assess the statistical significance of topic prevalence and sentiment scores against the presence of EAs for banks in a given year, I performed propensity score matching (PSM) following Ho et al. (2007) using the “MatchIt” package in R statistical software (Ho et al., 2018) to identify three nearest neighbors (*k*=3 NN) for each bank based on two factors, Total Assets and Net Income (Loss), for a given year.

PSM has its origins in the medical field where researchers want to identify a "control" subject to compare against a "treated" subject in order to assess the effectiveness of a medical treatment. Researchers identify two people who are alike (or ideally identical) in key aspects except that one has been treated, for example with the trial drug, and other not treated.

I applied this method by viewing an EA in a given year as the banks being "treated" to match them with a bank with similar Total Assets and Net Income (Loss) that did not have an EA in effect (the "untreated" or "control" sample) that year. Table 8 provides summary statistics of the banks with EAs and their nearest neighbors according to financials, Li Risk Measures and SIFI Risk Measures. Not all banks with EAs could be matched; therefore, the data in the table represents 177 of the 203 banks with EAs in effect at any point in the sample period and a total of 2,273 firm-year observations for each bank-neighbor pair (hence n=4,546 in Table 8).

On average, in comparing the values in Table 8 to those in Tables 6 and 7, one can see that the balance sheet measures in Table 8 are roughly 75% greater than their counterparts in Tables 6 and 7. For example, the mean *Total Assets* in Table 8 is $28.9 billion, while it is $16.5 billion in Table 6. This would indicate that the banks that received EAs across the sample period were on average 75% larger than the total sample population of banks. Similarly, the mean *Total_SIFI* value in Table 8, which represents only EA-impacted banks and their nearest neighbors, is $2.5 billion while this measure is $1.4 billion in Table 7, which represents the full sample population.

Table 8

Summary Statistics of Matched Banks: Those with EAs in Effect During the Sample Period and Their Nearest Neighbors Without EAs (in $000s)

Variable	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Total Assets	4,546	28,900,000	166,000,000	93,180	2,570,000,000
Net Income (Loss)	4,546	253,992	1,467,624	(3,094,179)	24,400,000
Credit Risk P&L	4,546	715,908	3,788,661	(200,160)	62,000,000
Credit Risk Exposure	4,546	22,600,000	136,000,000	40,929	2,500,000,000
Market Risk P&L	4,546	24,257	350,127	(3,642,565)	12,400,000
Market Risk Exposure	4,546	7,949,292	62,500,000	-	1,770,000,000
Liquidity Risk P&L	4,546	275,748	1,916,995	(58,182)	34,300,000
Liq Risk Exposure	4,546	11,300,000	81,400,000	-	2,040,000,000
Operational Risk P&L	4,546	(409,537)	2,166,398	(41,400,000)	644,895
Op Risk Exposure	4,546	35,500,000	220,000,000	95,100	3,810,000,000
SIFI_Sub1	4,546	289,308	2,097,941	-	39,700,000
SIFI_Sub2	4,546	340,551	2,018,157	(28,868)	33,100,000
SIFI_Sub3	4,546	1,074,628	5,982,230	(172,356)	113,000,000
SIFI_Sub4	4,546	322,393	2,881,042	(24)	57,700,000
SIFI_Sub5	4,546	501,585	3,573,510	-	65,600,000
Total_SIFI	4,546	2,528,465	15,100,000	6,334	275,000,000

Note: The number of unique banks represented is 778.

Method D: Structured topic modeling. Pollach (2010) provides a comprehensive introduction to the practice of corpus linguistics and how it can be applied to the computer-aided analysis of textual data. Corpus linguistics is a specialized branch of in the field of linguistics that focuses on the use of language within a specific body (hence "corpus") of literature. For example, Pollach created a corpus using the text of letters from banks to their shareholders in 2006 and 2008. She analyzed this corpus using multiple techniques beyond word frequency and word counts, including text collocations, word distribution and corpus comparisons to identify similarities / differences. Each subsequent technique built upon the prior step so that Pollach produced a list of seven themes from across the letters and used them to characterize tone and sentiment within the corpus.

Hogenboom et al. (2015) discuss the use of rhetorical structure (RS) techniques to ascertain the sentiment contained in both brief and long bodies of text, e.g., from Twitter feeds to annual reports. RS looks at text segments rather than individual words in order to understand the rhetorical role of that segment (as either a nucleus or a satellite) and its relationship to other segments, thereby creating a hierarchical tree structure from text samples.

For this work, I leveraged the Structured Topic Modeling (STM) method (Roberts et al., 2013, 2014) using the "stm" package (Roberts et al., 2018) in R statistical software to analyze 10-K reports and measure the prevalence of topics (see Figure 5). All 10-K reports were reduced to plain text and "cleansed": Numerals were removed, as were punctuation and letter case (converted all letters to lower case). All words were "stemmed" so that only the root of the word remained, e.g., finance, financial and financially would all stem to "financ." The year in which each 10-K report was filed was used as a "prevalence covariate" to track topic prevalence over time and the number of topics was set to k=10.

As shown in Figure 5 below, the stm package builds topics based on the contextual presence of words in documents until the model converges. The algorithm determines the topics, not the user in an application of unsupervised machine learning. Each 10-K report was considered a document (*D* in the figure).

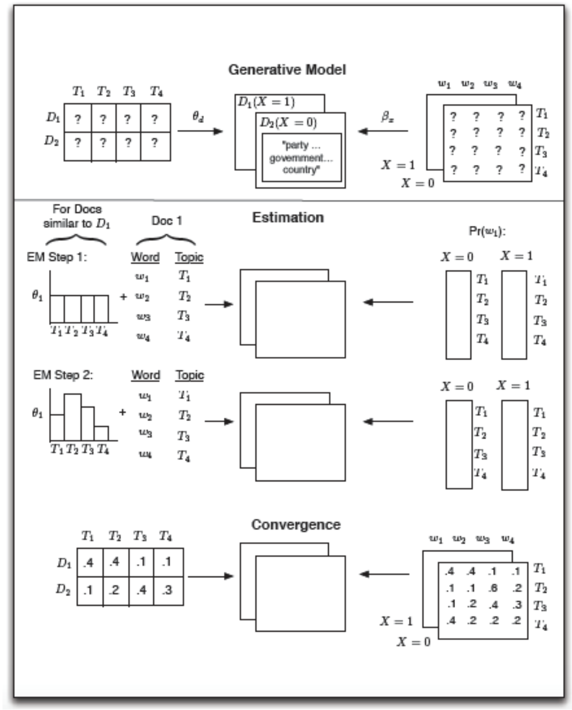

Figure 5. From the stm package vignette, this figure provides a heuristic of the STM process, where *D*=document, *T*=topic, *w*=word.

Four topic models were generated by the stm package, each with 10 topics. A single model was chosen from the four, based on "semantic coherence" (Mimno et al., 2011; Roberts et al., 2014), "exclusivity" (Bischof and Airoldi, 2012; Airoldi and Bischof, 2016), and correlation scores of the 10 topics, this latter measure included in order to provide meaningful topics that were not highly correlated. Based on reviewing words within each topic and their metrics (e.g., probability of occurring, frequency, exclusivity, etc.), titles were developed for each topic.

Leveraging the work of Hargie et al. (2010) in which bankers reverted to more organizational, foundational (and less technical) profiles of their banks following the Financial Crisis, I designated each topic as "technical" or "foundational" based on the same factors used to name each topic. Of the 10 topics, three were designated as "foundational" and seven as "technical."

Table 9 provides this list of topics with their "T/F" designation and statistics related to each topic's prevalence over the sample period. Topic prevalence is measured by how much of each 10-K's content is related to that topic.

Table 9

Summary Statistics of Topics and Related Prevalence Over the Sample Period

Topic and Designation		*n*	*Mean*	*Std Dev*	*Min*	*Max*
1 - Public Company Obligations	F	11,123	0.0278	0.0861	0.0001	0.9557
2 - Loans & Financial Assets	T	11,123	0.1225	0.1955	0.0000	0.9463
3 - Interest Rates' Effect on Income	T	11,123	0.1619	0.1664	0.0000	0.9138
4 - Loans & Capital Requirements	T	11,123	0.1884	0.2132	0.0000	0.9620
5 - Loans & Interest Rate	T	11,123	0.1262	0.1502	0.0000	0.7154
6 - Executives, Employees & Benefits	T	11,123	0.0758	0.1312	0.0001	0.9711
7 - Business of Bank Loans	F	11,123	0.1229	0.1903	0.0000	0.8289
8 - Property & Premises Obligations	T	11,123	0.0199	0.0793	0.0000	0.9708
9 - Business of Being a Public Company	T	11,123	0.1243	0.1706	0.0000	0.9746
10 - Credit Risk Management	F	11,123	0.0301	0.0906	0.0000	0.7682

Note: The number of unique banks represented is 1,037. "T" indicates the topic is designated as "technical" versus "F" indicating the topic is "foundational" in nature.

The correlation matrix shown in Table 10 confirms that the topics developed by the algorithm are not highly correlated, with the highest absolute value at 0.29.

Table 10

Correlation Table to Confirm Topics Are Not Highly Correlated

Topic	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	-0.11	1.00								
3	-0.11	-0.17	1.00							
4	-0.10	-0.24	-0.17	1.00						
5	-0.14	-0.05	-0.12	-0.18	1.00					
6	0.05	-0.12	-0.14	-0.15	-0.18	1.00				
7	-0.09	-0.15	-0.18	-0.20	0.02	-0.18	1.00			
8	0.02	-0.10	-0.10	-0.01	-0.09	0.03	-0.11	1.00		
9	0.00	-0.21	-0.02	-0.17	-0.17	-0.02	-0.29	-0.05	1.00	
10	0.07	-0.13	-0.04	-0.15	-0.17	0.05	-0.01	-0.03	-0.05	1.00

Figure 6 graphically presents the average topic prevalence over time since the 10-K filing year was specifically chosen as a prevalence covariate. The prevalence values that appear along the y-axis are a measure, at the individual 10-K level, of the portion of that report in which the specific topic was discussed.

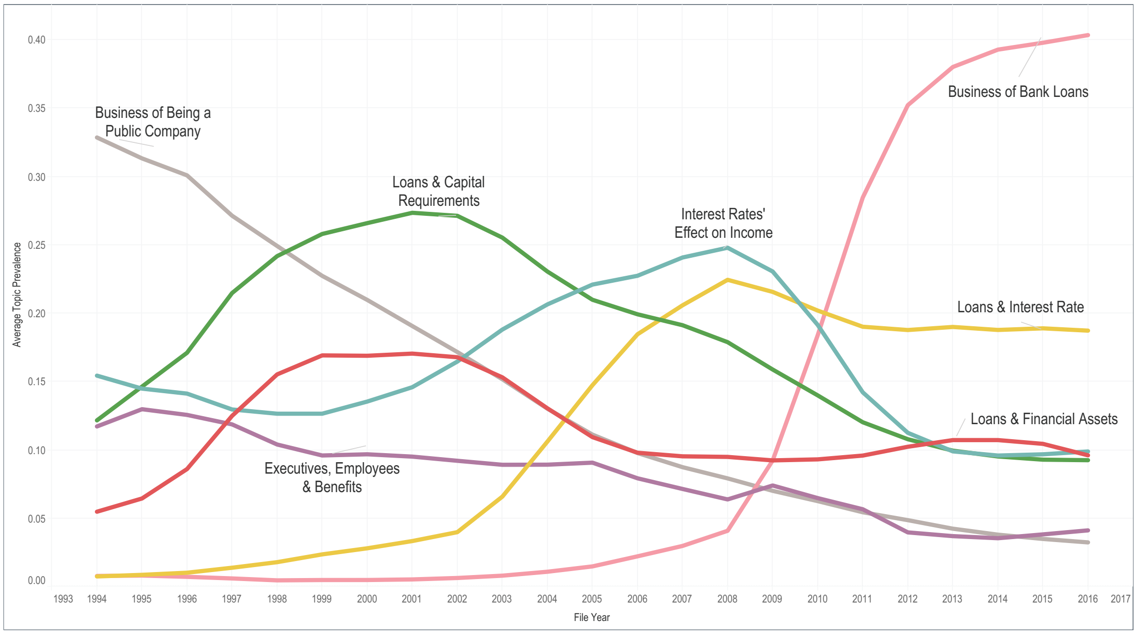

Figure 6. Average topic prevalence by year for each of the seven most prevalent topics developed by the stm package.

