Summer 2016

The Effect of XBRL and Social Media on Information Asymmetry: Evidence from Bank Loan Contracts

Dazhi Chong
Old Dominion University

THE EFFECT OF XBRL AND SOCIAL MEDIA ON INFORMATION ASYMMETRY: EVIDENCE FROM BANK LOAN CONTRACTS

by

Dazhi Chong
B.S. July 1998, Anhui Institute of Finance and Trade, China
M.S. June 2006, Hefei University of Technology, China

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

BUSINESS ADMINISTRATION- INFORMATION TECHNOLOGY

OLD DOMINION UNIVERSITY
August 2016

Approved by:

Harris Wu (Director)

Ling Li (Member)

Jong Park (Member)

ABSTRACT

THE EFFECT OF XBRL AND SOCIAL MEDIA ON INFORMATION ASYMMETRY: EVIDENCE FROM BANK LOAN CONTRACTS

Dazhi Chong
Old Dominion University, 2016
Director: Dr. Harris Wu

This study analyzes how two information technology advancements, the adoption of XBRL (eXtensible Business Reporting Language), and social media, affect bank loan contracting using a sample of 554 US bank loan contracts in 2011. I hypothesize that the adoption of XBRL and social media can enhance information dissemination and mitigate the information asymmetry problem between borrowers and lenders. Consistent with this hypothesize, I find that borrowers that adopt XBRL and/or receive positive social media user opinion in social media enjoy more favorable price and non-price terms of bank loan contracts. Additional analyses indicate that the relations among XBRL adoption, social media user opinion and bank loan price vary with the firm size, loan structure and availability of public information of borrowers. Overall, this research provides evidence that technology advancements, the adoption of XRBL and social media, reduce cost of bank loans by decreasing information asymmetry between borrowers and lenders.

This dissertation is dedicated to my parents and my wife for their unconditional love, support, and encouragement.

ACKNOWLEDGMENTS

It is a great pleasure to thank those who made this dissertation possible. I would never have been able to finish my dissertation without the guidance of my committee members and support from my family.

I would love to thank my PhD advisor, Dr. Harris Wu, for his excellent guidance, advice, and patience. Thank you so much for believing in my abilities and guiding me through the challenges of this dissertation. I would also like to thank my committee members, Dr. Ling Li and Dr. Jong Park, for their valuable expertise and their commitment to the improvement of this dissertation.

I would also love to thank my parents and my wife, for their unconditional and endless love and encouragement in all my efforts. Thank you for your love, patience, and support during my PhD study.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 BACKGROUND

Recent development in information technology, especially XBRL (eXtensible Business Reporting Language) and social media, has changed the way that firms communicate with stakeholders. As a standard for exchanging business information, XBRL enables the tagging of information using predefined taxonomy. This feature makes it easier for analysts, investors, and other stakeholders to access, analyze, and process financial information. Realizing the advantage of XBRL, on January 30, 2009, SEC (Securities Exchange Commission) adopted rules that require companies to submit their financial statements in XBRL format. After June 15, 2011, the proposed rules apply to domestic and foreign companies using U.S. GAAP and to foreign private issuers using International Financial Reporting(Securities Exchange Commission, 2009). Another new technology, social media, also shows its potential to facilitate information dissemination. Social media allows firms to push information to investors or customers simultaneously via postings, direct message et.al. This characteristic reduces the time and cost that investors or customers spend in sorting through various news sources(Blankespoor, Miller, & White, 2013). More importantly, social media communications are bi-directional which enable direct and immediate interactions among users(He, Zha, & Li, 2013). While some researchers have found that XBRL and social media can bypass the shortcomings of traditional information dissemination tools and reduce information inequities in the capital markets (Bollen, Mao, & Zeng, 2011; Farewell, 2006; Maina, 2015), few studies have been conducted on the relationship between these new technologies and loan contracts. This dissertation extends the literature by examining how the adoption of XBRL, and social media sentiment, affect bank loan contracts.

1.2 RESEARCH QUESTION

The importance of bank loans attracts lots of researchers to investigate the determinants of bank loan contracting(Bae & Goyal, 2009; Benmelech, Garmaise, & Moskowitz, 2004; Bharath, Sunder, & Sunder, 2008; Dennis, Nandy, & Sharpe, 2000; Diamond, 1991b; Hasan, Park, & Wu, 2012; Haselmann, Pistor, & Vig, 2010; Klock, Mansi, & Maxwell, 2005; Lin, Chen, & Yen, 2014; Qian & Strahan, 2007; Strahan, 1999; Sufi, 2007; Wu, Francis, Hasan, & Koetter, 2011). Some firm characteristics and loan characteristics such as default risk, liquidation value, and corporate board structure are found to be associated with loan contracts. However, there is a lack of current research on how the adoption of XBRL affects bank loan contracts. As XBRL has the potential to reduce information asymmetry in the capital markets, there is a possible link between XBRL adoption and loan contracts. Therefore, the first purpose of this study is to investigate the effect of XBRL on information asymmetry in capital markets, in particular, to evaluate the impact of XBRL adoption on the price and non-price terms of bank loan contracts. Hence, the first two research questions that need to be answered are:

1. Can the adoption of XBRL affect price and non-price terms of bank loan contracts?

2. How does the adoption of XBRL affect price and non-price terms of bank loan contracts?

Many studies have found that social media sentiment is an important indicator of firm's potential risks and values. For instance, Hu, Liu, and Zhang (2008) find that online product reviews are major information source for consumers to make buying decision. There is a positive relationship between favorable peer reviews and product sales. By analyzing the peer opinions on social media, Chen, De, Hu, and Hwang (2014) find that social media sentiment has predictive power over future stock returns and earnings surprises. Although prior studies suggest that social media sentiment is associated with firm's stock returns and sales, there is a lack of empirical evidence of the relationship between social media sentiment and loan contracts. Therefore, the second purpose of this study is to investigate how social media sentiment affects loan contracts. This objective leads to the other two research questions of this study:

1. Can social media sentiment affect price and non-price terms of bank loan contracts?
2. How does social media sentiment affect price and non-price terms of bank loan contracts?

1.3 THE CONTRIBUTIONS OF THE DISSERTATION

This dissertation contributes to the literature in the following aspects. First, this dissertation is the first to examine the impact of XBRL adoption and social media sentiment on the cost of bank loan. Using a sample of 554 US bank loan contracts in 2011, this study finds that borrowers that adopt XBRL and/or receive positive social media user opinion enjoy more favorable price and non-price terms of bank loan contracts. Second, this dissertation reveals that the effect of XBRL adoption and social media sentiment on bank loan price is not homogenous. The analyses show that these effects are stronger for small borrowers and syndicated loans. Furthermore, the analyses also find that borrowers that adopt XBRL are more likely to be offered with unsecured loan.

In sum, this study confirms the hypotheses that both the adoption of XBRL and social media sentiment are important indicators of borrower's risks and values. The study not only answers the question "Can XBRL and social media affect price and non-price terms of bank loan?", but also provides stakeholders with a better guide for how to use XBRL and social media to decrease information asymmetry in the capital markets.

1.4 OUTLINE OF THE DISSERTATION

The remainder of the dissertation is structured as follows: chapter 2 provides a brief review of the literature concerned with the determinants of bank loan contracts and develops the hypotheses of this study. This chapter contains four sections: Section 2.1 introduces prior findings related to the determinants of bank loan contracts. Section 2.2 discusses the framework and benefits of XBRL. Section 2.3 introduces the framework and benefits of social media. The final section, section 2.4 discusses knowledge gap in the literature and develops the hypotheses of this study.

The purpose of chapter 3 is to present the research methodology and the research approaches adopted by this study. This chapter contains five sections. First, section 3.1 describes the architecture of the Bank Loan Pricing Analytics System (BLPAS). Three key modules of BLPAS: data collection module, data integration module and data analysis module are introduced in this section. The second section, section 3.2 presents the data sources of this study. After introducing sampling procedure in section 3.3, section 3.4 describes the measurement of dependent variables, independent variables, and control variables in this study. Finally, the empirical models for the study are discussed in section 3.5.

By presenting results of the regression analyses, chapter 4 provides evidences to indicate how XBRL and social media sentiment affect bank loan price. This chapter contains five sections: Section 4.1 presents the descriptive statistics and correlation analysis of dependent and independent variables. Section 4.2 and 4.3 presents the results of the regression analysis related to XBRL, social media sentiment and loan price. After discussing the results of robust test in section 4.4, section 4.5 introduces the results of analysis on firm size. Section 4.6 analyzes the results of analysis on lending relationship. Section 4.7 reports the results of analysis on syndication. Finally, the results of analysis on Non-Price terms are discussed in section4.8.

The last chapter, chapter 5 presents the research conclusion of this research. This chapter contains three sections. Section 5.1 discusses implications for the research. Section 5.3 introduces future research opportunities. And section 5.4 includes a summary with conclusions.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1 DETERMINANTS OF BANK LOAN PRICING

In the bank loan literature, the determinants of bank loan contracting have attracted tremendous research interests. For instance, Zhang (2008) examines the benefits of accounting conservatism in the loan contracting process. He finds that accounting conservatism results in reduced default risk, which enables lenders to offer lower initial interest rates to borrowers. Bharath et al. (2008) focus on the relationship between borrower accounting quality and loan prices. They suggest that lenders have to raise interest rates and tighten non-price debt terms to compensate for the default risk arising from low accounting quality. Hasan et al. (2012) examine the impact of earnings predictability on both price and non-price terms. They find that lower predictable earnings are associated with higher interest rates, shorter maturities, and more covenants and collateral requirements. Benmelech et al. (2004) argue that higher liquidation value can not only lower the cost of liquidation but also increase the asset's durability and make longer maturity debt feasible. Thus, borrowers with high liquidation value are more likely to receive favorite loans prices and no-price terms. Sufi (2007) finds that high borrower reputation is associated with low default risk, which enables banks to offer favorite loan terms to borrowers. Some studies find that corporate board structure also affects bank loan price. As an important corporate governance mechanisms to control loan risk, independent board makes firms easier to control CEO overcompensation(Core, Holthausen, & Larcker, 1999) , limit over-investments(Richardson, 2006) and reduce earnings management and financial fraud(Agrawal & Chadha, 2005; Klein, 2002). From this point, the less independent a borrower's board is, the more likely it will experience financial distress and be offered high loan price(Daily & Dalton, 1994; Lee & Yeh, 2004).

In addition to firm characteristics, some external factors are also associated with debt contracting. Qian and Strahan (2007) study the impact factor of bank loan price from law and finance view. They argue that loan contracts are the reflection of differences in legal protection of lenders and the enforcement of contracts. When lenders' rights are well protected, the law enforces lenders' ability to the take collateral in the case of default. In this situation, borrowers are more likely to get favorable loan terms. Consistent with Qian's findings, Bae and Goyal (2009) find that firms in a region with strong creditor rights and high enforceability of loan contracts are more likely to receive lower loan spreads, longer maturities.

Besides the impact factors mentioned above, theoretical work suggests that information asymmetry is another key determinant of loan contracting. Information asymmetry is a situation where at least one party in a contract relationship has more information than others. The occurrence of asymmetric information creates an unbalanced transaction and result in "moral hazard" and "adverse selection" problems in the loan contracting(Kim, 1985). When the levels of information asymmetry are different, loan terms offered by lenders will also be different (Amihud & Mendelson, 1986; Diamond & Verrecchia, 1991; Easley & O'hara, 2004). According to the literature, information asymmetry is mainly caused by specific firm characteristics and loan characteristics. For instance, Barclay and Smith (1995) examine the relationship between firm size and information asymmetry. Their findings suggest that large firms are more efficient in reducing information asymmetries. As a result, large firms are more likely to get long-term debt. Bhojraj and Sengupta (2003) argue that the agency risk and information risk between management and outside stakeholders will affect bank loan price. In general, these research imply that when information asymmetry is high between borrowers and banks, banks are more likely to raise interest rates and tighten non-price debt terms to compensate for default risk and potential losses(Wu et al., 2011).

In order to reduce the information asymmetry between insiders and outside investors or stakeholders, firms often use disclosure to disseminate private information to the markets. As a popular tool, disclosure has been shown to influence a number of operations related factors such as firm's cost of capital (Botosan1997), analyst following (Lang and Lundholm 1996), institutional investor following (Bushee and Noe 2000), and stock price volume and volatility (Healy, Hutton, and Palepu 1999; Bushee and Noe 2000). The main drawbacks of disclosure are that it can only reach limited set of stakeholders and are often biased(Blankespoor et al., 2013; Miller & Skinner, 2015). For instance, business press is an important disclosure approach. Many investors rely on business press to acquire value relevant information. Bushee, Core, Guay, and Hamm (2010) find that greater press coverage around earnings announcements reduces information asymmetry in the form of spread reductions and depth improvements. Soltes (2010) finds that disclosure dissemination through the press reduces spreads, increases trading volume, and lowers idiosyncratic volatility. Kothari, Li, and Short (2009) find that positive (negative) press coverage decreases (increases) firms' cost of capital, return volatility, and analyst forecast dispersion. However, some studies find that only important corporate news releases are sufficiently monitored by traders, increased distribution of the information by the press only results in little increase in dissemination(Bushee et al., 2010). Further, much of the information disseminated to the market by the press doesn't contain editorial content, and only provides limited information (Soltes, 2010).

Beside business press, another main disclosure approach, conference call also has some shortcomings. Mayew (2008) argues that managers are more likely to call on analysts who have more favorable recommendations. Bushee, Matsumoto, and Miller (2003) find that firms with relatively more shareholders and relatively fewer institutional holders are more likely to open their conference calls. Their finding suggests that the scope of information dissemination is determined by the nature of the firm's investor base.

2.2 EXTENSIBLE BUSINESS REPORTING LANGUAGE (XBRL)

While the shortcomings of traditional disclosure approaches often limit the depth and breadth of information dissemination, recent changes in technology make it possible for investors, analysts, and banks to bypass the weaknesses of traditional tools. Among these technologies, XBRL is one of the most popular tools used in today's capital markets.