It is helpful to note some of the history of the banking environment to gain insights as to the variations in topic prevalence presented in Figure 6 above.[6] In 1994, the financial industry was completing its recovery from the Savings and Loan Crisis of the 1980s and was changing rapidly due to an acceleration of mergers and acquisitions. This is due in large part to the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 being enacted, paving the way for full interstate banking and allowing bank holding companies to acquire banks in any state. According to the FDIC, mergers or consolidations absorbed 550 banks in this year alone while 50 new banks became chartered. Because of their larger size, a number of banks became public companies; hence, we see the top topic from that year discussing the Business of Being a Public

[6] https://www.fdic.gov/about/history/timeline/

Company (Topic 9 in gray) in Figure 7 below, which ranks each topic by its average prevalence in each year of the sample period.

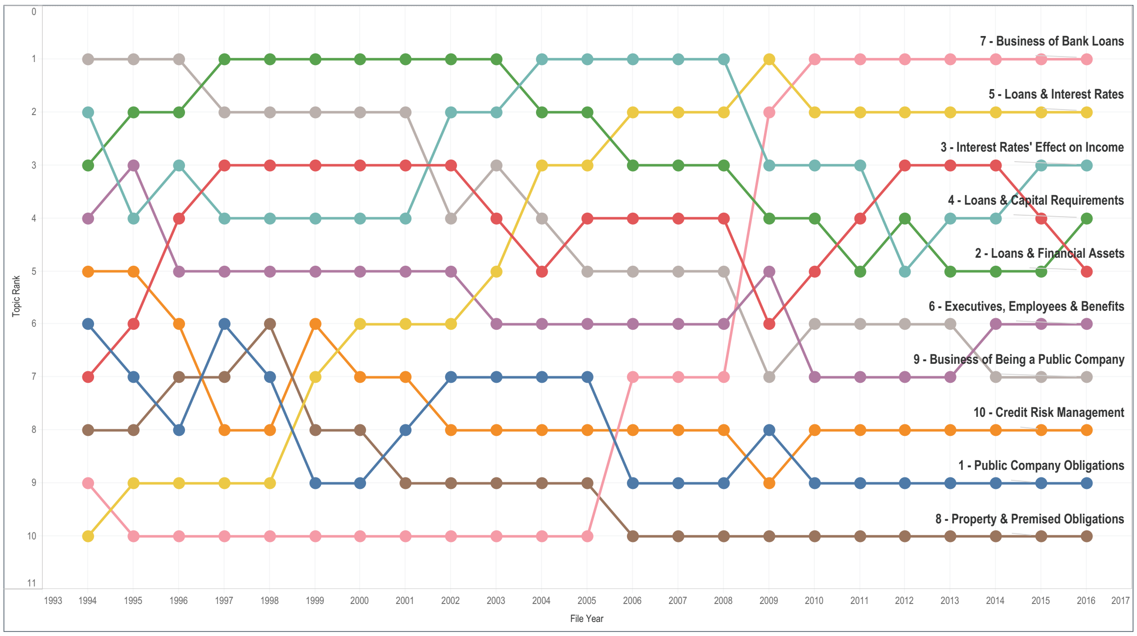

Figure 7. Rankings of average topic prevalence by year for each of the 10 topics developed by the stm package.

Also in Figure 7, we see several topics shift direction following the Dot-Com Bubble in the early 2000s, with banks taking more of a focus on Interest Rates' Effect on Income (Topic 3 in teal) and how rates affect the loans they can offer (Loans and Interest Rates, Topic 5 in yellow). The next large shift in topic prevalence follows the Financial Crisis in the late 2000s, with the Business of Bank Loans (Topic 7 in pink) jettisoning from the least discussed topic to the top topic over the short span of two years. We also see continued discussion on the effect of interest rates and loans, as well a slight increase in the discussion of other financial assets (Loans and Financial Assets, Topic 2 in red) likely as a result of banks' ability to hold and trade assets such as asset-backed securities and credit default swaps. These insights will be leveraged when ascertaining the difference in topic prevalence between banks with EAs in effect or initiated in a given year versus those that do not have EAs.

Method E: Sentiment analysis. As described in the Review of the Literature, sentiment analysis was one of the initial NLP methods and remains one of the most frequently methods employed in text analysis of 10-K reports. As a result, a number of automated tools are available to leverage this technique. For this research, I reviewed several sentiment analysis tools and chose the "syuzhet" package (Jockers, 2017) available via R statistical software. Three of the main criteria I used in selecting this package were:

1. the open source nature of the syuzhet package lends itself to transparency and reproducibility;
2. the algorithms employed by the syuzhet package are based on published peer-reviewed research (including Mohammed and Turney, 2010); and
3. this package provided sentiment scores for eight sentiment categories, whereas other packages provided either a single sentiment score or simply two sentiment scores, one positive and one negative.

All 10-K reports were "cleansed" as described in the STM method section above but they were not stemmed. Sentiment analysis was performed on all 10-K reports following the methodology of Mohammed and Turney (2010) via the syuzhet package.

Scores were generated for each 10-K in each of the following eight categories according to the frequency with which certain words appeared.

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust

Method F: Confirmatory factor analysis. Due to potential endogeneity concerns with the individual sentiment categories, following Bollen (1996), confirmatory factor analysis (CFA) was performed on individual sentiment category scores to create a latent variable *Total_Sent* to represent the full sentiment score for each 10-K. The CFA model and factor loadings appear in Figure 8 below. As will be seen in the summaries of the sentiment data collected, the model demonstrates that the *Total_Sent* score is driven more by the individual sentiment categories of <u>Trust</u> and <u>Fear</u>, while less so by the scores for <u>Surprise</u>, <u>Disgust</u> and <u>Joy</u>.

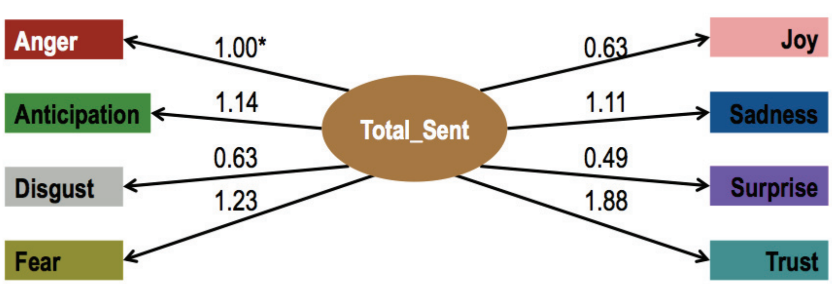

Figure 8. Total_Sent CFA model and factor loadings. * Note that Anger was constrained at 1.00.

This creation of a latent variable to represent a total sentiment score is a departure from prior literature in that typically scores for negative sentiment are subtracted from scores for positive sentiment to obtain an overall sentiment score. This CFA approach was taken, as the traditional approach did not provide meaningful differentiation between 10-K reports' sentiment scores. Many 10-Ks contained a balance of positive and negative sentiments. That balanced approach resulted in many banks' sentiment scores that tended toward zero. As a result of this CFA approach, a "High" *Total_Sent* would indicate the presence of a large amount of positive or negative sentiment, or some combination thereof. The illustrative examples in Figure 9 demonstrate the more meaningful differentiation enabled using a CFA-produced *Total_Sent* score versus the traditional approach.

	Illustrative Sentiment Scoring	
Sample Sentences	**Traditional**	**CFA**
Our loan portfolio is positively outperforming last year by large margins.	100	100
Our management is highly concerned by the receipt of our third enforcement action.	-98	98
Our loan portfolio is outperforming last year by large margins but our management is highly concerned by the receipt of our third enforcement action.	2	198
Our loan portfolio is performing equal to last year.	5	5

Figure 9. Illustrative examples of how the CFA approach in sentiment scoring differs from the traditional approach.

Table 11 summarizes the individual sentiment category scores and the resulting *Total_Sent* score from the CFA. As suggested by the CFA factor loading, Trust is the sentiment most frequently invoked in the 10-K reports across the samples. Note that at least part of this may be due to the frequency with which the word "trust" is used by banks, particularly in the case where the firm provides trust (fiduciary) services as part of an asset and/or wealth management business line. Later hypothesis testing that compares banks to other banks indicates that this double meaning of trust does not invalidate the results shown here.

Table 11

Summary Statistics of Sentiment Category Scores Over the Sample Period

Sentiment	*n*	*Mean*	*Std Dev*	*Min*	*Max*
Anger	11,123	50.29	19.56	4.00	181.00
Anticipation	11,123	96.77	23.32	21.00	234.00
Disgust	11,123	31.78	12.46	3.00	119.00
Fear	11,123	67.97	24.01	5.00	202.00
Joy	11,123	56.74	13.87	10.00	182.00
Sadness	11,123	62.96	21.71	4.00	194.00
Surprise	11,123	33.05	10.24	3.00	92.00
Trust	11,123	186.00	38.25	47.00	403.00
Total_Sent	11,123	1,020.23	265.41	184.00	2,489.00

Note: The number of unique banks represented is 1,037.

Figure 10 provides a graphical interpretation of the individual scores for all 10-K reports analyzed. This representation provides a sense of the dispersion of statistical scores across the 10-K reports over time. In general, the *Total_Sent* scores have drifted higher over time. As previously noted, the Trust sentiment category is the dominant sentiment invoked in the 10-K reports across the samples, shown in the lower half of the figure with its scores increasing over the sample period as demonstrated by the cluster of teal-colored dots pulling above the dots of the other sentiment categories.

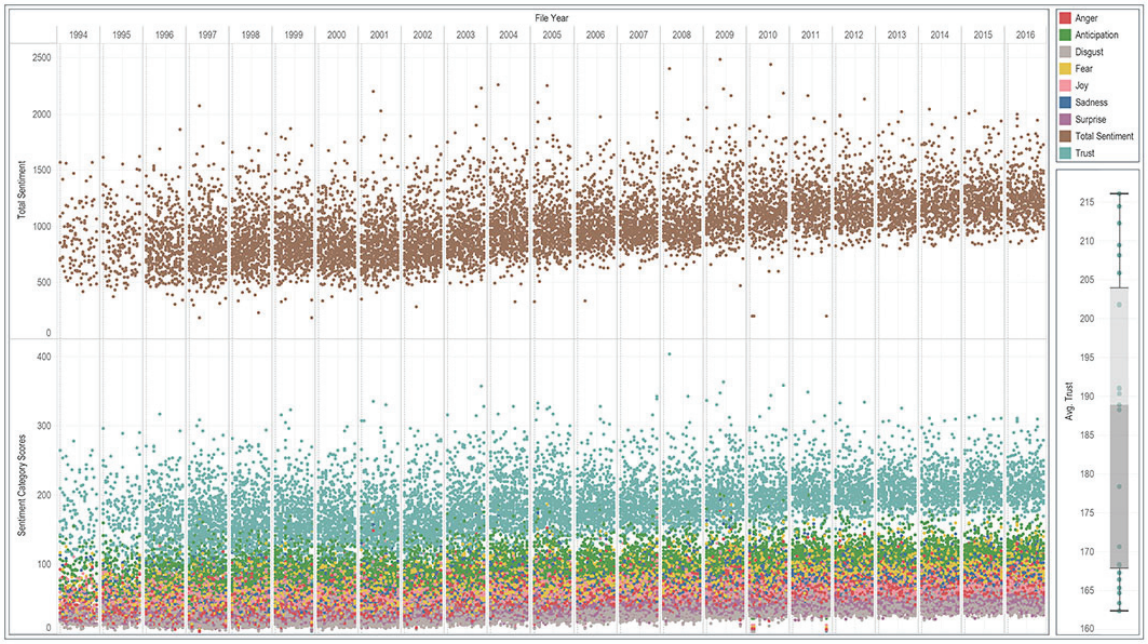

Figure 10. Range and magnitude of sentiment scores by sentiment category. Trust is highlighted as the sentiment category with the widest distribution over the sample period.

Method G: Portfolio construction. Portfolios of bank stocks were constructed leveraging the concept of factors introduced by Fama and French (1993) and executed by Tetlock et al. (2008), who developed their trading strategy by constructing a portfolio based on negative word counts in firm-specific news stories.

For this research, I constructed portfolios based on sentiment by ranking all banks each year according to their *Total_Sent* scores. Based on the 10-Ks filed each year, banks with the top 20% of the High(est) Sentiment scores and banks with the bottom 20% of the Low(est) Sentiment scores were selected for inclusion in several portfolios. Table 12 provides a high-level comparison of the bank stocks chosen for inclusion in High and Low sentiment portfolios. Over the entire sample period, the Low sentiment bank stocks returned nearly 2% more than the High sentiment stocks, which had an mean sentiment 579 points (or 76%) higher than the Low sentiment stocks.

Table 12

Summary of Bank Stocks Chosen for Inclusion in High and Low Sentiment Portfolios

	n	*Mean Returns*	*Mean Sentiment*
High Sentiment Stocks	1,494	0.1190	1,335
Low Sentiment Stocks	1,494	0.1378	756
Difference (High-Low)		-0.0188	579

Some portfolios went exclusively long on High or Low sentiment banks and some went exclusively short on High or Low sentiment banks. Some portfolios went long on High sentiment banks and short on Low sentiments banks and some were vice versa.

Each portfolio described above was constructed using equal weighting, e.g., if there were 10 stocks in the portfolio each was weighted 10%. Each portfolio was then re-created using linear weighting according to the bank's *Total_Sent* score, i.e., the bank with the Highest *Total_Sent* score received the largest weighting. Table 13 summarizes the 12 portfolios constructed on sentiment and includes details on the *Total_Sent* scores that were used as a basis for the bank stocks' selection.

Table 13

Summary of 12 Portfolios Constructed Based on Total_Sent Scores

Portfolio	*Long*	*Short*	*Weighting*	*Mean Sent*	*Min Sent*	*Max Sent*
A	High	Low	Equal	1376	184	2489
B	High	Low	Linear	1072	184	2489
C	High	-	Equal	1376	1007	2489
D	High	-	Linear	1376	1007	2489
E	-	Low	Equal	769	184	1107
F	-	Low	Linear	769	184	1107
G	Low	High	Equal	769	184	2489
H	Low	High	Linear	1072	184	2489
J	Low	-	Equal	769	184	1107
K	Low	-	Linear	769	184	1107
L	-	High	Equal	1376	1007	2489
M	-	High	Linear	1376	1007	2489

Annualized returns were calculated based on July 1 through the following June 30 returns. The portfolio was rebalanced each year based on the sentiment of the 10-K reports that were filed for the prior fiscal year as of March 31 that year, i.e., on July 1, 2015 the portfolio was rebalanced to reflect the sentiment scores from 10-Ks filed in the first quarter 2015 (which reflect fiscal year 2014). Expected and abnormal returns were calculated based on the Capital Asset Pricing Model (CAPM) (Sharpe 1964) as shown in Equation 1 below.

$$E\left(R_p\right) = R_f + \beta_p\left(E\left(R_m\right) - R_f\right) \tag{1}$$

Where $E(R_p)$ is the expected return on the portfolio, R_f is the risk-free rate of the interest, β_p is the portfolio beta (calculated as the weighted average of all individual bank stocks' etas within the portfolio), $E(R_m)$ is the expected return of the market.

To test the impact of an EA and sentiment on subsequent returns, 141 banks with EAs initiated during the sample period were ranked in the year following the fiscal year of their EA according to their sentiment score. The High(est) one third (33%) and the Low(est) 33% banks according to their *Total_Sent* scores were placed in separate portfolios (EA_High and EA_Low). This resulted in 48 firm-years in the High portfolio balanced with 48 firm-years in the Low portfolio. In order for each year to have at least one bank in both the High and Low sentiment portfolios, this sample is limited to the years 2003-2016. Returns were calculated by going long on each of these banks' stocks in the year following receipt of their EA.

As shown in Table 14 over the entire sample period, the EA_Low bank stocks returned over 16% more than the EA_High stocks, which had an mean sentiment 1,583 points (or 56%) higher than the EA_Low stocks. Compared with the values in Table 12, the returns for the EA_High portfolio are 71% lower than those from the High sentiment portfolio, while the returns from the EA_Low portfolio are 43% higher than those from the Low sentiment portfolio. The mean sentiment values for both EA-based portfolios are higher than their counterparts in Table 12 (19% for High and 35% for Low).

Table 14

Summary of Bank Stocks with EAs Chosen for Sentiment Portfolios

	n	*Mean Returns*	*Mean Sentiment*
EA_High Stocks	48	0.0337	1,583
EA_Low Stocks	48	0.1975	1,018
Difference (High-Low)		-0.1639	565

Method H: Event studies. Similar to the portfolio construction described above, based on the 10-Ks filed each year, banks with the top 20% of the High(est) Sentiment scores and banks with the bottom 20% of the Low(est) Sentiment scores were selected for event studies. The High(est) sentiment banks were placed in a "High" category while the Low(est) sentiment banks were placed in a "Low" category. For all banks, based on the 10-K filing year for which they were chosen, I used that 10-K filing date as the event date.

For each sentiment category (High and Low), I performed event studies using Eventus (version 9.0) through WRDS (Cowan, 2010) using 15-, 30-, 45-, 60-, 75- and 90-day windows (symmetrical) from the event date. For each sentiment category, I measured cumulative abnormal average returns (CAARs) for each event window. I also evaluated the CAARs of each group for significance and compared CAARs of the two groups in order to note significant differences. As shown in Table 15 for each event window, the High sentiment banks' mean CAARs were negative while the mean CAARs for the Low sentiment banks were positive.

Table 15

Summary Statistics of CAARS by Sentiment Category and Event Window

Sentiment	*Event Window (in Days)*	*Mean*	*Std Dev*	*Min*	*Max*
High	15	-0.0023	0.0057	-0.0094	0.0107
Low	15	0.0017	0.0026	-0.0022	0.0066
High	30	-0.0046	0.0066	-0.0161	0.0047
Low	30	0.0013	0.0039	-0.0044	0.0099
High	45	-0.0070	0.0066	-0.0211	0.0019
Low	45	0.0023	0.0040	-0.0037	0.0107
High	60	-0.0019	0.0061	-0.0171	0.0065
Low	60	0.0063	0.0041	0.0000	0.0154
High	75	-0.0033	0.0059	-0.0186	0.0059
Low	75	0.0072	0.0048	-0.0033	0.0173
High	90	-0.0112	0.0070	-0.0266	0.0000
Low	90	0.0056	0.0044	-0.0033	0.0157

To test the impact of an EA and sentiment on subsequent returns, event studies were also performed on the 96 banks with EAs initiated during the sample period used for portfolio construction, with 48 banks in the High Sentiment (EA_High) category and 48 banks in the Low Sentiment (EA_Low) category. The 10-K filing date was used as the event date. Similar to above, for each sentiment category, I measured CAARs using 15-, 30-, 45- and 60-, 75- and 90-day windows. In comparing the mean CAARs in Table 16 with those in Table 15, for all but two event window-sentiment category combinations (15-day High and 90-day High), the mean CAARs for the EA-receiving banks were higher.

Table 16

Summary Statistics of CAARS by EA Sentiment Category and Event Window

Sentiment	*Event Window (in Days)*	*Mean*	*Std Dev*	*Min*	*Max*
EA_High	15	-0.0056	0.0198	-0.0339	0.0301
EA_Low	15	0.0047	0.0094	-0.0128	0.0219
EA_High	30	0.0106	0.0223	-0.0285	0.0681
EA_Low	30	0.0224	0.0232	-0.0067	0.0869
EA_High	45	0.0184	0.0281	-0.0286	0.0811
EA_Low	45	0.0518	0.0310	-0.0087	0.1199
EA_High	60	0.0313	0.0307	-0.0211	0.0926
EA_Low	60	0.0522	0.0327	-0.0147	0.1250
EA_High	75	-0.0096	0.0299	-0.0621	0.0526
EA_Low	75	0.0681	0.0358	0.0000	0.1487
EA_High	90	-0.0348	0.0277	-0.0887	0.0243
EA_Low	90	0.0243	0.0337	-0.0363	0.1086

It should be noted that there is an inherent limitation to this approach of using the 10-K disclosure date as the event date. The bank's EA would have already been reported in the prior fiscal year both by the FBA issuing the EA and presumably via a press release or statement from the bank. Therefore, the impact of the reporting of an EA in a 10-K may be somewhat limited, particularly if several months had passed since the EA was issued. Despite this limitation, this approach was the most effective method to provide a direct bank-to-bank comparison of both the initiation of an EA in the prior fiscal period and a 10-K report issuance.

CHAPTER 5

RESULTS

A summary of the results for each hypothesis tested appears in Figure 11. All hypotheses were confirmed as detailed in the following pages. After detailing the results of my hypothesis testing, I will discuss the informative (and even predictive) power of the NLP methods employed here in assessing banks' individual and collective levels of risk relative to the U.S. Financial System.

Hypotheses	**Result**
H1: Technical topics will be <u>less</u> prevalent in the content of banks' 10-Ks that received one or more EAs in the prior fiscal year, compared to peer banks that did not receive EAs that year.	Confirmed
H2: Technical topics will be <u>less</u> prevalent in the content of banks' 10-Ks that have one or more EAs in effect during the sample period, compared to peer banks that did not receive EAs in the sample period.	Confirmed
H3: For banks that receive one or more EAs in a given fiscal year, the sentiment expressed in their 10-K will be <u>higher</u> than peer banks that did not receive EAs that year.	Confirmed
H4: For banks that have one or more EAs in effect during the sample period, the sentiment expressed in their 10-K will be <u>higher</u> than peer banks that did not receive EAs in the sample period.	Confirmed
H5: Based on the sentiment expressed in banks' 10-Ks, portfolios constructed on <u>high</u> (*low*) sentiment will have <u>lower</u> (*higher*) abnormal returns than their peers with <u>low</u> (*high*) sentiment.	Confirmed
H6: Even when comparing only those banks with EAs initiated in the prior fiscal year, based on the sentiment expressed in banks' 10-Ks, portfolios constructed on <u>high</u> (*low*) sentiment will have <u>lower</u> (*higher*) abnormal returns than their peers with <u>low</u> (*high*) sentiment.	Confirmed
H7: Pre- and post-10-K disclosure cumulative average abnormal returns (CAARs) will be <u>lower</u> (*higher*) for firms that employ <u>high</u> (*low*) sentiment in their 10-K report.	Confirmed
H8: Even when comparing only those banks with EAs initiated in the prior fiscal year, pre- and post-10-K disclosure cumulative average abnormal returns (CAARs) will be <u>lower</u> (*higher*) for firms that employ <u>high</u> (*low*) sentiment in their 10-K report.	Confirmed

Figure 11. Review of hypotheses tested and summary of results.

5.1 Hypothesis 1: 10-K Content 1

After performing STM on the full population of 10-K reports to identify the prevalence of the 10 chosen topics for each firm-year observation, I tested Hypothesis 1 by measuring the difference in topic prevalence scores for: a) banks that had an EA initiated in the prior fiscal period (*EA_Initiated*=1) versus all banks without an EA; and b) banks that had an EA initiated in the prior fiscal period versus their PSM-assigned nearest neighbor without an EA. Differences in topic prevalence were statistically significant for most of the 10 topics analyzed.

Across the two comparisons as shown in Table 17 and Figure 12, the topics that increased in prevalence for banks with EAs versus banks without EAs were foundational in nature, i.e., Business of Bank Loans (Topic 7) and Credit Risk Management (Topic 10). In addition, when compared to all banks, those with EAs showed a statistically significant increased prevalence of Public Company Obligations (Topic 1).

In contrast, the topics for which banks with EAs decreased in prevalence versus banks without EAs were largely more technical in nature, i.e., Loans and Financial Assets (Topic 2), Interest Rates' Effect on Income (Topic 3) and Loan and Capital Requirements (Topic 4). These results are consistent with Hargie et al. (2010) in which bankers testifying following the Financial Crisis reverted to more organizational profiles of their institutions rather than discussing their more technical attributes.

Table 17

Difference in Topic Prevalence for Banks with EAs Initiated in a Given Year

Topic and Designation		*EA_Initiated=1 vs All Other Banks*	*EA_Initiated=1 vs Nearest Neighbors*
1 - Public Company Obligations	F	** Increased	Increased
2 - Loans & Financial Assets	T	*** Decreased	*** Decreased
3 - Interest Rates' Effect on Income	T	*** Decreased	*** Decreased
4 - Loans & Capital Requirements	T	*** Decreased	*** Decreased
5 - Loans & Interest Rate	T	Increased	Increased
6 - Executives, Employees & Benefits	T	Decreased	Decreased
7 - Business of Bank Loans	F	*** Increased	*** Increased
8 - Property & Premises Obligations	T	** Decreased	** Decreased
9 - Business of Being a Public Company	T	*** Decreased	*** Decreased
10 - Credit Risk Management	F	*** Increased	*** Increased

Note: ** indicates significance at 0.01, while *** indicates significance at 0.001. “T” indicates the topic is designated as “technical” versus “F” indicating the topic is “foundational” in nature.