2.2.1 INTRODUCTION

XBRL(eXtensible Business Reporting Language) is a XML based reporting language for exchanging business information(Zhu & Wu, 2014). XBRL enables the tagging of information using predefined taxonomy. Therefore, XBRL tags can be easily accessed and interpreted by computer applications. Currently, XBRL has been widely used by banking regulators, stock exchanges regulators, investors, analysts, and statistical agencies.

XBRL is composed of three key components: specification, taxonomy, and instance documents. As a guideline of taxonomies and instance documents design, specification defines how to build XBRL instance documents and XBRL taxonomies. The standardization of syntax of instance documents, syntax of taxonomies, semantics of instance documents, and semantics of taxonomies enable different users to create, exchange financial data among various organizations(Wu & Vasarhelyi, 2004). Table 1 shows an example of XBRL specification.

[Insert Table 1 here]

The second component, taxonomy, contains the definitions of reporting data elements, reference to the shared meaning of elements, and business rules for validation and interrelation between data elements. The aim of taxonomy is to classify and standardize the accounting information generated by different types of accounting standards. Similar to the data dictionary, data elements and their interrelationships are well defined in taxonomy. As an extensible framework, XBRL taxonomy allows users to add data elements, redefine relationships and references. This feature enables companies to create, publish and transfer financial information without losing the integrity of the data(Aad & Paul, 2008). In XBRL, taxonomy is comprised of two major components: XML Schemas and linkbases. Taxonomy schemas define the names and types of concepts. Table 2 shows the examples of taxonomy schemas of US GAAP 2016.

[Insert Table 2 here]

Another component, linkbases, defines the relationships between data elements and resources related to them. The XBRL 2.1 specification defines five types of linkbases: Definition Linkbase, Calculation Linkbase, Presentation Linkbase, Reference Linkbase and Label Linkbase. The first three linkbases describe inter-concept relationships, and the last two linkbases define the relationships between concept and documentation(Engel et al., 2008).

[Insert Table 3 here]

Table 3 shows how the IncomeTaxesPaidRefund is calculated using two elements: IncomeTaxesPaidRefundClassifiedAsInvestingActivities and IncomeTaxesPaidRefundClassifiedAsOperatingActivities. While the Calculation Linkbase has described the relations and weights of these elements, it is easy for the users to calculate the value of IncomeTaxesPaidRefund using the formula below:

IncomeTaxesPaidRefund = IncomeTaxesPaidRefundClassifiedAsInvestingActivities + IncomeTaxesPaidRefundClassifiedAsOperatingActivities

The third component of XBRL, instance documents, is composed of instances of specific elements and tags defined in the taxonomies. Instance documents can be used to store and publish different kind of business reporting information including business facts, units, contexts, and footnotes(Nutz & Strauß, 2002). Table 4 shows an example of XBRL instance documents.

[Insert Table 4 here]

2.2.2 BENEFITS OF XBRL

Prior studies find that XBRL can facilitate the information sharing among various stakeholders (Debreceny et al., 2005; Khan, 2006; Pinsker & Li, 2008). For instance, traditional financial reports are often generated in different formats and various accounting standards. Only end users are familiar with each accounting standards and data formats. The exchange and analysis of the information in financial reports are extremely complicated. With the adoption of XBRL, accounting principles and financial reports can be mapped into standardized data elements, relationships, and references. As accounting information is more reliable and relevant in XBRL format, firms can easily adapt to various accounting principles and financial reports using corresponding XBRL rules(Cuneo, 2003). The comparison of accounting information is much easier than before(Vasarhelyi, Chan, & Krahel, 2010; Zhu & Wu, 2010).

XBRL can increase the efficiencies of the data integration and processing. The introduction of XBRL tags allows computers to process information independently, thus reducing the cost associated with data integration and processing(Altova, 2016; Rezaee & Turner, 2002). The increased efficiencies in the financial data processing also make it possible for auditors, regulators and banks to monitor firm's operations continuously(Rezaee & Turner, 2002), which can significantly reduce the information asymmetry between insiders and outsiders.

2.2.3 PRELIMINARY STUDY OF XBRL: FIRM CLUSTERING

2.2.3.1 INTRODUCTION

There are several ways to group firms. The most common method is to group them into different industries according to firm's core business. In the U.S., the Standard Industry Classification (SIC) codes, and more recently the North American Industry Classification System (NAICS) codes, have been widely used to classify a firm into a certain industry. These classification codes provide an efficient way for investors and policy-makers to analyze economic data and reveal the structure of economy(US Census Bureau, 2012). However, the code assignments are somewhat static, not capturing the evolvement of firm's business and industry structure. For example, as firms expand or shift their business fields, the codes assigned to firms often do not accurately reflect the "natural" grouping of firms. Thus, more "dynamic" and "efficient" classification approaches are required in today's highly competitive market.

When financial statements were not digitized or were in unstructured format, it was difficult to derive useful information from them. This has been fundamentally changed since the recent adoption of XBRL (XBRL International, 2006). In the U.S., the Securities and Exchange Commission has adopted the GAAP taxonomy as a data standard. Specified in XBRL, this taxonomy defines a set of financial reporting concepts and their relationships. Earlier studies(Zhu & Wu, 2011a; Zhu & Wu, 2011b) show that 87% of the reported data are defined in the GAAP taxonomy. The wide use of GAAP taxonomy and XBRL makes it possible for me to mine a large quality of financial statements to identify firm clusters.

In this study, I model firms and the GAAP Taxonomy elements used by the firms as a bipartite "social network". I implement a spectral clustering method and apply it to the "social network". The results demonstrate the feasibility of using financial data to identify firm clusters.

2.2.3.2 RESEARCH METHOD

In order to group firms based on tags used in their financial statements, I develop a Firm Clustering System to collect data, analyze the structure of the financial statements, and create clusters. Figure 1 shows the framework of this system.

[Insert Figure 1 here]

Data Collection

I collect the 10-K annual financial statements from 2009 to 2011. If a firm has more than one 10-K, I choose to use the earliest one. Since there is no reliable method to match custom data elements, my analysis focuses on the elements of GAAP Taxonomy. In the rest of the paper, I use the term tag and GAAP element interchangeably. The dataset has 10-K's of 1799 firms, which together use 7021 GAAP elements.

In the financial statement, some commonly used tags such as "Assets", "CostofRevenue" are used by most companies. The strong relationship created by these commonly used tags will make most companies belong to the same cluster. Since my main objective is to cluster companies based on their usage of specific tags, it is necessary to ignore these commonly used tags. Therefore, I delete tags that are used by more than 50 companies. After removing these tags, 5815 elements are used by this research in the final dataset.

Clustering Approach

Clustering is an efficient way to explore information within certain networks or groups. Many clustering algorithms are based on the assumption that the data within the dataset that has specific attributes or links (Zha, He, Ding, Simon, & Gu, 2001). In this study, one company uses a number of tags specified in the GAAP Taxonomy. Likewise, each tag is also used by several companies. By treating tags as attributes of companies, the company-tag relationship can be represented as an m-by-n matrix A. In this case, I can identify clusters based on the relationship between firms and tags. Figure 2 shows how the clustering works.

[Insert Figure 2 here]

There are numerous algorithms, such as agglomerative clustering (Voorhees, 1986) and k-means algorithm (Dhillon & Modha, 2001), which can be used to identify clusters. However, most of them can't guarantee global optimization of clusters (Shi & Malik, 2000). Spectral clustering is a solution to address this deficiency. The objective of spectral clustering is to find the partition of a graph so that the linkages between groups are minimized and the linkages within groups are maximized. In graph language, the linkages among groups are called "cuts", which can be computed through the total weight of the edges between connected groups. Shi and Malik (2000) suggest a co-clustering algorithm to minimize cuts globally. They argue that second smallest eigenvalue of Laplacian matrix can be used to find the minimum cut vertex partitions in a graph. Dhillon (2001) extends this algorithm. By using the singular value decomposition (SVD) approach, he finds that the second left and right singular vectors of a normalized matrix provide an optimal solution for co-clustering problem. While Dhillon's approach is more efficient and can cluster tags and companies simultaneously, I choose to use this approach to identify clusters in the dataset. The main procedures of Dhillon's approach are as follows:

1. A graph G=(V, E) is a set of vertices V={v_1, ,..., v_n}, and a set of edges (i, j). E_{ij} is equal to the edge weight between v_i and v_j. W is a symmetrical matrix, where W_{ij} =E_{ij}, if there is an edge between i and j, else W_{ij} =0 ; Let matrix D be an diagonal matrix, where $D_1(i,j)=\sum_j W_{ij}$, $D_2(i,j)=\sum_i W_{ij}$

$$D=\begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$$

2. Form matrix $Wn=D_1^{-1/2}WD_2^{-1/2}$,

3. Compute the second singular vectors of Wn and form the vector z_2, where u_2 and v_2 are the left and right singular vectors of Wn.

$$z2=\begin{bmatrix} D1^{-1/2}u_2 \\ D2^{-1/2}v_2 \end{bmatrix}$$

4. Run the k-means algorithm on z_2 to obtain the desired clusters.

In this study, there are 1799 firms and 5815 tags, so there are 1799 elements in V and 5815 elements in E, an edge <fi,tj> exists when firm fi uses tj tag. For simplicity, I set the weight of each edge as one and the edges are undirected. Table 5 shows an example of firm-tag matrix. For instance, the value of row 2 and column 2 is zero, which indicates that firm1 does not use tag1 in its financial statement.

[Insert Table 5 here]

My clustering algorithm uses k-means algorithm on the singular vectors to obtain the desired clusters, thus I need to determine the best value for k. A number of approaches have been developed to determine k value. In this study, I use approach suggest by Schaeffer (2007). He considers three measurements to evaluate the fitness of the k value. Among these measurements, relative criterion is used to measure the goodness of inter-cluster density. Local criterion focuses on the goodness of a clustering structure without external information. In order to optimize both local density and global density, Schaeffer (2007) uses the product of the local and relative densities to measure the fitness of cluster function. This approach provides an easy way to optimize k value in this study.

2.2.3.3 FINDINGS

Using the approach mentioned above, I cluster the 1799 companies into 20 clusters. Based on Schaffer's approach, I calculate local density, relative density and total density of different cluster numbers. Table 6 shows the distribution of local density, relative density and total density in different numbers of clusters.

[Insert Table 6 here]

In table 6, when k is equal to 20, the total density gets the highest value. Since total density is the tradeoff between local density and relative density, I set the k value as 20. Table 7 shows firm distribution among industry groups according to NAICS.

[Insert Table 7 here]

Next, I try to reveal potential relationship among firms and compare it with standard classification code. Figure 3 shows the distribution of major industry in 20 clusters.

[Insert Figure 3 here]

According to figure 3, cluster 17 only contains one firm, Verde Resources Inc (VRI), an exploration stage company. After removing popular tags, it only has seven tags. In addition to this outlier case, cluster 5 is also "pure", containing only "Finance and Insurance" (NAICS code 52) firms. Clusters 8, 14, and 15 are nearly pure with predominantly Finance and Insurance firms. An examination of the firms in these clusters shows that firms within each cluster have significant similarities in their core businesses. For example, firms in cluster 5 are primarily insurance companies, whereas firms in cluster 8 are primarily commercial banks. Utilities (code 22) firms are dominant in clusters 1 and 7. Similarly, Mining, Quarrying, Oil, Gas Extraction (code 21) firms are dominant in cluster 11.

This case study leads me to hypothesize that certain firm behaviors (such as operation mode, investment strategy) determine the contents and structures of firm financial statements. Conversely, the contests and structures should help us infer firm characteristics. To preliminarily test my hypothesis, I analyze the tags used by firms within each cluster. I rank the tags according to the number of firms that use them. The frequently used tags may indicate major financial behaviors in these clusters. Table 8 lists the top 10 tags in selected clusters (the list does not include the removed elements – those used by more than 50 firms).

[Insert Table 8 here]

The tags in cluster 1 indicate that firms are regulated public utilities, some of which are jointly owned (JointlyOwnedUtilityPlantProportionateOwnershipShare is used by more than 15 firms in cluster 1). The tags in cluster 2 indicate that most firms are partnerships. The tags in cluster 5 indicate that insurance premiums are important to insurance business and the firms are also in the reinsurance business. The tags in cluster 6 indicate that firms sell products with warranty, use elaborate financing and compensation methods, and are engaged in acquisition of other businesses. All these findings are useful but cannot be derived in any way from firms' NAICS codes. These preliminary findings are very promising to support our hypothesis. Since there are hundreds of frequently used tags in 20 clusters, I plan to analyze them all to better understand the clustering results and fully test my hypothesis.

2.2.3.4 CONCLUSION

Prior research attempt to identify firm groups based on the operation process and output of firms. The interrelation and interdependence among firms are not captured. In this study, I introduce a spectral clustering method to cluster firms based on XBRL- based financial statements. Similar to other "social networks", the firm-tag network has natural groupings of firms' financial statements. My work demonstrates the feasibility of clustering firms based on XBRL tags and shows that clusters exhibit interesting common features of the firms within the same cluster.

2.3 SOCIAL MEDIA

2.3.1 INTRODUCTION

According to Kaplan and Haenlein (2010)'s definition, social media is "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content." (pp. 61). This definition suggests that Web2.0 based applications and User Generated Content are the two key featuresof social media. Social media is distinct from traditional media as it emphasizes greater collaboration among users. At the same time, users on social media are more involved in the creation of content and have more control over it(Kaplan & Haenlein, 2009). As a collective of online communications channels, social media is composed of various platforms that can provide different social actions. For instance, Facebook is a popular social media website that allows users to send messages, post photos, and share various apps with friends. In contrast, LinkedIn is a business-oriented social networking site. The goal of this website is to help users establish connections with industry experts and business partners. Currently, social media has become an essential part of people's life. Noticing the potential benefits of social media, more and more businesses begin to use social media to conduct marketing research, promote products, and build long-term relationships with customers.