Direct comparisons shown in Figure 12 graphically demonstrate that the difference in the prevalence of the topics discussed by banks with EAs initiated in a given year as compared with all other banks (left side chart) and when compared with their nearest neighbors identified through PSM (right side chart). The most obvious difference in both charts is in the prevalence of Business of Bank Loans (Topic 7).

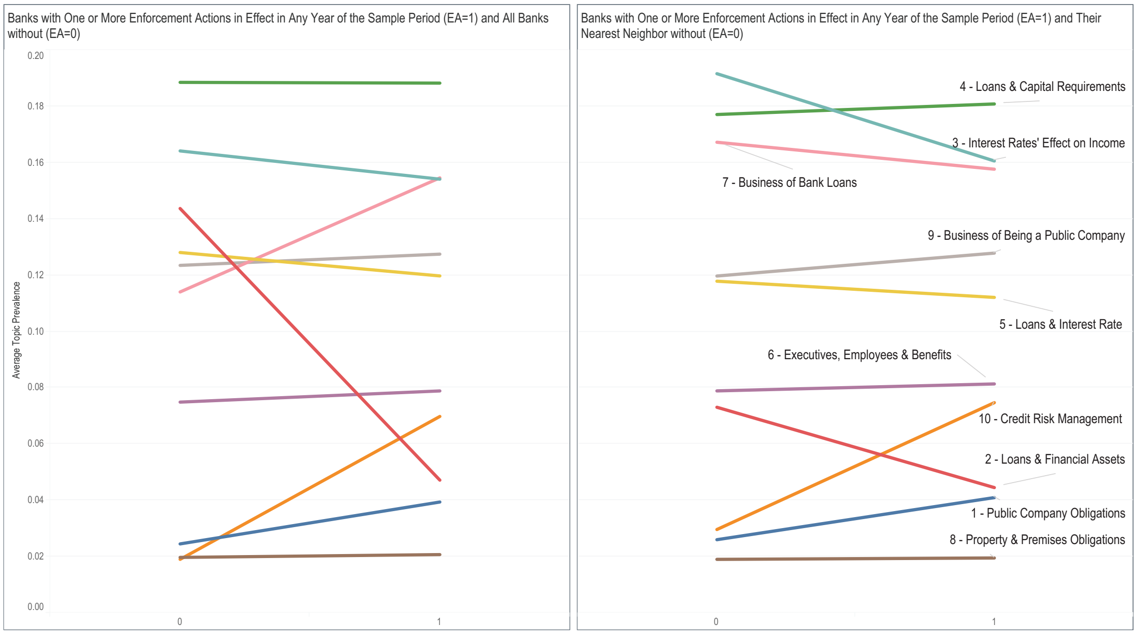

Figure 12. Graphical comparison of topic prevalence based on whether or not an EA was initiated in a given year.

Taken in aggregate, these findings suggests that banks with EAs are signaling to the market that they understand their foundational obligations of being a public company bank while lowering the profile of some technical areas that might signal risky undertakings as they work to address the bank's EA(s).

5.2 Hypothesis 2: 10-K Content 2

I tested Hypothesis 2 by measuring the difference in topic prevalence scores for: a) banks that had an EA in effect in any period (*EA_Period*=1) versus all banks without an EA; and b) banks that had an EA in effect in any period versus their PSM-assigned nearest neighbor without an EA.

Similar to the results from Hypothesis 1, across the two comparisons as shown in Table 18 and Figure 13, the topics that had statistically significant increases in prevalence for banks with EAs versus banks without EAs were foundational in nature, i.e., Public Company Obligations (Topic 1) and Credit Risk Management (Topic 10). In addition, when compared to all banks, those with EAs showed an increased prevalence of Business of Bank Loans (Topic 7). And when compared to their PSM nearest neighbors, banks with EAs in the sample period increased their discussion of Business of Being a Public Company (Topic 9).

Table 18

Difference in Topic Prevalence for Banks with EAs in Effect During the Sample Period

Topic and Designation		*EA_Period=1 vs All Other Banks*	*EA_Period=1 vs Nearest Neighbors*
1 - Public Company Obligations	F	*** Increased	*** Increased
2 - Loans & Financial Assets	T	*** Decreased	*** Decreased
3 - Interest Rates' Effect on Income	T	** Decreased	*** Decreased
4 - Loans & Capital Requirements	T	Decreased	** Decreased
5 - Loans & Interest Rate	T	** Decreased	Decreased
6 - Executives, Employees & Benefits	T	Increased	Increased
7 - Business of Bank Loans	F	*** Increased	** Decreased
8 - Property & Premises Obligations	T	Increased	Increased
9 - Business of Being a Public Company	T	Increased	** Increased
10 - Credit Risk Management	F	*** Increased	*** Increased

Note: ** indicates significance at 0.01, while *** indicates significance at 0.001. "T" indicates the topic is designated as "technical" versus "F" indicating the topic is "foundational" in nature.

Also similar to Hypothesis 1 results, the topics that decreased significantly in prevalence were more technical in nature, i.e., for both comparisons Interest Rates' Effect on Income (Topic 3) and Loans and Financial Assets (Topic 2) decreased in prevalence, while when compared to all banks, the EA-impacted banks' prevalence of Loans and Interest Rate (Topic 5) decreased.

Different from Hypothesis 1 testing, there was one topic that moved in an opposite direction in comparing banks with EAs first to all banks and then to the PSM-assigned nearest neighbors, Business of Bank Loans (Topic 7). In the comparison with all banks, those banks with EAs increased the prevalence of Topic 7 but when compared with their nearest neighbor, they decreased the prevalence of this topic. Upon a closer inspection as shown in Figure 13, this "decrease" a reflection of how much more the EA-impacted banks' nearest neighbors were discussing Business of Bank Loans (Topic 7) than the rest of the banks that did not receive EAs, suggesting that banks of this size and risk profile will discuss this topic more frequently in their 10-K.

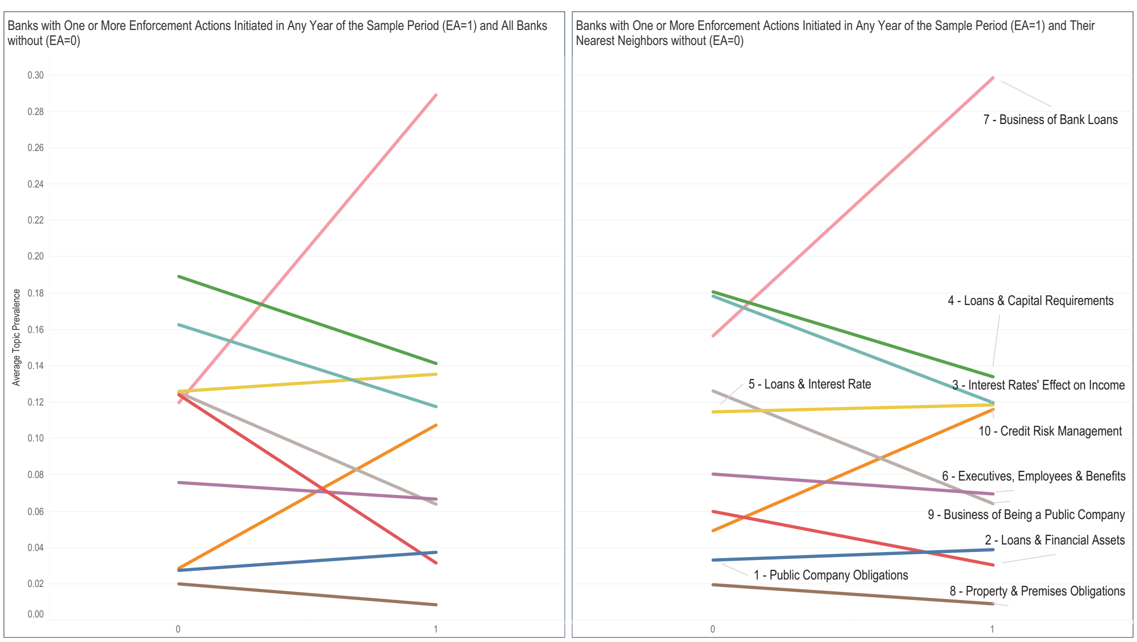

Figure 13. Graphical comparison of topic prevalence based on whether or not an EA was in effect during the sample period.

Direct comparisons shown in Figure 13 graphically demonstrate that the difference in the prevalence of the topics discussed by banks with EAs in effect at any time in the sample period as compared with all other banks (left side chart) and when compared with their nearest neighbors identified through PSM (right side chart).

The starkest contrast is the decrease in discussion of Loans and Financial Assets (Topic 2) shown on the left, in combination with the simultaneous increase in discussing Business of Bank Loans (Topic 7) and Credit Risk Management (Topic 10). Taken in combination, an interpretation could be that banks with EAs in effect are trying to de-emphasize loans as merely a balance sheet line item and communicate their management of the risk that loans represent.

5.3 Hypothesis 3: 10-K Sentiment 1

After performing sentiment analysis on the full population of 10-K reports to obtain eight sentiment category scores, I tested Hypothesis 3 by measuring the difference in sentiment category scores for: a) banks that had an EA initiated in the prior fiscal period (*EA_Initiated*=1) versus all banks without an EA; and b) banks that had an EA initiated in the prior fiscal period versus their PSM-assigned nearest neighbor without an EA.

As shown in Table 19, for both of these comparisons, banks with EAs had higher sentiment scores across all eight sentiment categories; all differences were statistically significant. The sentiment category with the greatest difference across these tests was Trust. In both tests, Fear was one of "Top 3" differences, while Anticipation and Sadness also appeared in the Top 3. In aggregate, banks with EAs are making greater efforts through the words of their 10-K reports to evoke trust from their stakeholders while still subtly expressing concerns, likely related to their EAs

Table 19

Difference in Sentiment for Banks with EAs Initiated in a Given Year

Sentiment Category	*EA_Initiated=1 vs All Other Banks*	*EA_Initiated=1 vs Nearest Neighbors*
Anger		
Anticipation	Top 3	
Disgust		
Fear	Top 3	Top 3
Joy		
Sadness		Top 3
Surprise		
Trust	Top 3	Top 3

Note: All differences were significant at 0.001.

Direct comparisons shown in Figure 14 confirm that the difference in the sentiment evoked by banks with enforcement actions initiated versus those without EAs is statistically significant for each of the sentiment categories. This held when banks with EAs initiated in a given year were compared with all other banks (chart on left side) and when compared with their nearest neighbors identified through propensity score matching (chart on right side). The slope of the lines in the graphs on both sides are very similar, with the slope for the Trust category being larger on the left, perhaps indicating that compared to all banks, banks with one or more EAs initiated in a given year are making more of an effort to evoke trust from their investors and stakeholders.

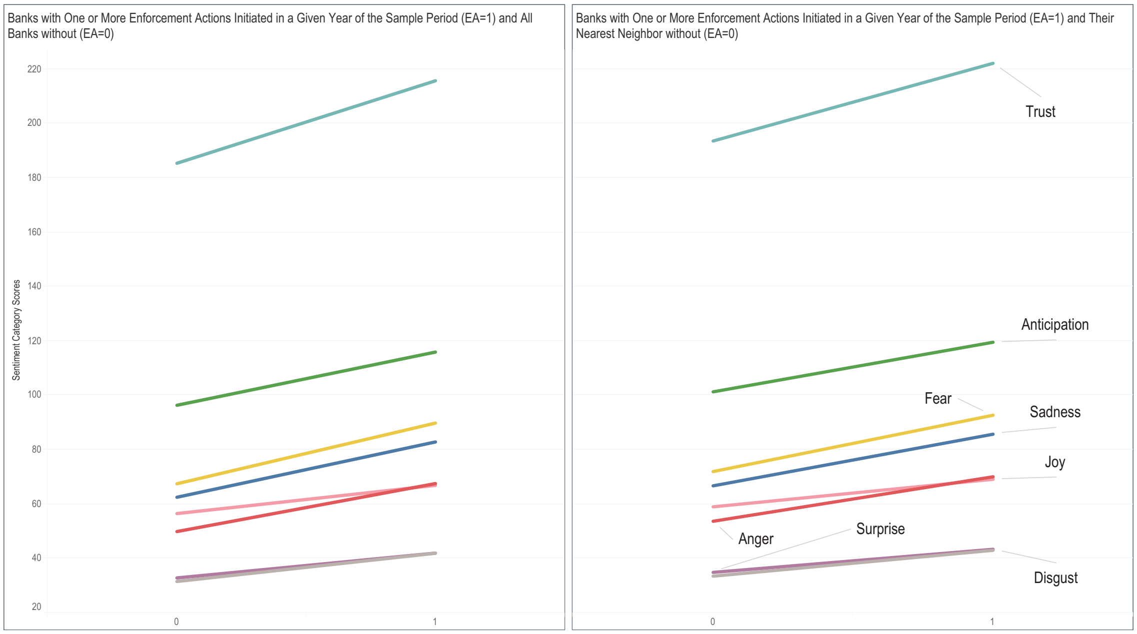

Figure 14. Graphical comparison of sentiment category scores based on whether or not an EA was initiated in a given year.