2.3.2 BENEFITS OF SOCIAL MEDIA

Compared to traditional communication channels, social media platforms allow firms to reach a large amount of audience at lower cost. This feature makes it easier for businesses to increase brand recognition, improve brand loyalty and build stronger relationships with customers(He, Tian, Chen, & Chong, 2016). De Vries, Gensler, and Leeflang (2012) analyze the social media data of eleven international companies. They find that brand loyalty is positively associated with positive posts on social media platform. Trainor, Andzulis, Rapp, and Agnihotri (2014) examine the relationship between social media technology usage and customer relationship. Their findings suggest that social media technology investments have a positive impact on social CRM capabilities as social media can facilitate the interaction between businesses and customers. Laroche, Habibi, Richard, and Sankaranarayanan (2012) also find the positive relationship between social media and brand loyalty. In general, these studies show that social media can enhance self-identity of the community, which in turn, create the value for both customers and companies.

While social media enable businesses to communicate with customers using two-way participatory media model, the speed and easiness in transporting and sharing information on social media allow users to acquire unbiased information much easier than before(Yi, Oh, & Kim, 2013). Blankespoor et al. (2013) use Twitter as the example to examine the relationships between social media information disclosure, firm's information environment and information asymmetry. They find that firm-initiated news disseminated by Twitter have negative influence on abnormal bid-ask spreads. Their finding suggests that dissemination of information via social media can significantly reduce information asymmetry in capital markets. He et al. (2016) analyze and compare the social media content on the Facebook sites of the three largest drugstore chains. They find that customers use social media comments to provide suggestions to these companies. He et al's study suggests that businesses can use the knowledge they gain from social media to develop better business strategies and improve competitive advantages in the market.

2.4 HYPOTHESIS DEVELOPMENT

Accounting information is a useful tool for banks to evaluate borrowers' default risk. The richer accounting information banks have, the more likely the banks can accurately assess the default risk of borrowers(Healy & Palepu, 2001). Among various tools providing accounting information, financial statement has been widely used by analysts, investors and the government in the capital markets(Chong & Zhu, 2012). Traditionally, financial statement was prepared and generated in unstructured format. Many resources were wasted on locating, converting, and understanding accounting information in financial statement. The application of XBRL makes it possible for users to retrieve accounting information in a more efficient way. Compared to traditional approaches, XBRL can define taxonomies to specify a set of concepts. When all users are using the same taxonomies, data in XBRL format can be easily shared across various applications and platforms(Farewell, 2006). A research conducted by Pinsker and Li (2008) indicates that XBRL benefits include cost savings, increased data processing capability, increased efficiency, decreased data redundancy, and decreased cost of bookkeeping. Overall, XBRL based disclosures have the potential to reduce information risk and information asymmetry in the capital markets(Kim, Kim, & No, 2011). While information risk and information asymmetry are closely associated with bank loan price, I expect a negative relation between the adoption of XBRL and cost of bank loan. Based on this argument, I hypothesize:

H1a: Borrowers that adopt XBRL enjoy more favorable loan prices

The main shortcomings of traditional media are that: 1) it can't disseminate the information to a broad set of stakeholders; 2) information disseminated to the markets are often biased (Bushee et al., 2010; Miller, 2006; Soltes, 2010). For instance, Heinle and Verrecchia (2015) suggest that the extent to which firms bias their disclosure depends on the content of the information disclosed by other firms or if there is a potential value to firms' stock prices. Frankel and Li (2004) suggest that large firms are concerned by more followers. Analysts' recommendations on large firms are often less biased than on small firms. Coombs (2007) finds that negative information created by hostile competitors often harm firm's image and value. Compared to traditional media, social media utilizes a two way broadcast model to disseminate information. This feature enable firms to efficiently reach a large number of stakeholders (Yi et al., 2013). Basically, the data available on social media can be classified into two categories: firm-initiated information and user generated information (Bushee et al., 2010; Lee, Hutton, & Shu, 2015). Firm-initiated information is created by firms and reflects firms' own opinion. On the other hand, postings or comments created by general users belong to user-generated information that represents users' opinions and views. Considering the huge amount of users on social media, user generated information can provide a source of value-relevant advices to other stakeholders.

For instance, Antweiler and Frank (2004) apply sentiment analysis to analyze 1.5 million messages from Yahoo! Finance. The result of their study shows that postings on social media are positively associated with firm's stock returns on the next day. Das and Chen (2007) analyze small investor's sentiment on stock message boards. They find that social media postings are related to stock index levels, volumes and volatility. Chen et al. (2014) use posts and comments on Seeking Alpha.com to examine the effect of users' sentiment on stock prices. They find that negative sentiment can negatively predict stock returns in a three-month period. Overall, the studies above imply that the sentiment in large amount of postings is an important indicator of firm's potential risks and values. Since borrower's risks are positively associated with bank loan prices, I have the hypothesis below:

H1b: Borrowers with more positive postings and comments on social media enjoy more favorable loan prices.

A number of studies have examined the impact of firm size on information asymmetry and loan terms. For instance, Dennis and Sharpe (2005) identify three influence factors related to firm size. The first factor is the bargaining power. Most large firms are well organized and they have significant profits. The reputation built over long time makes it easier for large firms to build a close relationship with lenders. As a result, the loan bargaining power of large firms is stronger than small firms. From lenders' perspective , it's reasonable for them to offer favorite loan contracts to large firms since the creditworthiness of large firms is easier to evaluate (Diamond, 1989). On the other hand, large firms are less likely to rely on the loans from single lender and market. More borrowing options enable large firms to receive relatively good terms from multiple lenders. The second influence factor is the transaction cost. Compared to larger firms, the amount of small firm loan is relatively small. The fixed costs incurred by the transaction make it harder for lenders to get economies of scale from small loans. Thus, lenders are reluctant to offer favorable loan terms to small firms. The third influence factor is information asymmetry. Shen and Reuer (2005) find that large firms have tighter regulation requirements to disclose information, thus large firms is more efficient at minimizing information asymmetry. Easley and O'hara (2004) suggest that large firms disclose more information than small firms, which result in reduced information asymmetry and lower cost of capital.

Beside the amount of information disclosed, firm size also has negative impact on the quality and scope of information disclosed. For instance, press is an important tool to disseminate information generated by sources like analysts, legal suits, and auditor changes. Miller (2006) examines the role of press in the rebroadcasting stage. He finds that press is often biased toward coverage of large firms because they tend to cater to interests of main readers. Das, Levine, and Sivaramakrishnan (1998) study the determinant of cross-section differences in analyst forecasting. They find that large firms are generally followed by more analysts. Analyst forecasts for large firms are less biased than those for small firms. Das et al's study shows that small firms have less opportunity to get favorable loan terms due to the disadvantages of information dissemination in traditional media channels. In general, prior findings indicate that 1) the information of the small firms is often opaque; 2) disclosures of small firms are often biased; 3) disclosures of small firms cannot reach the public in a broad and correct manner. These studies also imply that banks may have to rely more on alternative tools (e.g., social media and XBRL) to evaluate the risk of small firms. Based on this argument, I hypothesize:

H2a: The effect of XBRL adoption on bank loan prices is stronger for small firms than for large firms.

H2b: The effect of social media sentiment on bank loan prices is stronger for small firms than for large firms.

Some studies find that prior borrowing relationship is associated with loan price. For instance, Degryse and Van Cayseele (2000) find that the scope of a relationship increase the probability of getting loans from that bank. Berger and Udell (1995) suggest that prior relationships will generate valuable information of borrower quality. This argument is confirmed by their findings, which longer banking relationships are associated with lower interest rates and fewer collateral requirements. Petersen and Rajan (1994) examine the relationship among prior relationships, loan availability and loan cost. They find that while prior relationships do affect the loan availability, there are only very small effects on the loan price. Bharath, Dahiya, Saunders, and Srinivasan (2011) 's research focuses on the main risks in bank loan. They identify three effects of prior loan relationships on bank loan. The first effect is adverse selection concerns reduction. When the lenders have prior relationships with the borrowers, multiple interactions allow lenders to collect borrowers' inside information, which is hard to gather through other channels. In this case, adverse selection risks are reduced. The second effect is syndicate moral hazard reduction. Syndicate moral hazard results from information asymmetries among lenders. In the syndicated loan, the lead lender is responsible for the cost of monitoring borrowers. Since loan is divided among more than one lender, the lead lender may not endeavor to monitor the borrowers efficiently as the single loan. In order to compensate for the potential loss caused by inefficient monitoring, the other syndicate members will demand tighter loan terms. As prior relationships reduce information asymmetry and lower the cost of future monitoring, the possibility of syndicate moral hazard is reduced. The third effect is borrower moral hazard reduction. Borrower moral hazard is a risk results from information asymmetry between borrowers and lenders. The borrower moral hazard is raised when lenders cannot observe or predict the borrower's risky activities (Hölmstrom, 1979). As prior loan relationships make it easier for lenders to monitor and control the borrower, the possibility of borrower moral hazard is reduced. As lenders can benefit a lot from the information gathered from prior borrowing relationship, they will be less likely to rely on other information sources (Berger & Udell, 1995; Harhoff & Körting, 1998; Petersen & Rajan, 1994). On the other hand, when lenders have limited access to the private information of new borrowers, they may have to rely more on other information sources such as XRBL and social media to evaluate borrowers' risks. Based on this argument, I hypothesize:

H3a: The effect of XBRL adoption on bank loan prices is stronger for loans in new relationships than for loans in existing relationships

H3b: The effect of social media sentiment on bank loan prices is stronger for loans in new relationships than for loans in existing relationships

A syndicated loan is a loan issued to the borrowers jointly by more than one lender. The motivation of syndication stems from lenders' demands to spread risk, enhance income or reduce costs(Pennacchi, 1988). In a syndicated loan, each participant is responsible for a share of the loan. And, only the lead lender supervises the arrangement of the syndication including loan terms negotiation, borrower monitoring and administration of repayments(Simons, 1993). A risk will rise when there is information asymmetry between lead lender and other participants. Sufi (2007) suggests that the lead lender may not endeavor to monitor the borrowers since borrower monitoring is costly and lead lender owns only part of the loan. He calls this kind of risk as syndicate moral hazard. Consistent with Sufi's finding, Hasan et al. (2012) find that syndicate members rely more on public information due to risk of the syndicate moral hazard. Since XBRL and social media allow users to retrieve public information more efficiently, I expect that lenders will rely more on XBRL and social media sentiment in loan contracting. Thus, I hypothesize:

H4a: The effect of XBRL adoption on bank loan prices is stronger for syndicated loans than for sole-owner loans

H4b: The effect of social media sentiment on bank loan prices is stronger for syndicated loans than for sole-owner loans

Banks use various non-price terms to minimize information problems and control loan risks. The most popular non-price loan terms are loan maturities, collateral, and covenants(Hasan et al., 2012). Many studies find that higher information asymmetry, greater uncertainty and lower firm quality of the borrowers lead to shorter maturities and more requirements of collateral (Barclay & Smith, 1995; Ortiz-Molina & Penas, 2008; Rajan & Winton, 1995). For instance, Flannery (1986) suggests that the relationship between borrower's quality and loan maturity is linear. Borrowers with good quality may prefer shorter maturity when the transaction costs are high. Diamond (1991a) examines the determinants of loan maturities from two aspects: information asymmetries and the liquidity risk of refinancing. He argues that long maturity allows borrowers to minimize liquidity risk, and borrowers with average quality will prefer long maturity to avoid changes of loan prices. On the other hand, short-term loan enables lenders to monitor borrowers more frequently. Thus, low quality borrowers are more likely to receive loans with short maturities. Diamond (1991a) points out that even both high quality borrowers and low quality borrowers use short term loan, they are driven by different mechanisms: different bargaining power determinates that low quality borrowers are forced to accept short term loan while high quality borrowers choose short term loan on their own initiative. Barclay and Smith (1995) and Ortiz-Molina and Penas (2008) get similar results with Flannery's findings. Both of their studies suggest that shorter maturities are useful in solving information problems. The reason is that lenders can periodically gather borrowers' private information through short-term loan renewal processes. Applying Diamond's theory, Bharath et al. (2011) examines the effects of lending relationships on loan contract terms. They find that lenders are more likely to monitor low quality borrowers more intensively, and the quality of borrowers is key determination of loan maturity length. In sum, the literature suggests that shorter maturities facilitate continual monitoring, which in turn reduces the information asymmetry between lenders and borrowers. As I discussed above, the adoption of XBRL and social media can reduce information asymmetry between lenders and borrowers significantly. Thus, the substitution effect of XBRL and social media will make loan maturity less important to lenders. At the same time, since sentiment on social media sites is an important indicator of firm's risks, it is reasonable for us to believe that social media sentiment is positively associated with the length of loan maturity. Thus, I hypothesize:

H5a: Borrowers that adopt XBRL have longer loan maturities.

H5b: Borrowers with more positive postings and comments on social media have longer loan maturities.

The literature has found that many factors including limited contract enforceability(Albuquerque & Hopenhayn, 2004; Banerjee & Newman, 1993; Cooley, Marimon, & Quadrini, 2004), high monitoring cost(Border & Sobel, 1987; Boyd & Smith, 1993; Gale & Hellwig, 1985; Lacker, 1998; Townsend, 1979; Williamson, 1986), high loan risk(Holmstrom & Tirole, 1997), efficient monitoring(Rajan & Winton, 1995), and adverse selection(Berger, Espinosa-Vega, Frame, & Miller, 2011; Duarte, 2011) are related to the use of loan collateral. Most of these findings can be explained by adverse selection models and borrower moral hazard models(Bharath et al., 2011). According to adverse selection models, collateral and interest rate are complementary. Higher quality borrowers may choose lower premiums without collateral, while low quality borrowers will prefer higher premiums with collateral. At this point, collateral plays a role in signaling borrower quality (Beaudry & Poitevin, 1995; Stiglitz & Weiss, 1981; Whette, 1983). Borrower moral hazard models explain the motivation of collateral from another aspect. These models suggest that the use of collateral is to compensate for potential losses caused by information asymmetries. Banks are more likely to require collateral from borrowers with low information transparency, while borrowers with high information transparency are less likely to be required to do so(Holmstrom & Tirole, 1997; Stulz & Johnson, 1985). In general, the literature indicates that information asymmetry between borrowers and lenders play an important role in the use of loan collateral. As the adoption of XBRL can reduce information asymmetry, I expect that the adoption of XBRL will lessen lenders' need for collateral. On the other hand, positive sentiment on social media sites indicates better borrowers' quality and lower loan risks. Therefore, I expect that lenders are less likely to require collateral from the borrowers with more positive sentiment. Based on these arguments, I hypothesize:

H6a: Banks are less likely to require collateral from borrowers that adopt XBRL

H6b: Banks are less likely to require collateral from borrowers with more positive postings and comments on social media.