5.4 Hypothesis 4: 10-K Sentiment 2

I tested Hypothesis 4 by measuring the difference in sentiment category scores for: a) banks that had an EA in effect in any period (*EA_Period*=1) versus all banks without an EA; and b) banks that had an EA in effect in any period versus their PSM-assigned nearest neighbor without an EA.

Similar to the results for Hypothesis 3, as shown in Table 20, for both of these comparisons, banks with EAs had higher sentiment scores across all eight sentiment categories; all differences were statistically significant. The sentiment category with the greatest difference across these tests was Trust. Compared with all other banks, Anticipation and Fear were the other "Top 3" differences, while in comparison with PSM-assigned nearest neighbors, Anger and Sadness also appeared in the Top 3. As was seen in Hypothesis 3 results, banks with EAs are making greater efforts through the words of their 10-K reports to evoke trust from their stakeholders invoking negative sentiments, perhaps due to the EAs, in nuanced terms.

Table 20

Difference in Sentiment for Banks with EAs in Effect During the Sample Period

Sentiment Category	*EA_Period=1 vs All Other Banks*	*EA_Period=1 vs Nearest Neighbors*
Anger		Top 3
Anticipation	Top 3	
Disgust		
Fear	Top 3	
Joy		
Sadness		Top 3
Surprise		
Trust	Top 3	Top 3

Note: All differences were significant at 0.001.

As shown in Figure 15, even when a bank had an EA in effect during the sample period but not necessarily initiated in that (prior fiscal) year, the difference in the sentiment used in their 10-K was significant both compared to all other banks (chart on the left side) and their PSM-assigned nearest neighbors (chart on the right side). The slopes of the lines on both sides of next page are quite similar, with some subtle differences in that the slopes of lines for Trust, Anticipation and Anger are slightly steeper on the right side. Taken in combination, an interpretation could be that banks with EAs initiated in a given year are, compared with all those who have an EA in effect in the period, are venting some frustration at their receipt of an EA while signaling to their stakeholders that trust in them remains warranted.

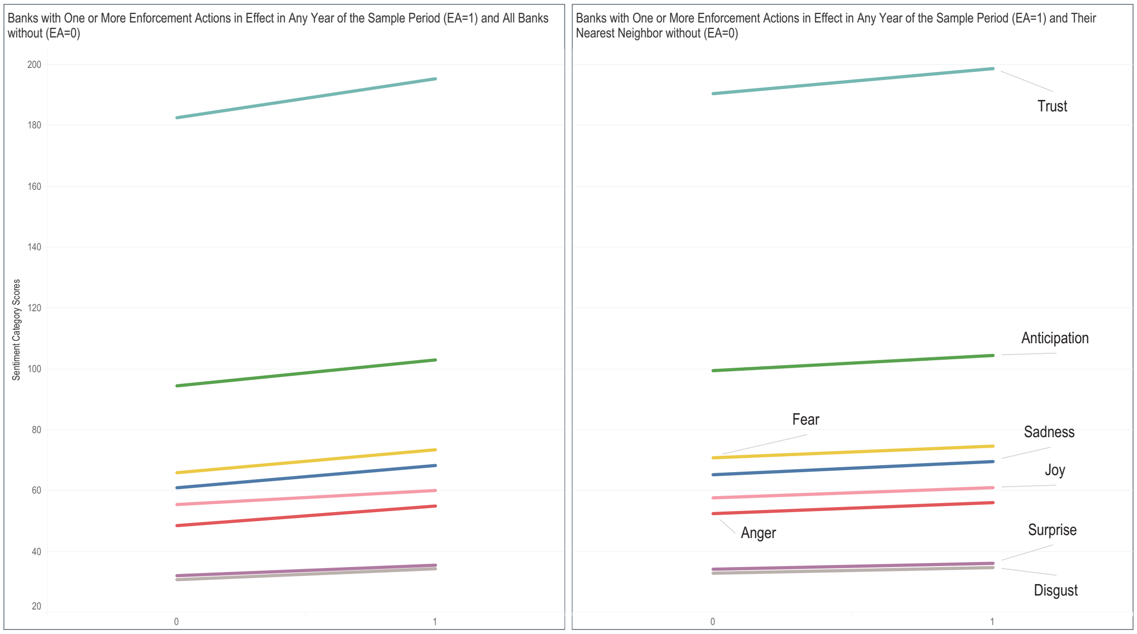

Figure 15. Graphical comparison of sentiment category scores based on whether or not an EA was in effect during the sample period.

5.5 Hypothesis 5: Sentiment-Based Portfolios 1

For Hypothesis 5 testing, I compared the portfolio returns, expected returns and abnormal returns for the 12 portfolios constructed on High versus Low Sentiment (described in Method G). Confirming the hypothesis, as shown in Table 21, the portfolios that went long on Low Sentiment banks fared better when directly compared to those that went long on High Sentiment banks. This indicates that banks that employed Low Sentiment in their 10-Ks experienced more favorable market performance over the sample period, despite some losses that will be described in the upcoming pages.

The best-performing portfolio Portfolio J, which went exclusively long on Low Sentiment, returned 1,403% in 2017 on $10,000 invested in 1994. The worst-performing portfolio Portfolio E, which went exclusively short on Low Sentiment, lost 98% of its value.

Table 21

Summary of Sentiment-Based Portfolio Results

Port	*Long*	*Short*	*Weighting*	*End Value*	*Port Return*	*Exp Return*	*Abn Return*	*Port Beta*	*Mean Sent*	*Min Sent*	*Max Sent*
A	High	Low	Equal	$ 6,570	-34%	659%	-693%	0.4124	1,376	184	2,489
B	High	Low	Linear	$ 7,627	-24%	807%	-830%	0.4866	1,072	184	2,489
C	High	-	Equal	$ 96,299	863%	696%	167%	0.7050	1,376	1,007	2,489
D	High	-	Linear	$ 103,906	939%	820%	119%	0.7681	1,376	1,007	2,489
E	-	Low	Equal	$ 198	-98%	-15%	-83%	(0.2926)	769	184	1,107
F	-	Low	Linear	$ 222	-98%	-13%	-85%	(0.2814)	769	184	1,107
G	Low	High	Equal	$ 13,554	36%	8%	27%	(0.4143)	769	184	2,489
H	Low	High	Linear	$ 11,142	11%	-15%	26%	(0.4869)	1,072	184	2,489
J	Low	-	Equal	$ 150,293	1403%	252%	1151%	0.2914	769	184	1,107
K	Low	-	Linear	$ 141,904	1319%	245%	1074%	0.2811	769	184	1,107
L	-	High	Equal	$ 273	-97%	-75%	-22%	(0.7056)	1,376	1,007	2,489
M	-	High	Linear	$ 237	-98%	-81%	-17%	(0.7680)	1,376	1,007	2,489

Table 22, as well as Figures 16 and 17 provide details on the two portfolios that went exclusively long on either High (Portfolio C) or Low (Portfolio J) Sentiment using an equal-weighting for the stocks in the portfolio each year. Both portfolios significantly outperformed the market as shown by their abnormal returns of 167% (C-High) and 1151% (J-Low).

Table 22

Comparison of Volatility Measures for Portfolios B, C (High Sentiment) and K, J (Low Sentiment)

Portfolio	*Std Dev*	*Peak-to-Trough*
C-High Sentiment Equal Weighting	0.204	1.74^
J-Low Sentiment Equal Weighting	0.178^	1.88
B-High Sentiment Linear Weighting	0.206	1.97
K-Low Sentiment Linear Weighting	0.175^	1.84^

Note: ^ indicates the lower value when comparing volatility measures between the High and Low sentiment portfolios.

Measures of volatility for these two equal-weighted sentiment portfolios (C-High and J-Low) were mixed as shown in Table 22 above, where lower values (noted by ^ in the table) indicate less volatility. The Low sentiment portfolio (J) had a lower standard deviation (0.178 versus 0.204 for Portfolio C) but a larger peak-to-trough measure (1.88 for J versus 1.74 for C). However, when directly comparing the two linear-weighted sentiment portfolios that went exclusively long on one sentiment (B-High and K-Low), the Low sentiment portfolio was less volatile by both measures.

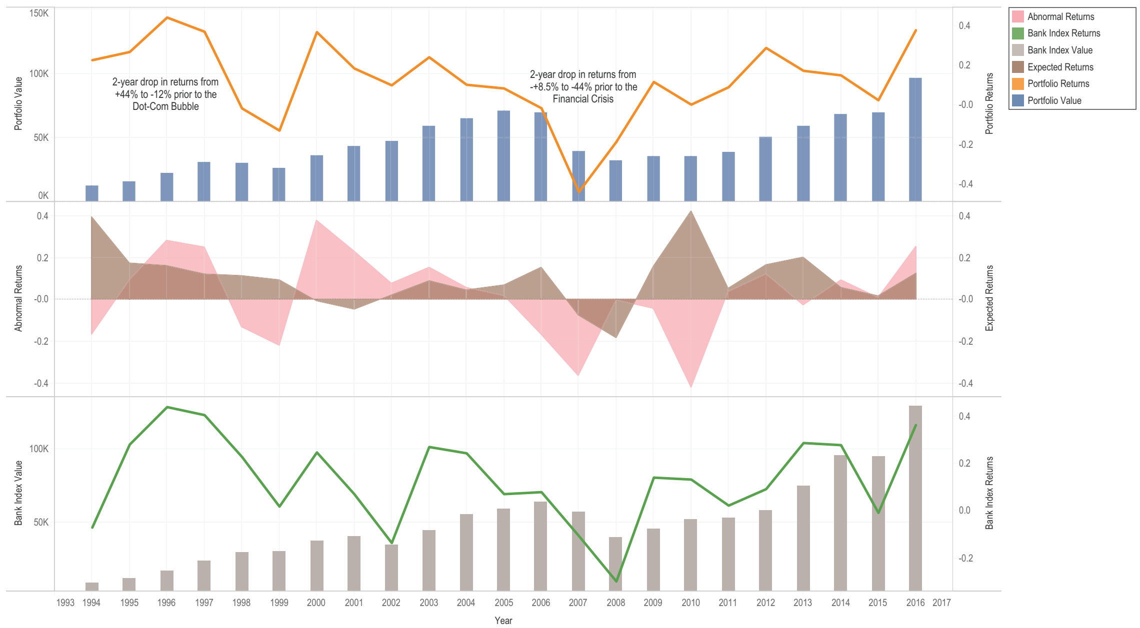

Figure 16. Detailed returns and Bank Index comparisons for Portfolio C (Long High Sentiment Equal-Weighted).