3. METHODOLOGY

3.1 BANK LOAN PRICING ANALYTICS SYSTEM

The objectives of this study are to investigate whether and how the adoption of XBRL and social media sentiment affect bank loan price. In order to achieve these objectives, I develop a Bank Loan Pricing Analytics System (BLPAS) to collect, integrate, and analyze data from multiple data resources. Figure 4 shows the framework for loan price analytics. This system is composed of three key components: data collection module, data integration module and data processing module.

[Insert Figure 4 here]

3.1.1 DATA COLLECTION MODULE

The Data Collection Module manages the data collected from various data sources. This module integrates a number of data collection tools.

1. Web-crawling: currently, many social media tools offer application programming interfaces (API) for users to retrieve data from their platforms. For instance, on Twitter platform, REST API allows developers to access core Twitter data including author profile, follower data, and user information. Streaming APIs enable users to retrieve updates of Tweets synchronously(Twitter, 2016). As the social media data sources used in this study do not provide APIs for data tracking, I develop a web-crawling application to retrieve the content from the websites. By analyzing the URL and particular HTML tags of web pages, this application allows users to retrieve posting and comments published on Seeking Alpha and Yahoo finance.

2. HTML parsing and RSS: Sec.gov offers RSS feeds which enable users to track the URL of submitted financial statements. In order to retrieve financial statements from Sec.gov, I develop a HTML parsing application to download and analyze the data. As XBRL based financial statement uses XML and XSD as the file extensions, this program can easily distinguish XBRL based financial statements from traditional financial statements.

3. Manual coping: Thomson and Compustat database offer extensive functions for users to retrieve information, generate report through pre-defined criteria and Excel add-in. Therefore, I manually collect bank loan data and firm accounting data from these two databases.

3.1.2 DATA INTEGRATION MODULE

A big challenge of loan data analytics is to integrate heterogeneous unstructured data collected from different sources. Built on Apache Hadoop and Hive platform, the Data Integration Module enables flexible data summarization, easy data querying, and analysis of large volumes of data. The Data Integration Module is composed of two components: ETL agents and Distributed Storage Platform.

1. ETL agents: The ETL agents aim to extract data from heterogeneous data sources, clean collected data, and load the processed data into distributed storage platform. ETL agents are deployed as server side applications using PHP and C# programming language. In order to process different types of unstructured data such as txt files, Excel files, html files, I develop multiple agents to extract and transform unstructured data. For instance, HTML pages contain HTML tags and ads in HTML pages, which cannot be processed by Distributed Storage Platform directly. Thus, HTML agent is used to analyze the structure of web pages, and to retrieve the content of postings and comments. After cleaning and mapping data collected from web pages, HTML agent loads the transformed social media data into Distributed Storage Platform.

2. Hive Based Distributed Storage Platform (HBDSP): Bank loan price analysis involves multiple big datasets including social media data, XBRL data, firm financial data, and bank loan data. The large amount of data generated from various data sources makes it hard for traditional storage platform to manage data with reliability and availability. In this case, Hive based distributed storage platform (HBDSP) provides a flexible solution with the capabilities required to support large scale datasets. The main advantages of HBDSP are that datasets are stored on Hadoop Distributed File System (HDFS), which offers key features such as scalability and redundancy on Hadoop platform. The SQL like Hive query language enables users who are familiar with SQL to query and manage the data more efficiently(Chong & Shi, 2015; Thusoo et al., 2009). Figure 5 shows the user interface of Hive Based Distributed Storage Platform.

[Insert Figure 5 here]

The main components of HBDSP include user interface, driver, compiler, metadata store , and execution engine. HBDSP allows users to use ODBC and JDBC API to create, insert, update, and query structured data. Based on Leverenz (2016) 's approach, I define the query flow of HBDSP which includes the following steps:

1. Data Analysis Module sends query to HBDSP through ODBC or JDBC API.

2. ODBC or JDBC driver sends query to compiler.

3. Compiler parses the query to check the syntax and the requirements of the query.

4. The compiler sends a request to Metadata store. After Metadata store returns the metadata, compiler generates the execution plan.

5. Execution engine executes the execution plan.

6. The execution engine sends the execution job to JobTracker, which is responsible for the job assignment.

7. Data nodes return the results to driver through execution engine.

8. Driver returns the results of query to Data Analysis Module through ODBC or JDBC API.

3.1.3 DATA ANALYSIS MODULE

The Data Analysis Module manages serials of analysis tools, which can be used to conduct content analysis, sentimental analysis, statistical analysis et al. The key components of this module include data exchange component, sentimental analysis tools and statistical analysis tools.

1. Data exchange component. Bank loan price analysis involves various data formats including txt file, csv file and database file et al. The transformation and sharing of data among different applications are extremely important to data analysis. Data exchange component provides a serial of functions to facilitate data exchange between analysis applications and Hive platform. On the one hand, this component uses ODBC and JDBC API to retrieve and manipulate data stored on Hive distributed storage system. On the other hand, this component outputs results with multiple data formats, which can be accessed by various analysis tools. For instance, SPSS is a major statistical tools used in this study. However, SPSS cannot access data stored on Hive platform directly. The data exchange component offers a bi-directional data exchange between SPSS and Hive platform with the following steps: 1). Users issue search requests using standard SQL language. 2). Data exchange component transforms and sends search requests to Hive platform. 3). Hive platform returns the results to Data exchange

component. 4). Data exchange component transforms the format of the results to the data format supported by SPSS. 5). Data exchange component sends the transformed results to SPSS.

2. Sentimental analysis tool. In this study, sentimental analysis tool is used to analyze users' opinions expressed within social media postings and comments. Two sentiment scores are calculated in the study. The first one is general sentiment score. This score is used to identify general opinions of the users. I use AlchemyAPI (AlchemyAPI Inc, 2016) to calculate general sentiment score. Compared to other Sentimental analysis tools, AlchemyAPI adopts hybrid approaches including linguistic techniques, statistical analysis techniques and large-scale learning techniques to identify the contents of sentences and phrases. These techniques enable AlchemyAPI to better understand the sentiment expressed in the content and mine key entities and topics more accurately(IBM, 2016). The second score is finance-related sentiment, which is used to identify opinions related to financial information. Following Nann, Krauss, and Schoder (2013)'s suggestion, I calculate finance-related sentiment based on predefined lists of positive(e.g. invest, long, earn, etc) and negative words(e.g. bailout, breakout, sell, etc). The final sentiment is the combination of general sentiment and finance-related sentiment.

3. Statistical analysis tools. Statistical analysis tools include two popular statistical packages: SPSS and R. As a frequently used statistical analysis tool, SPSS allows users to conduct most complex statistical analyses and easily generate summary reports, charts and descriptive statistics (Hayes & Matthes, 2009). In this study, SPSS is used to conduct most of the statistical analyses and generate reports of the analyses. Another statistical analysis tool used in this study is R. As an integrated suite of statistical analysis, R integrates most popular statistical techniques and has more than 4000 packages for data analysis. The R programming language allows users to add, use, and modify functionalities. This feature makes R more extensible than

other statistical analysis tools(The R Foundation, 2016). In this study, I conduct robustness analysis on R environment.

Besides the tools introduced above, data analysis module also provides simple and flexible interfaces for future system expansion. Future analytic applications can use JDBC API, ODBC API and XML to exchange data with current analysis tools and Hive storage platform.

3.2 DATA SOURCE

In this study, XBRL data is collected from the archive of financial statements on www.sec.gov. According to SEC's requirements, the adoption of XBRL is phased in three stages. In the first stage, all domestic and foreign large accelerated filers matching certain conditions (use U.S. Generally Accepted Accounting Principles (GAAP), have a worldwide public float of greater than $5 billion as of the end of the second fiscal quarter of their most recently completed fiscal year, fiscal periods ending on or after June 15, 2009) are required to provide financial statement information in XBRL. In the second stage, all other domestic and foreign large accelerated filers matching certain conditions (use U.S. GAAP, fiscal period ending on or after June 15, 2010), are required to include interactive data (XBRL) in their financial statements. In the third stage, all remaining filers including foreign private issuers using International Financial Reporting Standards (IFRS) matching certain conditions(fiscal periods ending on or after June 15, 2011), are required to include interactive data(XBRL) in their financial reporting(Securities Exchange Commission, 2009). Since the adoption of XBRL was not mandatory for all companies from 2009 to 2011, some companies adopted XBRL while others did not during three years. This circumstance makes it possible for me to examine the impact of XBRL adoption on bank loan price. Before June 15, 2009, no companies were required to provide financial statement information in XBRL. The adoption of XBRL did not have any impact on bank loan price before 2009. On the other hand, the rule of SEC requires all companies to submit their financial information in XBRL after June 15, 2011. Bank loan price were not related to the adoption of XBRL after 2011 since all borrowers included XBRL in their financial statement after that deadline. For the above reasons, I collect data of XBRL adoption in the period from 2009 to 2011.

Bank loan data is obtained from Thomson one database. Thomson one provides detailed financial information such as earnings estimates, financial news, transaction data, mergers and acquisitions, ownership profiles and analysts' reports. More importantly, this database contains over 92,000 syndicated loans since 1982. Key loan terms such as corporate profile of borrowers and lenders, deal dates, interest rate, collateral, covenants, maturities are well covered by this database(Thomson Reuters, 2016). While Thomson one provides detailed loan information, some firm-specific accounting information such as tangibility, profitability, financial ratios, leverage are not included in the database. In this case, I use Compustat to obtain firm-specific accounting information. As I only examine the adoption of XBRL from 2009 to 2011, the bank loans made at the end of 2011 enable me to evaluate the comprehensive impact of XBRL on bank loan price. Thus, I collect bank loan data in a period from August 1, 2011 to December 31, 2011. To ensure data integrity and data consistency, I use ticker symbols to match the records collected from Thomson one, Compustat and SEC.gov.

Social media data is collected from two financial social networking websites, Yahoo Finance and Seeking alpha. Yahoo Finance is a leading financial data website, which has more than 70 million visitors each month. This website provides financial news, stock quotes, press releases, financial reports and financial analysis(Wikipedia, 2016). As an important part of Yahoo finance, Yahoo Message Board allows investors and analysts to share their opinions and views on companies all over the world. Figure 6 shows the message board of Microsoft on Yahoo finance. One key feature of Yahoo message board is that for each post on the board, users can reply and click the sentiment label to indicate their preferences to the post. The huge amounts of posts and comments on Yahoo Message Board make it easier for me to mine sentiments of millions of users. The second social media data source, Seeking Alpha is one of the biggest financial social media websites in the U.S. The aim of Seeking Alpha is to provide professional opinion and analysis written by experts(Chen et al., 2014). In order to assure quality of the articles, the articles submitted to Seeking Alpha are reviewed by a panel. The breadths and depth of the articles on Seeking Alpha are much better than other websites(Seeking Alpha, 2016a). Similar to Yahoo finance, users of Seeking Alpha are free to leave comments on articles that they are interested in. Figure 7 shows the example articles and comments of Microsoft on Seeking Alpha.

[Insert Figure 6 here]

[Insert Figure 7 here]

Before making a loan decision, it requires a long time for lenders to collect borrower's information and evaluate potential loan risks. Therefore, social media views and opinions published in a short time period may not be fully noticed or evaluated by lenders. In order to improve the accuracy of this study, I collect social media data in the four-month period prior to the loan announcement date.

3.3 SAMPLE SELECTION

In order to ensure the accuracy and consistency of the data collected from multiple data sources. I use the following procedures to form the sample:

1. First, I retrieve all loan-facility-level records from Thomson one database in a period from August 1, 2011 to December 31, 2011. While the data of private or non-US companies are not completely covered by SEC.gov and social media websites, I limit my sample to listed US companies.

2. Following Hasan et al. (2012)'s definition, I measure the loan price as the basis points over LIBOR or LIBOR equivalent. In this step, I exclude all records without available LIBOR price.

3. Collect borrowers' 10-K submission from SEC.gov in a period from January 1, 2009 to December 31, 2011. Exclude all borrowers without submission of 10-k during this period and borrowers being delisted during this period. I also manually check the names of all borrowers. Borrowers without consistent names on SEC.gov and Thomson one are excluded.

4. Collect borrower-related articles, and comments from Yahoo message board and Seeking Alpha in a four-month period prior to each loan announcement date. Borrowers without posts on these two sites or have inconsistent company names are excluded.

Finally, I get 554 loan records. Table 9 shows the details of sample selection.

[Insert Table 9 here]

3.4 VARIABLE MEASUREMENT

3.4.1 BANK LOAN PRICE

Loan price is one of the key variables in this study. A number of prior studies use spread as the measurement of loan price (Beatty, Weber, & Yu, 2008; Hasan et al., 2012; Liu, Seyyed,

& Smith, 1999; Zhang, 2008). Following these studies, I use the initial interest rate spread over London Interbank Offered Rate (LIBOR) to measure bank loan price.

3.4.2 THE ADOPTION OF XBRL

Following Kim et al. (2011)'s approach, I construct the adoption of XBRL measure by examining borrowers ' 10-K submission from 2009 to 2011. The value of this variable equals to the frequency of the adoption of XBRL between 2009 and 2011. For instance, if a firm submitted its financial information in XBRL once between 2009 and 2011, then the value of this variable is equal to one. Since the time period only covers three fiscal years, the scale of XBRL adoption is from 0 to 3. Based on the discussion in chapter 2, I expect that the frequency of XBRL adoption is negatively related to loan prices.

3.4.3 SOCIAL MEDIA SENTIMENT

The sentiment on social media websites reflects user's opinions about firm's risks and values. According to prior discussion, I expect that social media sentiment is negatively related to loan prices. Based on the algorithm provided by IBM(AlchemyAPI Inc, 2016), I calculate social media sentiment using the following stages:

1. Calculate financial terms sentiment score of each post

$$\mathbf{Finsen_i} = \frac{(Count\_P(post_i) - Count\_N(post_i))}{Count\_All(post_i)} \tag{1}$$

Where $Count\_P(post_i)$ returns the number of positive financial terms in post *i*. $Count\_N(post_i)$ returns the number of negative financial terms in post *i*. $Count\_All(post_i)$

returns the number of all financial terms in post *i*. This formula returns the finance-related sentiment score of post *i*.

2. Calculate general sentiment score,

$$\mathbf{StandSen_i} = \mathit{AlchemyAPI}(\text{ post}_i\ ) \tag{2}$$

Where $post_i$ is the content of post i. By calling *AlchemyAPI,* this formula returns the general sentiment score of post *i*.

3. Calculate total sentiment score,

$$\mathbf{Sentiment\_Score} = \frac{(\sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} FinSen_{\boldsymbol{i}} + \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} StandSen_{\boldsymbol{i}})}{2n} \tag{3}$$

Where *FinSeni* is the finance-related sentiment score of post i. StandSeni is general sentiment score of post i. n is the number of all posts over a specific time period. This formula returns the total sentiment score of a company over a specific time period.

3.4.4 CONTROL VARIABLES

Prior studies find that some firm characteristics also affect loan price (Graham, Li, & Qiu, 2008; Qian & Strahan, 2007; Wu et al., 2011). Following these studies, I control for several firm characteristics and loan characteristics in my models.

Log(Asset), prior studies find that firm size has positive impact on information transparency(Dennis & Sharpe, 2005). The information of the small firms is often opaque. While information transparency plays an important role in loan prices, I expect that firm size has a negative impact on bank loan prices. In this study, I use the natural logarithm of the total assets of borrowers as the measurement of firm size.

Leverage, I define leverage as the total debts including long-term debt and short-term debt divided by firm book assets. As an important indicator of default risk and liquidity risk(Diamond, 1991a; Graham et al., 2008), high leverage often suggests that firms have higher default risk and liquidity risk. Thus, I expect leverage is positively associated with loan prices.

Current Ratio and Cash to Debt Ratio, in this study, current ratio is defined as the current assets divided by current liability, cash to debt ratios is measured by total cash divided by total debt. According to the literature, lower current ratios and lower cash to debt ratios suggest that firms have higher default risk (Graham et al., 2008). Thus, I expect a negative relation between current ratio, cash to debt ratio and loan prices.

Profitability, I measure profitability as Net income over total sales. While higher profitability means lower loan risks, I expect that profitability is negatively associated with loan prices.

Interest Coverage, I measure interest coverage as Earnings Before Interest & Tax (EBIT) divided by total interest expense. This variable reflects firm's ability to pay the interest charges on time. Low interest coverage often suggests that firm does not have enough cash to pay off interest charges. Thus, I expect that interest coverage is negatively associated with loan prices.

Tangibility, I define tangibility as property, plant, and equipment divided by total asset. As tangible assets are potential guarantee for banks to recover from default (Hasan et al., 2012), I expect a negative relationship between tangibility and loan prices.

MB, this variable is defined as market to book ratio that indicates growth opportunities of a firm. Since better growth opportunities result in lower loan cost(Kothari et al., 2009), I expect that *MB* is negatively associated with loan prices.

Loan Size, I define Loan Size as the total amount of bank loan. Hasan et al. (2012) suggest that lenders can achieve economies of scale when offering large amount of loan. As large loan size is associated with lower lending cost, I expect a negative relationship between Loan Size and loan prices.

Prior Relations, is a dummy variable that is equal to one when there is a previous lending relationship between lenders and borrowers, it equals zero otherwise. A number of studies have examined the relationship between repeated borrowing and loan prices. Boot (2000) argues that repeated interaction between the same lender and borrower will facilitate production of durable and reusable information. Bharath et al. (2011) suggest that previous borrowing relationship can reduce information asymmetries, and lending cost. According to their studies, previous borrowing relationship contributes to 10–17 bps lower loan spreads. Based on the findings of literature, I expect a negative relationship between Prior Relations and loan prices.

Z-Score, Z-Score is a measurement of default risk. Follow Hasan et al. (2012) 's approach, I define Z-Score as (1.2*Working capital+1.48Retained earnings + 3.3*EBIT + 0.999*Sales)/Total assets. As lower Z-Score is associated with higher default risk, I expect a negative relationship between Z-Score and loan prices

The literature finds that some other variables such as loan type, loan purpose and industry also have impact on loan prices(Graham et al., 2008; Hasan et al., 2012). Following these studies, I control for loan type, loan purpose, and industry effects. In this study, loan type is classified into five categories: term loan, term loan B-D, revolver, 364-Day Facility and others. Loan purpose is classified into seven categories: acquisition lines, LBO, Takeover, debt repayment, corporate purpose, working capital, and others. For the industry effects, I use one digit SIC dummies to control for it.

3.5 EMPIRICAL MODELS

To examine the effect of adoption of XBRL and social media on bank loan price, I specify the following empirical models:

$$\text{Log(Spread)} = f(\text{Constant, XBRL adoption, Firm characteristics, Loan characteristics, Industry effects}) \quad (4)$$

The first model is used to test whether and how the adoption of XBRL affect bank loan price. The explicit form of equation (1) above is represented as follows:

$$\begin{aligned}\text{Log(Spread)} = {} & \alpha_0 + \alpha_1(\text{XBRL adoption}) + \alpha_2 \log(\text{Assets}) + \alpha_3\text{Leverage} \\ & + \alpha_4\text{Current Ratio} + \alpha_5\text{Cash to Debt Ratio} + \alpha_6\text{Interest Coverage} \\ & + \alpha_7\text{Tangibility} + \alpha_8\text{Profitability} + \alpha_9\text{M/B} + \alpha_{10}\text{Log(Loan size)} \\ & + \alpha_{11}\text{Prior Relation} + \text{Loan type effect} + \text{Loan purpose effect} \\ & + \text{Industry effect} + \varepsilon \qquad (5)\end{aligned}$$

Where Log(Spread) is the natural logarithm of initial interest rate spread over London Interbank Offered Rate (LIBOR). XBRL adoption equals to the frequency of the adoption of XBRL between 2009 and 2011. I expect a negative relationship between XBRL adoption, firms size(Assets), Current Ratio, Cash to Debt Ratio, Interest Coverage, Tangibility, Profitability, M/B, Loan Size, Prior Relation and loan prices and a positive relationship between Leverage and loan prices.

The second model is used to examine the effect social media sentiment on loan prices by using the following model:

$$\text{Log(Spread)} = f(\text{Constant, Social Media Sentiment, Firm-specific variables, loan-specific variables, other control variables}) \quad (6)$$

The explicit form of equation (6) above is represented as follows:

$$\begin{aligned}\text{Log(Spread)} = {} & \alpha_0 + \alpha_1(\text{Social Media Sentiment}) + \alpha_2 \log(\text{Assets}) + \alpha_3\text{Leverage} \\ & + \alpha_4\text{Current Ratio} + \alpha_5\text{Cash to Debt Ratio} + \alpha_6\text{Interest Coverage} \\ & + \alpha_7\text{Tangibility} + \alpha_8\text{Profitability} + \alpha_9\text{M/B} + \alpha_{10}\text{Log(Loan size)} \\ & + \alpha_{11}\text{Prior Relation} + \text{Loan type effect} + \text{Loan purpose effect} \\ & + \text{Industry effect} + \varepsilon \qquad (7)\end{aligned}$$

Where social media sentiment is the sentiment score calculated by formulas mentioned in section 3.4.3. I expect a negative relationship between social media sentiment, firms size(Assets), Current Ratio, Cash to Debt Ratio, Interest Coverage, Tangibility, Profitability, *M/B*, Loan size, Prior Relation and loan prices and a positive relationship between Leverage and loan prices.

In the third model, I add the interaction of social media sentiment and XBRL adoption in the regression. The model examines the comprehensive effects of XBRL adoption and social media sentiment on bank loan price:

$$\begin{aligned}\text{Log(Spread)} = f(& \text{Constant, XBRL adoption, Social Media Sentiment,} \\ & \text{XBRL adoption} \times \text{social media sentiment, Firm-specific variables,} \\ & \text{loan-specific variables, other control variables}) \qquad (8)\end{aligned}$$

The explicit form of equation (8) above is represented as follows:

$$
\begin{aligned}
\text{Log(Spread)} = {} & \alpha_0 + \alpha_1(\text{XBRL adoption}) + \alpha_2(\text{Social Media Sentiment}) \\
& + \alpha_3(\text{XBRL adoption} \times \text{ Social Media Sentiment}) + \alpha_4 \log(\text{Assets}) \\
& + \alpha_5\text{Leverage} + \alpha_6\text{Current Ratio} + \alpha_7\text{Cash to Debt Ratio} \\
& + \alpha_8\text{Interest Coverage} + \alpha_9\text{Tangibility} + \alpha_{10}\text{Profitability} + \alpha_{11}\text{M/B} \\
& + \alpha_{12}\text{Log(Loan size)} + \alpha_{13}\text{Prior Relation} + \text{Loan type effect} \\
& + \text{Loan purpose effect} + \text{Industry effect} + \varepsilon \qquad (9)
\end{aligned}
$$

I expect that interaction of social media sentiment and XBRL adoption is negatively associated with loan cost.

4. FINDINGS AND ANALYSES

In this chapter, I will discuss the results of the regression analysis: The descriptive statistics of the dependent and independent variables are provided in section 4.1. The correlation analyses between dependent and independent variables are also discussed in this section. Section 4.2 and section 4.3 presents the regression results related to the effects of XBRL adoption, and social media sentiment on loan prices. After discussing the results of robust test in section 4.4, section 4.5, 4.6, and 4.7 investigate whether the effect of XBRL and social media sentiment on bank loan prices varies with borrower characteristics and loan characteristics (e.g. new relationship, syndication, and firm size). Finally, section 4.8 discusses the regression results regarding the relations between the adoption of XBRL, social media sentiment and non-price terms.

4.1 DESCRIPTIVE STATISTICS

Table 10 presents descriptive statistics for dependent variables, independent variables and a set of control variables in our models. The mean of bank loan spread is 199.621 with a median value of 175, minimum value of 5 and maximum value of 1150. The bank loan spread highly dispersed from its mean value with the standard deviation of 123.194. The mean of XBRL is 0.894 with a median value of 1, minimum value of 0 and maximum value of 2. This result indicates that the maximum times borrowers adopted XBRL between 2009 and 2011 is two, and most of the borrowers only used XBRL one time during that period. The mean of Seeking Aplaha sentiment is 0.217 with a median value of 0.221, minimum value of -0.706, and maximum value of 1.200. The mean of Yahoo finance sentiment is 0.026 with a median value of -0.002, minimum value of 1.161, and maximum value of 1.350. The statistics of two social media sites show that the sentiment of Seeking Aplaha is more positive than the sentiment of

Yahoo finance. Some other borrower firm characteristics, such as Current Ratio, Tangibility, and Interest Coverage, also vary across our sample.

[Insert Table 10 here]

For the loan characteristics, the result reveals that the sample's average loan highly dispersed from its mean value as the mean of loan size $700M with a median value of $350m, minimum value of $5M and maximum value of $15,000M. Other loan characteristics including maturities, collateral, and relations vary considerably across the sample. The mean value of maturities is 53 months. The maximum and minimum values of maturities are 85 months and 3 months respectively. The standard deviation of 14 shows little dispersion of maturities from its mean. The mean and median value of secured is 0.128 and 0, only 12.8% of the sample loans have collateral requirements. This result also indicates that most of the borrowers (69%) have prior relations with lenders. The mean of relation is 0.690 with a median value of 1, minimum value of 0 and maximum value of 1.

Correlation is a way to measure whether and how two or more variables are related to each other. In this study, Pearson correlation coefficient is used to measure the correlation between variables. Table 11 provides the result of Pearson correlation analysis. As expected, the bank loan spreads are negatively correlated with Social Media postings at 5% level, and negatively correlated with XBRL adoption at 1% level. This result shows the preliminary evidence about the effect of XBRL and social media on bank loan spreads. Also, the highest variance inflation factor (VIF) in our regression is only 5.800, which is below the suggested multicollinearity problem threshold of 10(Gefen, Straub, & Boudreau, 2000) . This result indicates that multicollinearity is unlikely to be a problem in this study.

[Insert Table 11 here]

4.2 XBRL ADOPTION AND BANK LOAN PRICE

In this section, I use OLS regression analysis to examine the relationship between XBRL adoption and bank loan price, and the relationship between social media sentiment and bank loan price. The regression results are reported in table 12.The dependent variable is the natural logarithm of loan spread. In column 1, I use *XBRL* to measure the adoption of XBRL. Firm and loan characteristics such as Asset, Leverage, Current Ratio, Cash to Debt Ratio, Profitability, Interest Coverage, Tangibility, MB, zscore, Loan Size, Prior Relations are also included in the regression. Based on the discussion above, I control for industry effect, loan-type effect, and loan-purpose effect in the regressions.

[Insert Table 12 here]

I find that the coefficient of XBRL is -0.132 and is significant at the 1% level (t = -3.242), indicating that a 1% increase in XBRL adoption is related to about a 0. 132% decrease in bank loan spread. This result shows that the effect of XBRL adoption on the bank loan price is statistically significant. Other firm and loan characteristics including log(asset), Profitability, MB, zscore are significantly negatively related to loan spread indicating that banks charge lower interest rates to borrowers with higher earnings quality, more growth opportunities, and lower loan risk. The coefficient of loan size is also negative and significant suggesting that the increase of loan size will reduce monitoring costs of bank loan. However, while Cash to Debt Ratio, Tangibility, and Prior Relations are negatively related to loan spread. These relations are not significant. This result indicates that some firm and loan characteristics only have a weak effect on loan spread.

In sum, the results of Table 12 are consistent with H1a that adoption of XBRL reduces information asymmetry between borrowers and lenders, which enables lenders to offer favorable loan price to borrowers.