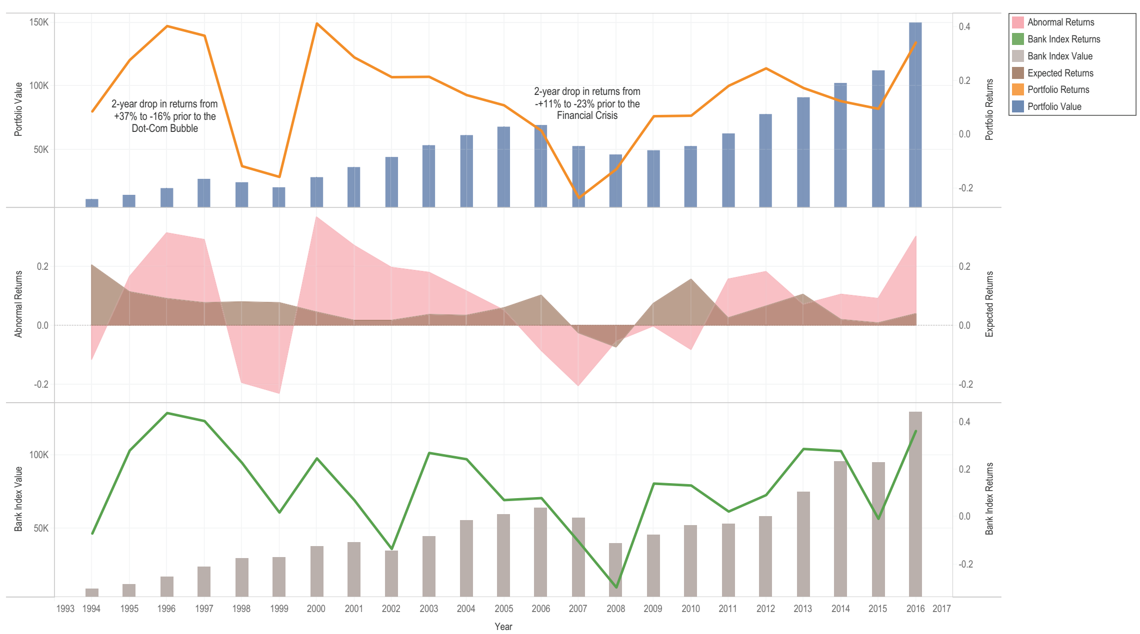

Figure 17. Detailed returns and Bank Index comparisons for Portfolio J (Long Low Sentiment Equal-Weighted).

An interesting and perhaps informative feature of both Portfolios C and J, shown in Figures 16 and 17 respectively, is that both of them suffered significant losses 18 months before the Dot-Com Bubble in the late 1990s / early 2000s and the Financial Crisis of 2008. The drop in the returns of Portfolio C-High in each of those periods (pre-Dot-Com returns fell from +44% to -12%; pre-Financial Crisis returns fell from +8.5% to -44%) were greater than those of Portfolio J-Low (pre-Dot-Com returns fell from +37% to -16%; pre-Financial Crisis returns fell from +11% to -23%) in both cases, offering a potential vehicle for insights into future market crashes. In looking at the timing of losses in the Bank Index portfolio in relation to these two events (Dot-Com Bubble and the Financial Crisis), we can see that those occurred after the declared events. Because the Bank Index is comprised of bank stocks of large firms with comparatively greater market capitalization than many of the banks in the Sample Bank portfolios created, the losses that occurred prior to each crash could be attributed to losses in mid-size bank stocks, suggesting that overall performance of these banks could be a leading indicator of the overall health of the financial system.

5.6 Hypothesis 6: Sentiment-Based Portfolios 2

Figure 18 shows the results of Hypothesis 6 testing with the portfolios constructed with banks with EAs initiated in the prior fiscal year. Similar to the previous portfolio results, the Low sentiment portfolio out-performed the High portfolio significantly. The two long charts within the figure with red (upper) and green (lower) bubbles show the distribution of individual bank returns in the High and Low sentiment portfolios respectively. The two tall charts with the box and whiskers plots show the range of portfolio returns (left) and *Total_Sent* scores (right) each year. The two boxes at the bottom of the page show the portfolio values in which one can see that the value of the Low portfolio was over \$133,000 on an initial \$10,000 investment in 2003, compared to the ending value of the High portfolio of just over \$8,200.

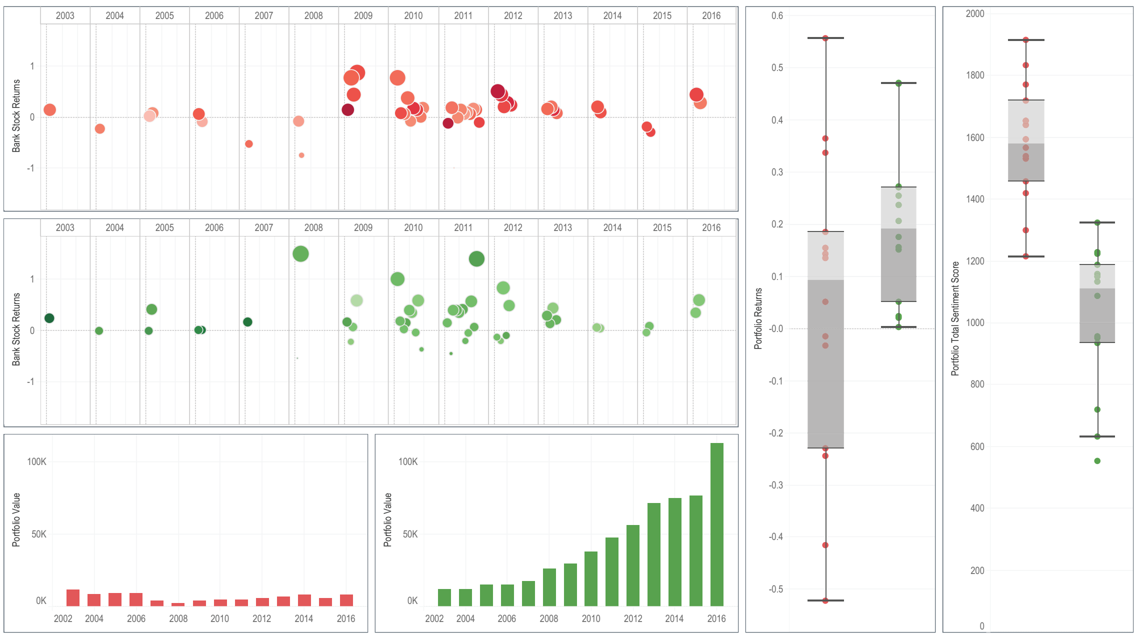

Figure 18. Detailed returns for portfolios of banks with EAs initiated in a given year, divided in High (red) and Low (green) sentiment portfolios.

5.7 Hypothesis 7: Sentiment-Based Event Studies 1

For Hypothesis 7, Figures 19 and 20 show the results of the event studies based on banks placed in the High or Low sentiment portfolios. The event date used is the 10-K disclosure date. In Figure 19, the box and whiskers plots demonstrate that the differences in cumulative average abnormal returns (CAARs) are statistically significant for each of the event windows: 15, 30, 45, 60, 75 and 90 days between the High and Low sentiment banks. For all event windows the CAARs for the Low sentiment banks were higher than those of the High sentiment banks. It should also be noted that the CAARs for both the High and Low sentiment banks were statistically significant for all event windows.

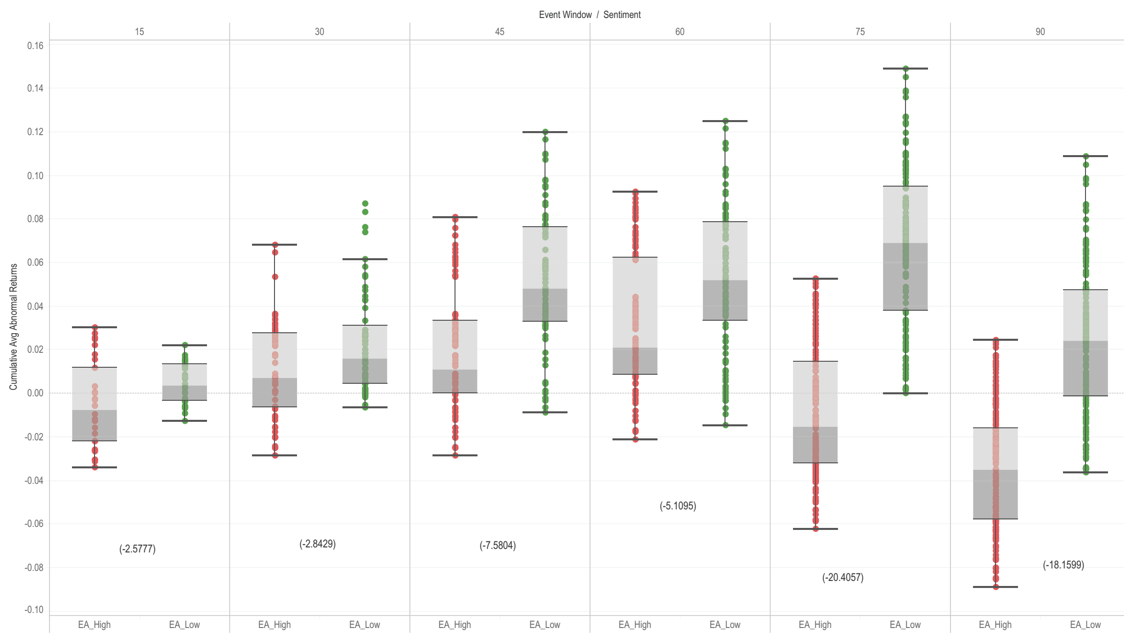

Figure 19. Direct comparison of CAARs across all event windows studied for High (red, left side of each chart) and Low (green, right side) sentiment bank stocks (in all cases, differences are statistically significant, with t-statistics noted).

In Figure 20, the following pattern emerges for the High sentiment bank stocks:

- They experience losses (from the zero basis starting point) leading up to their 10-K disclosure date.
- They experience gains immediately following their 10-K release, in line with Griffin (2003).
- The gains flatten around Day 15 following disclosure.
- By Day 35 following their 10-K release, they begin experiencing losses.
- This trend continues for the duration of the event studies.

In contrast, the Low sentiment bank stocks are flat through their 10-K disclosure date, experience modest gains through Day 30 after their disclosure and then decrease slightly before flattening.

The High sentiment patterns are consistent with the literature in market's ability to absorb information from textual disclosures (Beaver, 1968; You and Zhang, 2009; Kravet and Muslu, 2013). We see that the market reacts positively and quickly to the High sentiment but then that reaction is short-lived and even reverses. The more subdued but sustained positive reaction to the Low sentiment bank stocks could be demonstrating that the market is reacting to the bank's company fundamentals, perhaps in terms of their financials, more so than the sentiment that they are conveying in their 10-K report.

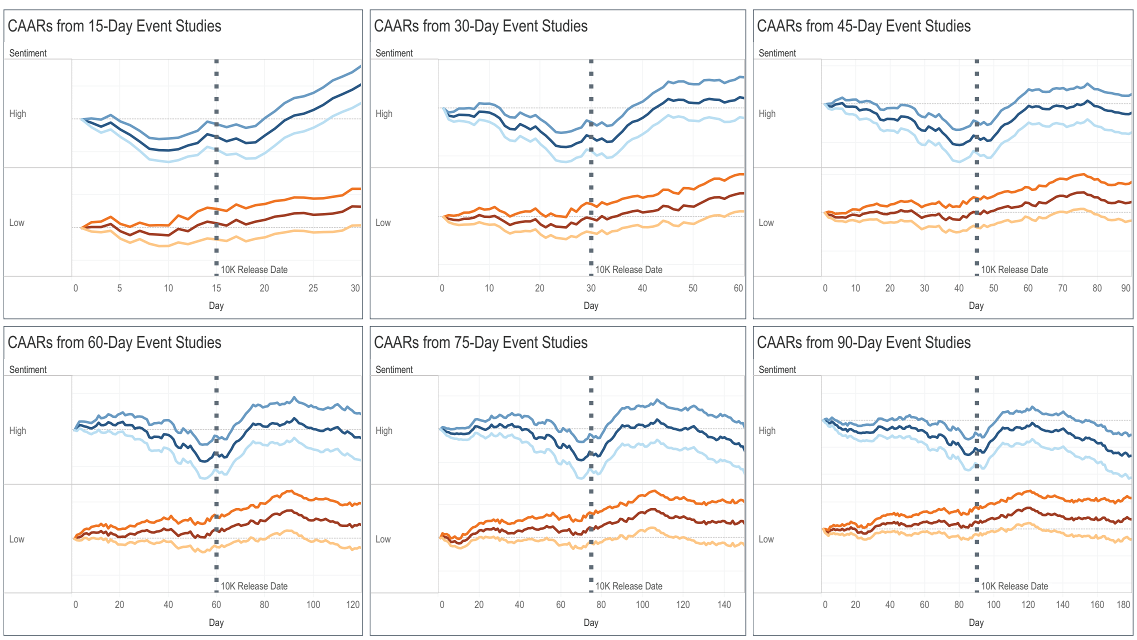

Figure 20. Graphs of CAARs for all event windows studied for High (upper half of each block) and Low (lower half) sentiment bank stocks.

5.8 Hypothesis 8: Sentiment-Based Event Studies 2

For Hypothesis 8, the event studies summarized in Figures 21 and 22 focus on the banks that received an enforcement action in the fiscal year just prior to the 10-K disclosure date that is used as the event date. Banks were split into High and Low sentiment groups, following the methodology used for portfolio construction.

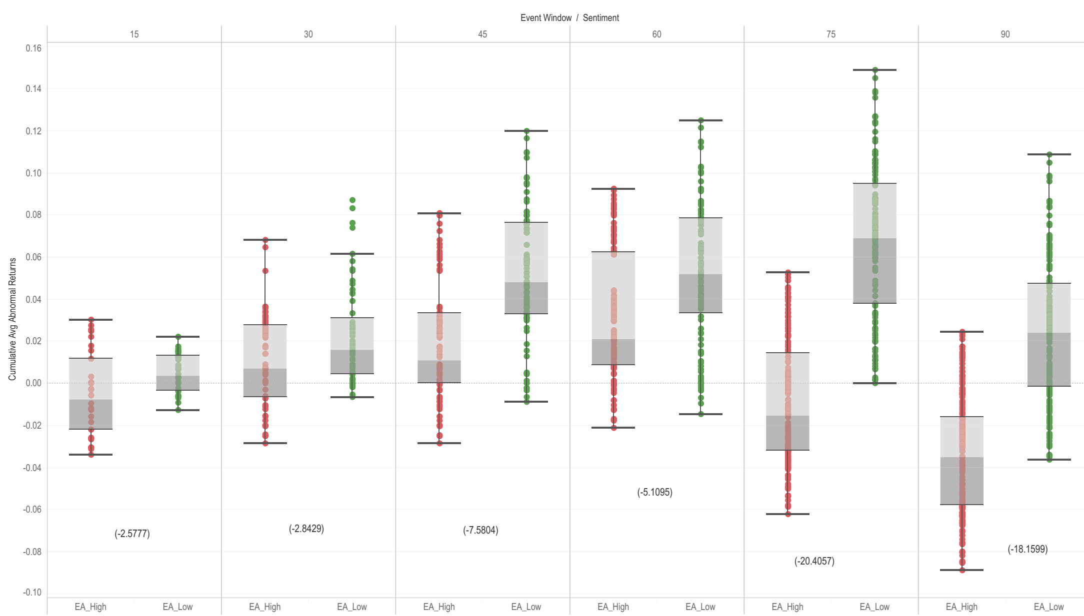

Figure 21. Direct comparison of CAARs across all event windows studied for High (red, left side of each chart) and Low (green, right side) sentiment bank stocks for those banks with an EA initiated in the year prior to the event date (in all cases, differences are statistically significant, with t-statistics noted).

In Figure 21, the box and whiskers plots demonstrate that the differences in cumulative average abnormal returns (CAARs) are statistically significant between the High and Low sentiment banks for each of the event windows: 15, 30, 45, 60, 75 and 90 days. For all event windows the CAARs for the Low sentiment banks were higher than those of the High sentiment banks. It should also be noted that the CAARs for both the High and Low sentiment banks were statistically significant for all event windows.

In Figure 22, a pattern emerges for the High sentiment bank stocks that is similar to the prior event studies:

- They experience losses (from the zero basis starting point) leading up to their 10-K disclosure date.
- They experience gains immediately following their 10-K release, in line with Griffin (2003).
- The gains flatten around Day 35 following disclosure, which is a little later than in the prior studies.
- By Day 50 following their 10-K release, they begin experiencing losses.
- This trend continues for the duration of the event studies.
- In contrast, the Low sentiment banks with EAs:
    - Are flat through their 10-K disclosure date.
    - Experience modest gains that peak around Day 28 after their disclosure.
    - Decrease slowly to return to the zero basis at Day 90 following their 10-K disclosure.

The patterns for banks with EAs demonstrates that the market still reacts positively to the high sentiment initially but then seems to become aware of (or reminded of) the negative news of the EA, since EAs would have been previously disclosed publicly by an FBA and presumably discussed by the bank in a press release or company statement. Low sentiment banks still experience higher CAARs than High sentiment banks and they return to a zero base once the market absorbs their news.

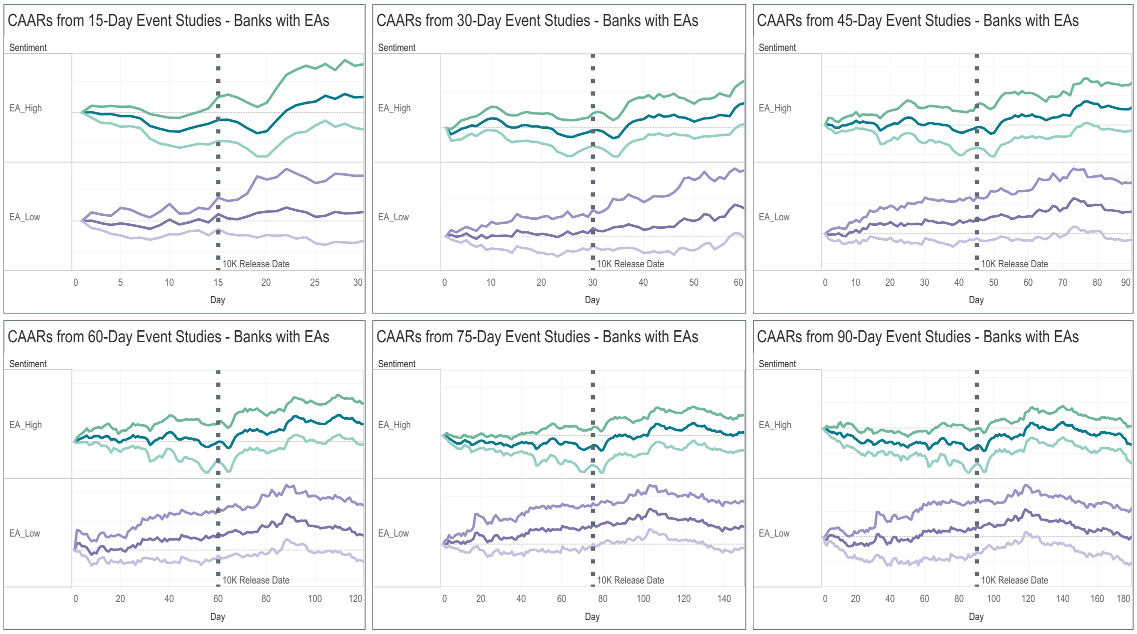

Figure 22. Graphs of CAARs for all event windows studied for High (upper half of each block) and Low (lower half) sentiment bank stocks with EAs initiated in the year prior to the event date.

5.9 Risk in the U.S. Financial System

To turn to the prospect of using the output from NLP techniques as means to help measure the level of risk in the U.S. Financial System, results from each of the tests performed in this research are summarized in the figure below with commentary on the insights that the natural language processing (NLP) method(s) employed provided on systemic risk.

Test	Commentary on Results
H.1 and H.2: 10-K Content	• Differences in the topic prevalence between banks with enforcement actions (EAs) in effect and initiated in the sample period and those without EAs initiated were statistically significant. • Hanley and Hoberg (2018) identified the "build-up" of systemic risk leading up to the Financial Crisis through the prevalence of key topics. • Similarly, one can see that the movement of topics prior to and in reaction to the Financial Crisis, such as *Topic 7-Business of Bank Loans*, could be a proxy for systemic risk.
H.3 and H.4: 10-K Sentiment	• Differences in the *Total_Sent* scores and individual sentiment categories between banks with enforcement actions (EAs) in effect and initiated in the sample period and those without were statistically significant. • Among all banks, increases in their *Total_Sent* scores and in their evocation of sentiments such as *Trust, Anger and Anticipation*, could provide insights into their individual and collective level of risk.
H.5 and H.6: Sentiment-Based Portfolio Construction	• These tests provided the most stark revelations on the potential use of NLP techniques in assessing systemic risk with dual crashes of the Long-High and Long-Low sentiment portfolios (Portfolios C and J) 18 months before the Dot-Com Bubble and the Financial Crisis. • Comparing these results to those of the Bank Index suggest that mid-size banks could be leading indicators for the next systemic crisis. • With consistently higher returns on Low sentiment bank stocks when compared directly to High sentiment bank stocks, it also stands to reason that when they falter, the market will follow.
H.7 and H.8: Sentiment-Based Event Studies	• For all event windows, whether or not a bank had an EA in effect or initiated in the sample period, Low sentiment bank stocks had higher CAARs than High sentiment bank stocks (statistically significant). • Similar to the observation above, changes in the pattern of CAARs for Low sentiment banks could be an indicators of shifts in systemic risk.

Figure 23. Commentary on the systemic risk insights from the tests performed.

These results indicate opportunities exist to further refine these methods to develop indicators, and potentially predictors, of systemic risk in the U.S. financial system.

CHAPTER 6

CHALLENGES AND OPPORTUNITIES FOR FUTURE RESEARCH

While the results of this research clearly demonstrated the utility and capability of NLP methods to provide unique insights as to risks in the U.S. financial system by way of the publicly traded banks represented in the sample, there were challenges encountered as well as future opportunities in this space.

6.1 Challenges

Several robustness checks were employed throughout the tests performed to assess to goodness of fit and model quality. This rigor precluded the use of any form of regression in assessing the utility of the output from NLP techniques in predictive or informative models. For example, despite using two-stage-least-squares regression (by way of instrumental variables) with fixed effects, a robust model of the NLP results (sentiment scores and/or topic prevalence) could not be developed to predict or inform on the level of systemic risk in the U.S. financial system. Even efforts at using neural networks and machine learning were not fruitful, though this type of non-linear / parametric analysis may hold the most promise for future endeavors.

6.2 Opportunities for Future Research

Among the opportunities for future research, the sample of banks could be expanded to include American Depository Receipts (ADRs) to capture the impact of foreign banks on the U.S. financial system.

Further event studies could be performed on EAs' dates with analysis of the text of the banks' press releases to complement existing studies of EA releases, such as Jordan et al. (2000). Additional event studies could be performed to determine if CAAR patterns hold over different time periods, such as pre- and post-Financial Crisis. We know that the rate at which the market absorbs information has increased exponentially in recent years so that intraday event studies may provide insights previously unavailable to researchers.

Further exploration of the predictive power of sentiment and banks with regard to EAs would assist investors and regulators, i.e., Could the content and sentiment used by a bank describe behaviors that predict the initiation of an EA?

One choice that may be considered a limitation of this research was the decision to focus on 10-K reports, which are only published annually. These were selected as a starting point for developing, testing and verifying methods. There is an opportunity to apply NLP techniques to (quarterly) 10-Q disclosures and potentially to other more frequent firm and non-firm, public and non-public disclosures to assess real-time content released by banks and other market participants, such as press releases and social media feeds for real-time indicators of risk to the U.S. financial system.

As shown by these examples, developments in NLP could facilitate further expansion of the content analyzed to enable broader and deeper insights into financial institutions.

CHAPTER 7

CONCLUSION

Financial institutions play a critical role in maintaining the stability of U.S. financial market, both as financial intermediaries and large publicly traded firms, yet the research to apply NLP techniques to their words (via 10-K reports) has been largely limited to word counts, with some exploration of topic modeling to identify emerging systemic risks (notably Hanley and Hoberg, 2018).

This research has: advanced the application of NLP including unsupervised machine learning to distill valuable information from banks' 10-K reports; identified and measured differences in what topics banks discuss and the sentiment they invoke after receiving enforcement actions to provide insights to their risk profiles; applied the results of NLP to portfolio construction to create a model with potential predictive power for systemic risk and market crashes; and applied results of NLP to discern informative patterns in cumulative average abnormal returns post-10-K disclosures.

Overall, this research has provided unique insights to various U.S financial market stakeholders, including investors, analysts and regulators, who can now build on and further these explorations.

APPENDIX A

DISCLAIMER AND RESEARCH NOTE

Disclaimer: The author, Sandra J.H. Rolnicki, is an employee of the Federal Reserve Bank of Chicago. The opinions expressed in this dissertation are her own, and are not formal opinions of, nor binding on, the Federal Reserve Bank of Chicago or the Board of Governors of the Federal Reserve System.

Research Note: All data was obtained from public sources or subscription-based services purchased by Illinois Institute of Technology's Stuart School of Business. This work does not contain confidential supervisory information in detail or in aggregate.

APPENDIX B

CREATING A MATCHED DATASET OF FINANCIAL FIRMS

The following steps resulted in over 1,100 matched financial firms.