4.3 SOCIAL MEDIA SENTIMENT AND BANK LOAN PRICE

In this study, I collect social media data from two websites: Yahoo Finance message board and Seeking Alpha. As discussed above, the postings and comments on social media sites represents users' opinions and views about specific firms. As the sentiment of online postings is an indicator to firm's risks and market performance, I predict that the sentiment of online postings is negatively associated with bank loan price. Column 2 and column 3 of table 12 report the regression results related to two social media sites. The coefficient of Yahoo sentiment is -.047(t=-.701). The p value is .484, which is not significant. This result indicates that Yahoo user opinion only has a weak impact on bank loan price. Interestingly, I find that the coefficient of Seeking Alpha sentiment is -0.157, and is significant at the 5% level (t = -1.894). These results indicate that a 1% increase in Seeking Alpha sentiment is related to about a 0.157 % decrease in bank loan spread. Column 4 shows the how the interaction of XBRL adoption and social media sentiment affect bank price. As I expected, the coefficients of the interaction is -0.284 and is significant at the 5% level (t = -2.337). The result indicates that borrowers adopting XBRL and receiving positive postings are more likely to get favorable loan price. In sum, the results of Seeking Alpha data are consistent with H1b that social media sentiment is negatively associated with loan price.

According to the results in column2 and column3, predictive power between two social media sites is different. In order to identify the cause of this difference, I perform further analysis on the postings of these two sites. First, I conduct a pair t-test analysis to compare the postings means. Table 13 shows the result of the analysis. The result indicates that there is a statistically significant difference between the mean of Seeking Alpha sentiment and Yahoo sentiment. The mean of seekingscore - yahoo is 0.26 indicating the mean of Seeking Alpha sentiment is 0.26

[Insert Table 13 here]

greater than the mean of Yahoo sentiment. Second, I randomly select 100 postings of the same firm from both two websites. I find that the average word of Seeking Alpha is 34 while the average word of Yahoo posting is 15. Table 14 shows the sample of Yahoo postings and Seeking Alpha postings. By manually analyzing the content of these postings, I find that 23% of Yahoo postings are unrelated to particular firm's risks or values while only 8% of Seeking Alpha postings are unrelated to particular firm's risks or values. This result implies that focus and quality of postings also have an impact on the predictive power of social media sentiment.

[Insert Table 14 here]

4.4 ROBUSTNESS CHECKS

4.4.1 CONTROLLING FOR FIRM CLUSTERING

In prior regressions, I control for industry effect, which is based on Standard Industrial Classification (SIC) System. A potential disadvantage of SIC is that it is a static system, which cannot capture the evolvement of firm’s business and industry structure. As an additional robustness check, I use a dynamic classification approach defined in section 2.2.3 to control for industry effect. I first collect the 10-K annual financial statements of all listed firms from 2009 to

2011. As some firms have more than one XBRL based 10-K from 2009 to 2011, I only choose the earliest submission. A spectral clustering method is used to analyze the links between companies and XBRL tags. Finally, 20 clusters are identified from 1799 filings. I then control for cluster effect and re-estimate the regressions. Table 15 reports the regressions results. Similar to the results in table 12, the coefficient of XBRL adoption is -0.160 and is significant at the 1% level, the coefficient of Seeking Alpha sentiment is -0.231 and is significant at the 5% level, the coefficient of interaction between XBRL adoption and Seeking Alpha sentiment is -0.162 and is significant at the 1% level. These results suggest that SIC classification system did not influence the primary results.

[Insert Table 15 here]

4.4.2 EXCLUSION OF FINANCIAL AND UTILITY FIRMS

Prior studies find that financial and utility firms are in regulated industries and may have different loan costs compared to other firms(Hasan et al., 2012; Wu et al., 2011). To remove the effect of these firms, I perform an analysis using a sample that excludes financial and utility firms. Table 16 reports the regressions results. The coefficient of XBRL adoption is -0.128 and is significant at the 1% level, the coefficient of Seeking Alpha sentiment is -0.087 and is significant at the 10% level, the coefficient of interaction between XBRL adoption and Seeking Alpha sentiment is -0.087 and is significant at the 5% level. As the coefficients are very similar to those in Table 12, I conclude that my results are not driven by financial and utility firms.

[Insert Table 16 here]

4.4.3 MEDIAN REGRESSION

Following Hasan et al. (2012)'s suggestion, I perform a median regression to investigate whether loans with extreme interest rates mislead the results. Table 17 shows the results of median regression. The coefficient of XBRL adoption is -0.108 and is significant at the 1% level, the coefficient of Seeking Alpha sentiment is -0.0009 and is significant at the 5% level, the coefficient of interaction between XBRL adoption and Seeking Alpha sentiment is -0.0007 and is significant at the 10% level. These results indicate that outlier is unlikely to be a problem in this study.

[Insert Table 17 here]

4.5 EFFECT OF FIRM SIZE

In this section, I first test H2a to see whether XBRL adoption has a stronger effect on small-sized borrowers. I construct a dummy variable, Small firms, which equals one if a firm's assets are less than the sample median of total assets, and zero otherwise. I add Small firms, the interaction of Small firms and XBRL adoption to the model and run the new regression. The results are in column 1 of table 18. The coefficient of the interaction term between XBRL adoption and Small firms is -0.052, p value is 0.421, which is not significant. The results suggest that XBRL adoption does not have a stronger effect on loan prices of small-sized firms. In this case, H2a is not supported.

[Insert Table 18 here]

Next, I test H2b to see whether the effect of social media sentiment on bank loan prices is stronger for small firms. I add Small firms, the interaction of Small firms and social media sentiment to the model and rerun the regression. Column2 of table 18 shows that the coefficient of the interaction term between social media sentiment and Small firms is 0.370 and is significant at the 5% level. This result suggests that social media sentiment has a stronger effect on loan prices of small-sized firms. Hence, H2b is supported.

4.6 EFFECT OF NEW RELATIONSHIP

To verify H3a that XBRL adoption has a stronger impact on new-relationship lending, I construct a dummy variable, New Loans, which equals one if there is no previous lending relationship between lenders and borrowers, and zero otherwise. I add New Loans, the interaction of New Loans and XBRL adoption to the model and run the new regression. The results are in column 1 of table 19. The coefficient of the interaction term between XBRL adoption and New Loans is -0.05, p value is 0.385 that is not significant at the 10% level. The result suggests that the effect of XBRL adoption on bank loan prices is not stronger for loans in new relationships. In this case, H3a is not supported.

[Insert Table 19 here]

Next, I test H3b to see whether social media sentiment has a stronger impact on new-relationship lending. I add New Loans, the interaction of New Loans and social media sentiment to the model and rerun the regression. Column2 of table 19 shows that the coefficient of the interaction term is 0.270, and p value is 0.145, which is not significant at the 10% level. Hence, H3b is not supported.

4.7 EFFECT OF SYNDICATION

In this section, I first test H4a to see whether XBRL adoption has a stronger impact on syndicated lending. I construct a dummy variable, Syndication, which equals one if a loan is offered by more than one lender, and zero otherwise. I add the Syndication, the interaction of

Syndication and XBRL adoption to the model and run the new regression. The results are in column 1 of table 20. The coefficient of the interaction term between XBRL adoption and Syndication is -0.577, p value is significant at the 1% level. The results suggest the effect of XBRL adoption on bank loan prices is stronger for syndicated loans than for sole-owner loans. In this case, H4a is supported.

[Insert Table 20 here]

Next, I test H4b to see whether social media sentiment has a stronger impact on syndicated lending. I add the Syndication, the interaction of Syndication and social media sentiment to the model and rerun the regression. Column2 of table 20 shows that the coefficient of the interaction term is 1.040, and p value is significant at the 10% level. The results suggest the effect of social media sentiment on bank loan prices is stronger for syndicated loans than for sole-owner loans Hence, H4b is supported.

4.8 XBRL ADOPTION, SOCIAL MEDIA SENTIMENT, AND NON-PRICE LOAN TERMS

Lenders use various non-price terms such as loan maturities and collateral to control loan risk and minimize information problems. The literature has confirmed that higher information asymmetry, greater uncertainty often result in shorter maturity, more requirements of collateral. As XBRL and social media can reduce information asymmetry between lenders and borrowers significantly, I hypothesize that borrowers that adopt XBRL and receive more positive postings on social media are more likely to enjoy favorable non-price terms. This hypothesis is tested in the following sections.

4.8.1 LOAN MATURITY

In this section, I use a new regression model to test whether XBRL Adoption, Social media sentiment affect loan maturities. The dependent variable is the natural logarithm of maturity. I use *XBRL* to measure the adoption of XBRL. Firm and loan characteristics such as asset, leverage, Current Ratio, Cash to Debt Ratio, Profitability, Interest Coverage, Tangibility, MB, zscore, loan size, Prior Relations are also included in the regression. I also control for industry effect, loan-type effect, and loan-purpose effect in the regression. The results are in column1 and column2 of table 21. The coefficient of XBRL adoption is -0.007, p value is 0.872, which is not significant at the 10% level. The result suggests the XBRL adoption is not related to loan maturities. In this case, H5a is not supported. For the social media sentiment, The coefficient is -0.029, p value is 0.718 that is also not significant at the 10% level. The result indicates the social media sentiment does not affect loan maturities. Hence, H5b is not supported.

[Insert Table 21 here]

4.8.2 COLLATERAL

To test the effect of XBRL Adoption, Social media sentiment on collateral, I construct a dummy variable, Secured, which equals one if a loan is secured by collateral, and zero otherwise. Column 1 of table 22 shows the results. I find that the coefficient of XBRL adoption is -0.062 and is significant at the 10% level (t = -1.635), indicating that borrowers adopting XBRL are less likely to be required to provide collateral. Hence, H6a is supported.

[Insert Table 22 here]

Next, I test the relationship between Secured and social media sentiment. Column 2 of table 22 shows the results. I find that the coefficient of social media sentiment is -0.079, p value is 0.204, which is not significant at the 10% level. This result suggests that social media sentiment does not have significant effect on the use of collateral. Hence, H6b is not supported.

5. DISCUSSION AND CONCLUSION

Many studies have demonstrated that XBRL and social media can facilitate the information sharing and minimize information asymmetry in the stock markets. However, few studies empirically test the relations between XRBL and bank loan contracting or relations between social media and bank loan contracting. This study aims to fill this gap by examining the influence of XBRL adoption and social media sentiment on bank loan contracting especially the cost of loan. In the following sections, I summarize the results of this study and offer recommendations for further research.

5.1 RESEARCH IMPLICATIONS

The results of analyses support the idea that XBRL and social media can decrease data redundancy, increase information-processing capability, and facilitate information sharing in the capital markets. The main findings and implications of this study are listed as follows: First, the analysis on XBRL adoption indicates that a 1% increase in XBRL adoption is related to about a 0.132% decrease in bank loan spreads after controlling for certain firm characteristics and loan characteristics. This finding reveals that the XBRL can enhance accounting disclosures and mitigate the information asymmetry problem between borrowers and lenders.

Second, the analysis on social media sentiment shows two different results. 1) While both Yahoo sentiment and Seeking Alpha sentiment are negatively associated with loan spread, the coefficient of Yahoo sentiment is insignificant, indicating that Yahoo user opinions only have a very weak influence on loan price. 2) The coefficient of Seeking Alpha sentiment is significant at the 5% level, suggesting that borrowers receive more positive postings on Seeking Alpha are more likely to enjoy favorable loan price. Further analysis on the postings of these two sites shows that Yahoo postings are more "general" than Seeking Alpha postings. As most articles and comments on Seeking Alpha were published and reviewed by professionals, the information shared on this site has more influence on loan contracting. From this point, the focus and quality of postings may result in different predictive power between Yahoo and Seeking Alpha. Overall, the results of analyses confirm that social media facilitate information sharing and dissemination, and social media sentiment is an important indicator of firm's potential risks and values.

Third, I find that XBRL adoption and social media user opinions have stronger effect on syndicated lending than for sole-owner lending. Hasan et al. (2012) suggest that the members of syndicated loans rely more on public information due to the risk of the syndicate moral hazard. As XBRL and social media facilitate information sharing and retrieving, lenders of syndicated loans are more likely to use these technologies to assess borrower's firm risk and default risk. Therefore, XBRL and social media play a more important role in the syndicated loans.

Fourth, this study confirms the hypothesis that social media sentiment is more important for small-sized borrowers. Prior studies indicate that small-sized firms have less opportunity to get favorable loan terms due to high information asymmetry between lenders and them(Das et al., 1998; Dennis & Sharpe, 2005). My findings suggest that social media can help mitigate information asymmetry between lenders and small-sized borrowers. In this case, small-sized firms can benefit more from social media in loan contracting.

Finally, according to the view of borrower moral hazard models, lenders use collateral to compensate for potential losses caused by information risk and default risk. My study indicates that the adoption of XBRL can reduce the incidence of collateral. I interpret this finding as the evidence that XBRL has the potential to reduce lenders' reliance on traditional non-price loan terms because it can improve transparency and efficiency for loan contracting.

Besides the hypotheses confirmed by this study, some non-significant results also have implications for the literature. First, the analysis shows that XBRL adoption does not have a stronger effect on loan price for small-sized firms. One possible explanation of this result is that XBRL adoption affects loan contracting in multiple ways. While XBRL makes it easier for lenders to evaluate the risk of small firms, large companies also benefit a lot from XBRL adoption. For instance, Roohani (2003) suggest that advantage of adopting XBRL is greater for large companies because XBRL facilitates the integration of business reporting procedures. At this point, XBRL can provide significant benefits to both small and large firms. Second, this study suggests that the substitution effects of XBRL and social media are not as strong as I expected. In H3a and H3b, I predict that the adoption of XBRL and social media sentiment are more important for new-relationship lending. However, these hypotheses are not supported by the analyses. This result implies that when lenders have limited access to the private information of borrowers, they may rely on other traditional tools such as business press or professional databases to evaluate loan risks. In this case, the adoption of XBRL and social media sentiment only has a weak influence on new-relationship lending. This finding may also explain the non-results for the hypotheses concerning the effects of XBRL adoption and social media sentiment on loan maturities (H5a, H5b) and collateral (H6b).

5.2 RESEARCH LIMITATIONS

This study is limited in several ways. First, this study uses two social media websites to analyze the influence of social media sentiment on loan contracts. Some popular websites such as Google Finance and Facebook are not included in the dataset. In addition, the social media dataset used in the experiments is from the postings of two websites in 2011. This dataset may not be able to fully reveal the influence of social media on today's financial markets. As more stakeholders rely on various social media to express their opinions and findings, future research should explore more social media websites and employ more current data to improve the quality of the results.

Second, the models in this study only contain limited control variables. For instance, the access to public debt markets is suggested to have an influence on loan contracting because this access increase borrowers' bargaining power with banks(Hasan et al., 2012). Similarly, Accruals is also found to be associated with bank loan prices. Due to limited data access, these variables are not included in the models. In the future, I will include more variables in the models and investigate how these variables affect the relations between XBRL adoption and loan contracting, and relations between social media user opinions and loan contracting.

Third, prior studies find that the quality of sentiment analysis depends on how the sentiment analysis algorithms specialize to the particular domain(Hogenboom, Bal, Frasincar, & Bal, 2013; Nann et al., 2013). This study uses a predefined domain dictionary to determine financial sentiment in the postings. In the future, more valuable predefined words should be added to the domain dictionary to increase the accuracy of sentiment analysis. In addition, this study only applies single sentiment analysis method, which cannot efficiently extract sentiment from various social media data sources. For instance, the length of posts on Twitter is relatively short. As short tweets do not provide enough word occurrence, it is unsuitable to apply sentiment analysis method used in this study to identify and categorize sentiment in tweets. In this case, a method using the author information and features within the tweets may achieve higher quality in sentiment recognition(Sriram, Fuhry, Demir, Ferhatosmanoglu, & Demirbas, 2010). To address this issue, future research should apply a variety of sophisticated sentiment analysis methods to improve the performance of sentiment analysis.

5.3 FUTURE DIRECTIONS

There are many possibilities exist for future research. First, future work could extend the research to private firms. Prior studies find that there is less information asymmetry in private firms than in public firms because major investors can easily access internal information of private firms(Chen, Hope, Li, & Wang, 2011; Kim & Kwon, 2015). While major investors often manage private firms directly, lenders in the capital markets may not have the same access as major investors have. Hence, it is necessary to investigate how XBRL and social media affect the information asymmetry between private borrowers and lenders.

Second, future research could explore the impact of XBRL and social media on loan contracts in markets of other countries. Compared to US market, other markets have different culture and social structures that may result in various levels of information asymmetry. Further investigation on other markets would help practitioners to have a better understanding of how to use XBRL and social media to minimize information risk under different circumstances.

Finally, analytic tools used in this study can only extract attitude, and feelings from social media websites. A lot of useful information including concepts, keywords, relations, and social structures is excluded. Therefore, future research could apply more analytic approaches such as content analysis or social network analysis to further explore the effects of social media on capital markets.

5.4 CONCLUSION

The purpose of this study is to investigate whether two information technology advancements, the adoption of XBRL (eXtensible Business Reporting Language), and social media, affect bank loan contracting. Using a sample of 554 US bank loan contracts in 2011, I find that borrowers that adopt XBRL and/or receive more positive social media user opinion enjoy more favorable price and non-price terms of bank loan contracts. Additional analyses indicate that the relations between XBRL adoption and bank loan price, and relations between social media user opinion and bank loan price vary with the firm size, loan structure, and availability of public information of borrowers. Overall, this study provides empirical evidence that technology advancements, the adoption of XRBL and social media, reduce cost of bank loans by decreasing information asymmetry between borrowers and lenders.

REFERENCE

Aad, B., & Paul, S. (2008). Understanding XBRL Challenges for Software vendors A roadmap. *XBRL Nederland,* (1). http://www.sbr.gov.au/__data/assets/pdf_file/0020/7391/20080501-understanding-xbrl-a-challenge-for-software-vendors-a-roadmap-netherlands.pdf

Agrawal, A., & Chadha, S. (2005). Corporate governance and accounting scandals*. *Journal of law and economics, 48*(2), 371-406.

Albuquerque, R., & Hopenhayn, H. A. (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies, 71*(2), 285-315.

AlchemyAPI Inc. (2016). Sentiment Analysis API. Retrieved 03/15, 2016, from http://www.alchemyapi.com/products/alchemylanguage/sentiment-analysis

Altova. (2016). XBRL Mapping. Retrieved 05/02, 2016, from http://www.altova.com/mapforce/xbrl-mapping.html

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17*(2), 223-249.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259-1294.

Bae, K., & Goyal, V. K. (2009). Creditor rights, enforcement, and bank loans. *The Journal of Finance, 64*(2), 823-860.

Banerjee, A. V., & Newman, A. F. (1993). Occupational choice and the process of development. *Journal of political Economy*, 274-298.

Barclay, M. J., & Smith, C. W. (1995). The maturity structure of corporate debt. *The Journal of Finance, 50*(2), 609-631.

Beatty, A., Weber, J., & Yu, J. J. (2008). Conservatism and debt. *Journal of accounting and economics, 45*(2), 154-174.

Beaudry, P., & Poitevin, M. (1995). Competitive screening in financial markets when borrowers can recontract. *The Review of Economic Studies, 62*(3), 401-423.

Benmelech, E., Garmaise, M. J., & Moskowitz, T. (2004). Do liquidation values affect financial contracts? Evidence from commercial loan contracts and zoning regulation: National Bureau of Economic Research.

Berger, A. N., Espinosa-Vega, M. A., Frame, W. S., & Miller, N. H. (2011). Why do borrowers pledge collateral: New empirical evidence on the role of asymmetric information. *Journal of financial intermediation, 20*(1), 55-70.

Berger, A. N., & Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *Journal of Business*, 351-381.

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *Review of Financial Studies, 24*(4), 1141-1203.

Bharath, S. T., Sunder, J., & Sunder, S. V. (2008). Accounting quality and debt contracting. *The Accounting Review, 83*(1), 1-28.

Bhojraj, S., & Sengupta, P. (2003). Effect of corporate governance on bond ratings and yields: The role of institutional investors and outside directors*. *The Journal of Business, 76*(3), 455-475.

Blankespoor, E., Miller, G. S., & White, H. D. (2013). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *The Accounting Review, 89*(1), 79-112.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

Boot, A. W. (2000). Relationship banking: What do we know? *Journal of financial intermediation, 9*(1), 7-25.

Border, K. C., & Sobel, J. (1987). Samurai accountant: A theory of auditing and plunder. *The Review of Economic Studies, 54*(4), 525-540.

Boyd, J. H., & Smith, B. D. (1993). The equilibrium allocation of investment capital in the presence of adverse selection and costly state verification. *Economic Theory, 3*(3), 427-451.

Bushee, B. J., Core, J. E., Guay, W., & Hamm, S. J. (2010). The role of the business press as an information intermediary. *Journal of Accounting Research, 48*(1), 1-19.

Bushee, B. J., Matsumoto, D. A., & Miller, G. S. (2003). Open versus closed conference calls: the determinants and effects of broadening access to disclosure. *Journal of accounting and economics, 34*(1), 149-180.

Chen, F., Hope, O.-K., Li, Q., & Wang, X. (2011). Financial reporting quality and investment efficiency of private firms in emerging markets. *The Accounting Review, 86*(4), 1255-1288.

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

Chong, D., & Shi, H. (2015). Big data analytics: a literature review. *Journal of Management Analytics, 2*(3), 175-201.

Chong, D., & Zhu, H. H. (2012). *Firm Clustering based on Financial Statements.* Paper presented at the 22nd Workshop on Information Technology and Information Systems (WITS'12).

Cooley, T., Marimon, R., & Quadrini, V. (2004). Aggregate consequences of limited contract enforceability. *Journal of political Economy, 112*(4), 817-847.

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate reputation review, 10*(3), 163-176.

Core, J. E., Holthausen, R. W., & Larcker, D. F. (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics, 51*(3), 371-406.

Cuneo, E. (2003). XBRL: Still A Ways Away From Saving The Day'. *Information Week, viewed, 10*, 4-5.

Daily, C. M., & Dalton, D. R. (1994). Corporate governance and the bankrupt firm: An empirical assessment. *Strategic Management Journal, 15*(8), 643-654.

Das, S., Levine, C. B., & Sivaramakrishnan, K. (1998). Earnings predictability and bias in analysts' earnings forecasts. *Accounting Review*, 277-294.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science, 53*(9), 1375-1388.

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing, 26*(2), 83-91.

Debreceny, R. S., Chandra, A., Cheh, J. J., Guithues-Amrhein, D., Hannon, N. J., Hutchison, P. D., . . . Lymer, A. (2005). Financial Reporting in XBRL on the SEC's EDGAR System: A Critique and Evaluation. *Journal of Information Systems, 19*(2), 191-210.

Degryse, H., & Van Cayseele, P. (2000). Relationship lending within a bank-based system: Evidence from European small business data. *Journal of financial intermediation, 9*(1), 90-109.

Dennis, S., Nandy, D., & Sharpe, L. G. (2000). The determinants of contract terms in bank revolving credit agreements. *Journal of financial and quantitative analysis, 35*(01), 87-110.

Dennis, S. A., & Sharpe, I. G. (2005). Firm size dependence in the determinants of bank term loan maturity. *Journal of Business Finance & Accounting, 32*(1 - 2), 31-64.

Dhillon, I. S. (2001). *Co-clustering documents and words using bipartite spectral graph partitioning*. Paper presented at the the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.

Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine learning, 42*(1), 143-175.

Diamond, D. W. (1989). Reputation acquisition in debt markets. *The Journal of Political Economy*, 828-862.

Diamond, D. W. (1991a). Debt maturity structure and liquidity risk. *The Quarterly Journal of Economics*, 709-737.

Diamond, D. W. (1991b). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of political Economy*, 689-721.

Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. *The Journal of Finance, 46*(4), 1325-1359.

Duarte, F. D. (2011). *The role of collateral and relationship lending in loan pricing: evidence from United Kingdom SMEs.*

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

Engel, P., Hamscher, W., Advantage, S., Shuetrim, G., vun Kannon, D., & Pryde, C. (2008). Extensible Business Reporting Language (XBRL) 2.1. *Jul, 2*, 1-165.

Farewell, S. M. (2006). An introduction to XBRL through the use of research and technical assignments. *Journal of Information Systems, 20*(1), 161-185.

Flannery, M. J. (1986). Asymmetric information and risky debt maturity choice. *The Journal of Finance, 41*(1), 19-37.

Frankel, R., & Li, X. (2004). Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders. *Journal of accounting and economics, 37*(2), 229-259.

Gale, D., & Hellwig, M. (1985). Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies, 52*(4), 647-663.

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*(1), 7.

Graham, J. R., Li, S., & Qiu, J. (2008). Corporate misreporting and bank loan contracting. *Journal of Financial Economics, 89*(1), 44-61.

Harhoff, D., & Körting, T. (1998). Lending relationships in Germany–Empirical evidence from survey data. *Journal of Banking & Finance, 22*(10), 1317-1353.

Hasan, I., Park, J. C., & Wu, Q. (2012). The impact of earnings predictability on bank loan contracting. *Journal of Business Finance & Accounting, 39*(7 ‐ 8), 1068-1101.

Haselmann, R., Pistor, K., & Vig, V. (2010). How law affects lending. *Review of Financial Studies, 23*(2), 549-580.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior research methods, 41*(3), 924-936.

He, W., Tian, X., Chen, Y., & Chong, D. (2016). Actionable Social Media Competitive Analytics For Understanding Customer Experiences. *Journal of Computer Information Systems, 56*(2), 145-155.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464-472.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of accounting and economics, 31*(1), 405-440.

Heinle, M. S., & Verrecchia, R. E. (2015). Bias and the Commitment to Disclosure. *Management Science*.

Hogenboom, A., Bal, M., Frasincar, F., & Bal, D. (2013). *Towards cross-language sentiment analysis through universal star ratings.* Paper presented at the 7th International Conference on Knowledge Management in Organizations: Service and Cloud Computing.

Hölmstrom, B. (1979). Moral hazard and observability. *The Bell journal of economics*, 74-91.

Holmstrom, B., & Tirole, J. (1997). Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics*, 663-691.

Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management, 9*(3), 201-214.

IBM. (2016). Sentiment analysis with AlchemyAPI: A hybrid approach. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&htmlfid=LBW03035USEN&attachment=LBW03035USEN.PDF&cm_mmc=Email-_-IBM+Watson_Watson+Developer+Cloud-_-WW_WW-_-Sentiment+Analysis+WP_ov43093&cm_mmca1=000000OF&cm_mmca2=10000409&cm_mmca3=M00000597

Kaplan, A., & Haenlein, M. (2009). Consumers, companies, and virtual social worlds: A qualitative analysis of Second Life. *Advances in consumer research, 36*(1), 873-874.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Khan, T. (2006). *Financial reporting disclosure on the internet: An international perspective.* Victoria University.

Kim, J. C. (1985). The market for" lemons" reconsidered: A model of the used car market with asymmetric information. *The American economic review, 75*(4), 836-843.

Kim, J. W., Kim, J., & No, W. G. (2011). *The effects of XBRL disclosures on information environment in the market: Early evidence.* Paper presented at the 2011 AAA Annual Meeting.

Kim, K., & Kwon, O. (2015). The Investment Efficiency Of Private And Public Firms: Evidence From Korea. *The Journal of Applied Business Research, 31*(4), 1387-1402.

Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. *Journal of accounting and economics, 33*(3), 375-400.

Klock, M. S., Mansi, S. A., & Maxwell, W. F. (2005). Does corporate governance matter to bondholders? *Journal of financial and quantitative analysis, 40*(04), 693-719.

Kothari, S., 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.

Lacker, J. M. (1998). Collateralized debt as the optimal contract.

Laroche, M., Habibi, M. R., Richard, M.-O., & Sankaranarayanan, R. (2012). The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior, 28*(5), 1755-1767.

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: evidence from consumer product recalls. *Journal of Accounting Research, 53*(2), 367-404.

Lee, T. S., & Yeh, Y. H. (2004). Corporate governance and financial distress: evidence from Taiwan. *Corporate Governance: An International Review, 12*(3), 378-388.