1) Obtain the *Banking Research Dataset* by the FRB-NY, 2017. https://www.newyorkfed.org/research/banking_research/datasets.html. For each financial institution, this file lists the matched PERMCO (CRSP identifier) and RSSD ID (Regulatory Reporting Bank identifier) numbers. The current edition of this file contains 1,412 records.

2) Obtain the Loughran and McDonald 10-K Summaries: http://sraf.nd.edu/textual-analysis/resources/#LM_10X_Summaries (the McDonald file) from the University of Notre Dame Software Repository for Accounting and Finance (SRAF). The McDonald file used here covers the (reporting) years 1993 up to and including 2016.

3) Create a text file of the unique SEC Central Index Keys (CIKs) that appear in the file.

4) Go to the Wharton Research Data Services site, https://wrds-web.wharton.upenn.edu/wrds/, (subscription required). Navigate to CRSP → CRSP/Compustat Merged → Bank Annual. Bank Annual was chosen to limit the output to financial firms.

5) On that web page in Step 1, select the dates for the desired data. The McDonald file used here covers the (reporting) years 1993 up to and including 2016, so these years were selected. In Step 2, browse to the text file of CIKs to use as inputs. In Step 3, be sure to select All Link Types. In Step 4, on the Identifying Information tab, select All. Select other variables and run the query with your desired output parameters, e.g., Stata (.dta) file with zip compression.

6) Open the WRDS-generated file and extract the columns with the following identifiers: GVKEY (Standard & Poor identifier), LPERMNO (CRSP identifier), LPERMCO (CRSP identifier), LINKDT (Date of Link), LINKENDDT (Link End Date), TIC (stock ticker symbol), CUSIP (security identifier for debt and capital markets), CIK, CONM (Firm name). Remove duplicate rows. At this step, there were over 1,400 unique records that had data in all of these identifier fields.

7) Use the LPERMCO field in the WRDS-extract to join that list with the PERMCO field in the FRB-NY *Banking Research Dataset*. Perform a reasonableness check with the Link Dates (including Link End Dates) to confirm an appropriate match.

BIBLIOGRAPHY

Abraham, S., & Cox, P. (2007). Analysing the determinants of narrative risk information in UK FTSE 100 annual reports. *The British Accounting Review, 39*, 227–248.

Abrahamson, E., & Park, C. (1994). Concealment of negative organizational outcomes: An agency theory perspective. *Academy of management journal*, *37*(5), 1302-1334.

Airoldi, E.M., & Bischof, J.M. (2016). “Improving and Evaluating Topic Models and Other Models of Text.” *Journal of the American Statistical Association*, 111(516), 1381–1403.

Akerlof, G. (1970). The market for lemons. *Quarterly Journal of Economics*, *84*(3), 488-500.

Balvers, R.J., Gaski, J.F., & McDonald, B. (2015). Financial disclosure and customer satisfaction: Do companies talking the talk actually walk the walk? *Journal of Business Ethics*, 1-17.

Barakat, A., & Hussainey, K. (2013). Bank governance, regulation, supervision, and risk reporting: Evidence from operational risk disclosures in European banks. *International Review of Financial Analysis*, *30*, 254-273.

Basel Committee (2013). Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement. *Bank for international settlements, Basel.*

Beaver, W.H. (1968). The information content of annual earnings announcements. *Journal of accounting research*, 67-92.

Beretta, S., & Bozzolan, S. (2004). A framework for the analysis of firm risk communication. *The International Journal of Accounting, 39*(3), 265–288.

Berger, P.G. (2011). Challenges and opportunities in disclosure research—A discussion of ‘the financial reporting environment: Review of the recent literature’. *Journal of Accounting and Economics*, *51*(1), 204-218.

Bettman, J.R., & Weitz, B.A. (1983). Attributions in the board room: Causal reasoning in corporate annual reports. *Administrative Science Quarterly*, 165-183.

Bischof J., Airoldi E. (2012). “Summarizing topical content with word frequency and exclusivity.” In J. Langford, J. Pineau (eds.), *Proceedings of the 29th International Conference on Machine Learning (ICML-12),* ICML ’12, pp. 201–208. Omnipress, New York, NY, USA. ISBN 978-1-4503-1285-1.

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-K text to gauge financial constraints. *Journal of Financial and Quantitative Analysis, 50*(4), 623-646.

Bollen, K.A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*(1), 109-121.

Brockman, P., & Cicon, J. (2013). The information content of management earnings forecasts: an analysis of hard versus soft information. *Journal of Financial Research*, *36*(2), 147-174.

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.

Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H.M., & Steele, L.B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, *19*(1), 396-455.

Cowan, A.R. (2010). Eventus software, version 9.0. *Cowan Research LC, Ames, Iowa*.

Craig, B., & von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23, 322-347.

Dobler, M (2008). Incentives for risk reporting—A discretionary disclosure and cheap talk approach. *The International Journal of Accounting*, *43*(2), 184-206.

Dobler, M., Lajili, K., & Zéghal, D. (2011). Attributes of corporate risk disclosure: An international investigation in the manufacturing sector. *Journal of International Accounting Research*, *10*(2), 1-22.

Easley, D., & O'Hara, M. (2004). Information and the cost of capital. *The Journal of Finance*, *59*(4), 1553-1583.

Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis*, *52*(2), 811-836.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3-56.

Gandhi, P., Loughran, T. & McDonald, B. (March 28, 2018). *Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks.* Available at SSRN: https://ssrn.com/abstract=2905225 or http://dx.doi.org/10.2139/ssrn.2905225 Retrieved from Social Science Research Network (SSRN) database.

Gompers, P.A., Ishii, J.L., & Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics, 118*(1), 107-156.

Griffin, P.A. (2003). Got information? Investor response to Form 10-K and Form 10-Q EDGAR filings. *Review of Accounting Studies*, *8*(4), 433-460.

Groth, S.S., & Muntermann, J. (2011). An intraday market risk management approach based on textual analysis. *Decision Support Systems*, *50*(4), 680-691.

Hanley, K.W. & Hoberg, G. (February 28, 2018). *Dynamic Interpretation of Emerging Risks in the Financial Sector*. Available at SSRN: https://ssrn.com/abstract=2792943 or http://dx.doi.org/10.2139/ssrn.2792943. Retrieved from Social Science Research Network (SSRN) database.

Hargie, O., Stapleton, K., & Tourish, D. (2010). Interpretations of CEO public apologies for the banking crisis: attributions of blame and avoidance of responsibility. *Organization*, *17*(6), 721-742.

Helbok, G., &Wagner, C. (2006). Determinants of operational risk reporting in the banking industry. *Journal of Risk, 9*(1), 49-74.

Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199-236.

Ho, D.E., Imai, K., King, G., Stuart, E.A., & Whitworth, A. (2018). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. R package version 3.0.2, URL: https://CRAN.R-project.org/package=MatchIt.

Hogenboom, A., Frasincar, F,. de Jong, F., & Kaymak, U. (2015). Using rhetorical structure in sentiment analysis. *Communications of the ACM*, *58*(7), 69-77.

Huang, A.H., Zang, A.Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, *89*(6), 2151-2180.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305-360.

Jockers, M. (2017). syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text. R package version 1.0.4, URL: https://CRAN.R-project.org/package=syuzhet.

Jordan, J.S., Peek, J., & Rosengren, E.S. (2000). The market reaction to the disclosure of supervisory actions: Implications for bank transparency. *Journal of Financial Intermediation*, *9*(3), 298-319.

Kothari, S.P., Li, X., & Short, J.E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, *84*(5), 1639-1670.

Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, *18*(4), 1088-1122.

Lang, M.H., & Stice-Lawrence, L. (2014). *Textual Analysis and International Financial Reporting: Large Sample Evidence*. Paper presented at Journal of Accounting & Economics Conference, Philadelphia PA.

Lajili, K., & Zéghal, D. (2005). A content analysis of risk management disclosures in Canadian annual reports. *Canadian Journal of Administrative Sciences*, *22*(2), 125.

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

Li, J., Wei, L., Lee, C.F., Zhu, X., & Wu, D. (2017). Financial statements based bank risk aggregation. *Review of Quantitative Finance and Accounting*, 1-22.

Linsley, P.M., & Shrives, P.J. (2006). Risk reporting: A study of risk disclosures in the annual reports of UK companies. *The British Accounting Review, 38*, 387–404.

Linsley, P.M., Shrives, P.J., & Crumpton, M. (2006). Risk disclosure: An exploratory study of UK and Canadian banks. *Journal of Banking Regulation*, *7*(3), 268-282.

Loughran, T., McDonald, B., & Yun, H. (2009). A wolf in sheep's clothing: The use of ethics-related terms in 10-K reports. *Journal of Business Ethics*, *89*(1), 39-49.

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. (2014a). Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643-1671.

Loughran, T., & McDonald, B. (2014b). Regulation and financial disclosure: The impact of plain English. *Journal of Regulatory Economics*, *45*(1), 94-113.

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

McClelland, P.L., Liang, X., & Barker, V.L. (2010). CEO commitment to the status quo: Replication and extension using content analysis. *Journal of Management*, *36*(5), 1251-1277.

Michalisin, M.D. (2001). Validity of annual report assertions about innovativeness: an empirical investigation. *Journal of Business Research*, *53*(3), 151-161.

Miihkinen, A. (2012). What drives quality of firm risk disclosure? The impact of a national disclosure standard and reporting incentives under IFRS. *The International Journal of Accounting*, *47*(4), 437-468.

Mimno D., Wallach H.M., Talley E., Leenders M., & McCallum A. (2011). Optimizing Semantic Coherence in Topic Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, pp. 262–272. Association for Computational Linguistics, Stroudsburg, PA, USA. ISBN 978-1-937284-11-4.

Mohammad, S.M., Turney, P.D., (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34). Association for Computational Linguistics.

National Information Center. (December 31, 2017). *Holding Companies with Assets Greater Than $10 Billion*. Retrieved from https://www.ffiec.gov/nicpubweb/nicweb/HCSGreaterThan10B.aspx

Nelson, K.K., & Pritchard, A.C. (August 2007). Litigation risk and voluntary disclosure: The use of meaningful cautionary language. In *2nd Annual Conference on Empirical Legal Studies Paper*. Available at SSRN: https://ssrn.com/abstract=998590 or http://dx.doi.org/10.2139/ssrn.998590 Retrieved from Social Science Research Network (SSRN) database.

Nguyen, D. D., Hagendorff, J., & Eshraghi, A. (2015). Can bank boards prevent misconduct? *Review of Finance*, 20(1), 1-36.

Perry, J., & de Fontnouvelle, P. (2005). Measuring reputational risk: The market reaction to operational loss announcements. *Federal Reserve Bank of Boston.*

Pollach, I. (2012). Taming textual data: the contribution of corpus linguistics to computer-aided text analysis. *Organizational Research Methods*, *15*(2), 263-287.

Roberts M.E., Stewart B.M., Tingley D., Airoldi E.M. (2013). "The Structural Topic Model and Applied Social Science." *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., & Rand, D.G. (2014). "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science*, 58(4), 1064–1082.

Roberts, M.E., Stewart., B.M., Tingley, D., & Benoit, K. (2018). stm: Estimation of the Structural Topic Model. R package version 1.3.3, URL: https://CRAN.R-project.org/package=stm.

Ross, S. (1973). The Economic Theory of Agency: The Principal's Problem. *The American Economic Review, 63*(2), 134-139.

Rothschild, M., & Stiglitz, J. (1978). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In *Uncertainty in economics* (pp. 257-280).

Roulstone, D.T. (1999). Effect of SEC financial reporting release No.48 on derivative and market risk disclosure. *Accounting Horizons, 13*(4), 343–363.

Rutherford, B.A. (2003). Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of management and governance*, *7*(2), 187-210.

Schumaker, R.P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, *45*(5), 571-583.

Securities and Exchange Commission. (2010). *Form 10-K instructions*. Retrieved from www.sec.gov/about/forms/form10-k.pdf

Securities and Exchange Commission. (2012). *Filings and Forms*. Retrieved from http://www.sec.gov/edgar.shtml

Sharpe, W.F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425-442.

Solomon, J.F., Solomon, A., & Norton, S.D. (2000). A conceptual framework for corporate risk disclosure emerging from the agenda for corporate governance reform. *The British Accounting Review, 32*, 447–478.

Tetlock, P.C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437-1467.

Wharton Research Data Services. “Eventus Basic Event Study,” wrds.wharton.upenn.edu, accessed 01/26/2018.

You, H., & Zhang, X.J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies*, *14*(4), 559-586.