Leverenz, L. (2016). Hive design and architecture. Retrieved 04/20, 2016, from https://cwiki.apache.org/confluence/display/Hive/Design#Design-HiveArchitecture

Lin, C.-Y., Chen, Y.-S., & Yen, J.-F. (2014). On the determinant of bank loan contracts: The roles of borrowers’ ownership and board structures. *The Quarterly Review of Economics and Finance, 54*(4), 500-512. doi: http://dx.doi.org/10.1016/j.qref.2014.04.005

Liu, P., Seyyed, F. J., & Smith, S. D. (1999). The independent impact of credit rating changes – the case of Moody's rating refinement on yield premiums. *Journal of Business Finance & Accounting, 26*(3 - 4), 337-363.

Maina, A. (2015). Why Social Media Can Help (or Hurt) For a Small Business Loan. Retrieved 03/12, 2016, from http://smallbiztrends.com/2015/12/social-media-can-impact-business-credit-score.html

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

Miller, G. S. (2006). The press as a watchdog for accounting fraud. *Journal of Accounting Research, 44*(5), 1001-1033.

Miller, G. S., & Skinner, D. J. (2015). The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. *Journal of Accounting Research, 53*(2), 221-239.

Nann, S., Krauss, J., & Schoder, D. (2013). *Predictive Analytics On Public Data-The Case Of Stock Markets.* Paper presented at the ECIS.

Nutz, A., & Strauß, M. (2002). eXtensible Business Reporting Language (XBRL). *Wirtschaftsinformatik, 44*(5), 447-457.

Ortiz-Molina, H., & Penas, M. F. (2008). Lending to small businesses: The role of loan maturity in addressing information problems. *Small Business Economics, 30*(4), 361-383.

Pennacchi, G. G. (1988). Loan sales and the cost of bank capital. *The Journal of Finance, 43*(2), 375-396.

Petersen, M. A., & Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance*, *49*(1), 3-37.

Pinsker, R., & Li, S. (2008). Costs and benefits of XBRL adoption: Early evidence. *Communications of the ACM*, *51*(3), 47-50.

Qian, J., & Strahan, P. E. (2007). How laws and institutions shape financial contracts: The case of bank loans. *The Journal of Finance, 62*(6), 2803-2834.

Rajan, R., & Winton, A. (1995). Covenants and collateral as incentives to monitor. *The Journal of Finance, 50*(4), 1113-1146.

Rezaee, Z., & Turner, J. L. (2002). XBRL-based financial reporting: Challenges and opportunities for government accountants. *The Journal of Government Financial Management, 51*(2), 16.

Richardson, S. (2006). Over-investment of free cash flow. *Review of Accounting Studies, 11*(2-3), 159-189.

Roohani, S. J. (2003). Trust and data assurances in capital markets: The role of technology solutions. *PriceWaterhouseCoopers research monograph, ed. DR. SAREED J. ROOHANI, Bryant College, RI, 2917*.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review, 1*(1), 27-64.

Securities Exchange Commission. (2009). Interactive Data to Improve Financial Reporting Retrieved 06/14/2015, from http://www.sec.gov/rules/final/2009/33-9002.pdf

Seeking Alpha. (2016a). About Seeking Alpha. Retrieved 04/10, 2016, from http://seekingalpha.com/page/about_us

Seeking Alpha. (2016b). The analysis of Microsoft Corporation. Retrieved 4/26, 2016, from http://seekingalpha.com/symbol/MSFT/focus

Shen, J.-C., & Reuer, J. J. (2005). Adverse selection in acquisitions of small manufacturing firms: A comparison of private and public targets. *Small Business Economics, 24*(4), 393-407.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(8), 888-905.

Simons, K. (1993). Why do banks syndicate loans? *New England Economic Review*(Jan), 45-52.

Soltes, E. (2010). Disseminating firm disclosures. *Unpublished Working Papers. Harvard Business School*.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., & Demirbas, M. (2010). *Short text classification in twitter to improve information filtering.* Paper presented at the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.

Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review, 71*(3), 393-410.

Strahan, P. E. (1999). Borrower risk and the price and nonprice terms of bank loans. *FRB of New York staff report*(90).

Stulz, R., & Johnson, H. (1985). An analysis of secured debt. *Journal of Financial Economics, 14*(4), 501-521.

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance, 62*(2), 629-668.

The R Foundation. (2016). Introduction to R. Retrieved 04/01, 2016, from https://www.r-project.org/about.html

Thomson Reuters. (2016). Discover Thomson Reuters. Retrieved 04/02, 2016, from http://banker.thomsonib.com/

Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Anthony, S., . . . Murthy, R. (2009). Hive: a warehousing solution over a map-reduce framework. *Proceedings of the VLDB Endowment, 2*(2), 1626-1629.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic theory, 21*(2), 265-293.

Trainor, K. J., Andzulis, J. M., Rapp, A., & Agnihotri, R. (2014). Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research, 67*(6), 1201-1208.

Twitter. (2016). Twitter Documentation for Developers. Retrieved 04/15, 2016, from https://dev.twitter.com/overview/documentation

US Census Bureau. (2012). Conceptual issues of economic classification. Retrieved 02-20, 2012, from http://www.census.gov/eos/www/naics/history/docs/issue_paper_1.pdf

Vasarhelyi, M., Chan, D., & Krahel, J. (2010). XBRL Consequences to Financial Reporting, Data Analysis, Decision Support, and Others: Rutgers University.

Voorhees, E. M. (1986). The effectiveness and efficiency of agglomerative hierarchic clustering in document retrieval. *Dissertation Abstracts International Part B: Science and Engineering[DISS. ABST. INT. PT. B- SCI. & ENG.], 47*(2).

Whette, H. (1983). Collateral in credit rationing in markets with imperfect information. *American Economic Review, 73*(3), 442-445.

Wikipedia. (2016). Yahoo Finance. Retrieved 04/02, 2016, from https://en.wikipedia.org/wiki/Yahoo!_Finance

Williamson, S. D. (1986). Costly monitoring, financial intermediation, and equilibrium credit rationing. *Journal of Monetary Economics, 18*(2), 159-179.

Wu, J., & Vasarhelyi, M. (2004). XBRL: A new tool for electronic financial reporting *Business Intelligence Techniques* (pp. 73-92): Springer.

Wu, Q., Francis, B., Hasan, I., & Koetter, M. (2011). Corporate Boards and Bank Loan Contracting. *Journal of Financial Research, Forthcoming*.

XBRL International. (2006). Extensible Business Reporting Language (XBRL) 2.1: XBRL International.

Yahoo Finance. (2016). Microsoft Corporation Message Board. Retrieved 04/21, 2016, from http://finance.yahoo.com/mb/MSFT/

Yi, M., Oh, S. G., & Kim, S. (2013). Comparison of social media use for the US and the Korean governments. *Government Information Quarterly, 30*(3), 310-317.

Zha, H., He, X., Ding, C., Simon, H., & Gu, M. (2001). *Bipartite graph partitioning and data clustering.* Paper presented at the the tenth international conference on Information and knowledge management, New York, USA.

Zhang, J. (2008). The contracting benefits of accounting conservatism to lenders and borrowers. *Journal of accounting and economics, 45*(1), 27-54.

Zhu, H., & Wu, H. (2011a). Interoperability of XBRL Financial Statements in the U.S. *International Journal of E-Business Research, 7*(2), 18-33.

Zhu, H., & Wu, H. (2011b). Quality of data standards: framework and illustration using XBRL taxonomy and instances. *Electronic Markets, 21*(2), 129-139.

Zhu, H., & Wu, H. (2014). Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. *Decision Support Systems, 59*, 351-360.

Zhu, H. H., & Wu, H. (2010). Quality of XBRL US GAAP taxonomy: Empirical evaluation using SEC filings. *Harris, Quality of XBRL US GAAP Taxonomy: Empirical Evaluation Using SEC Filings (March 3, 2010).*

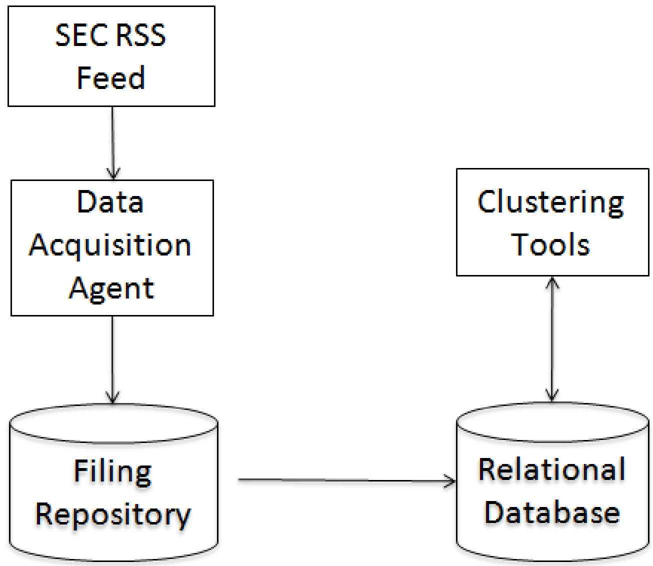

Figure 1 Firm Clustering System

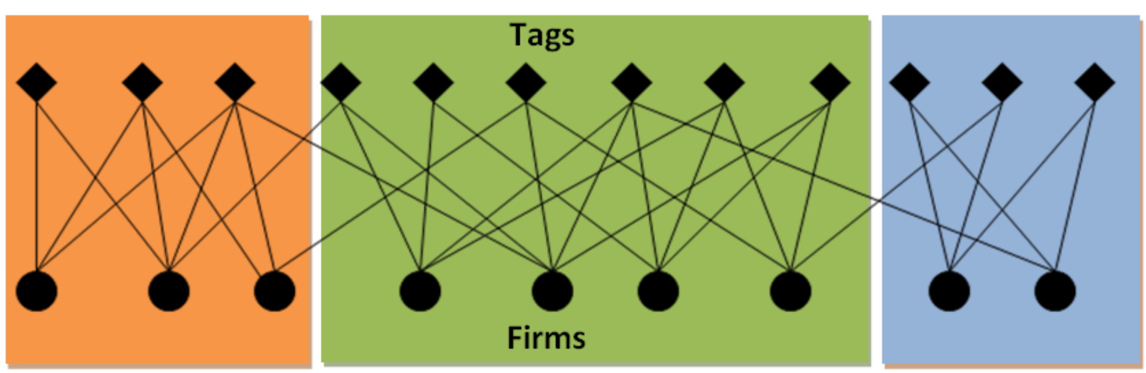

Figure 2 Clustering Firms Based on the Tags Firms Used

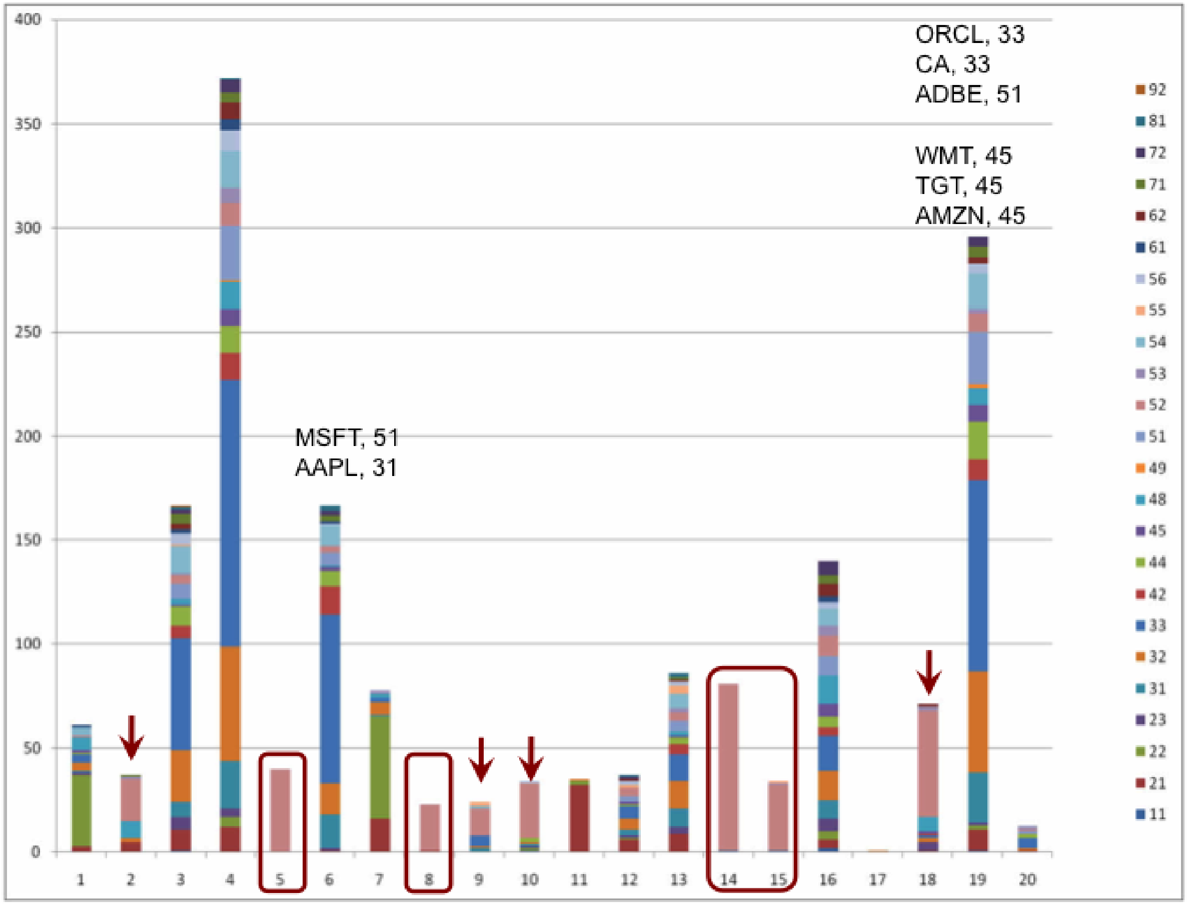

Figure 3 Distribution of Major Industry in 20 Clusters

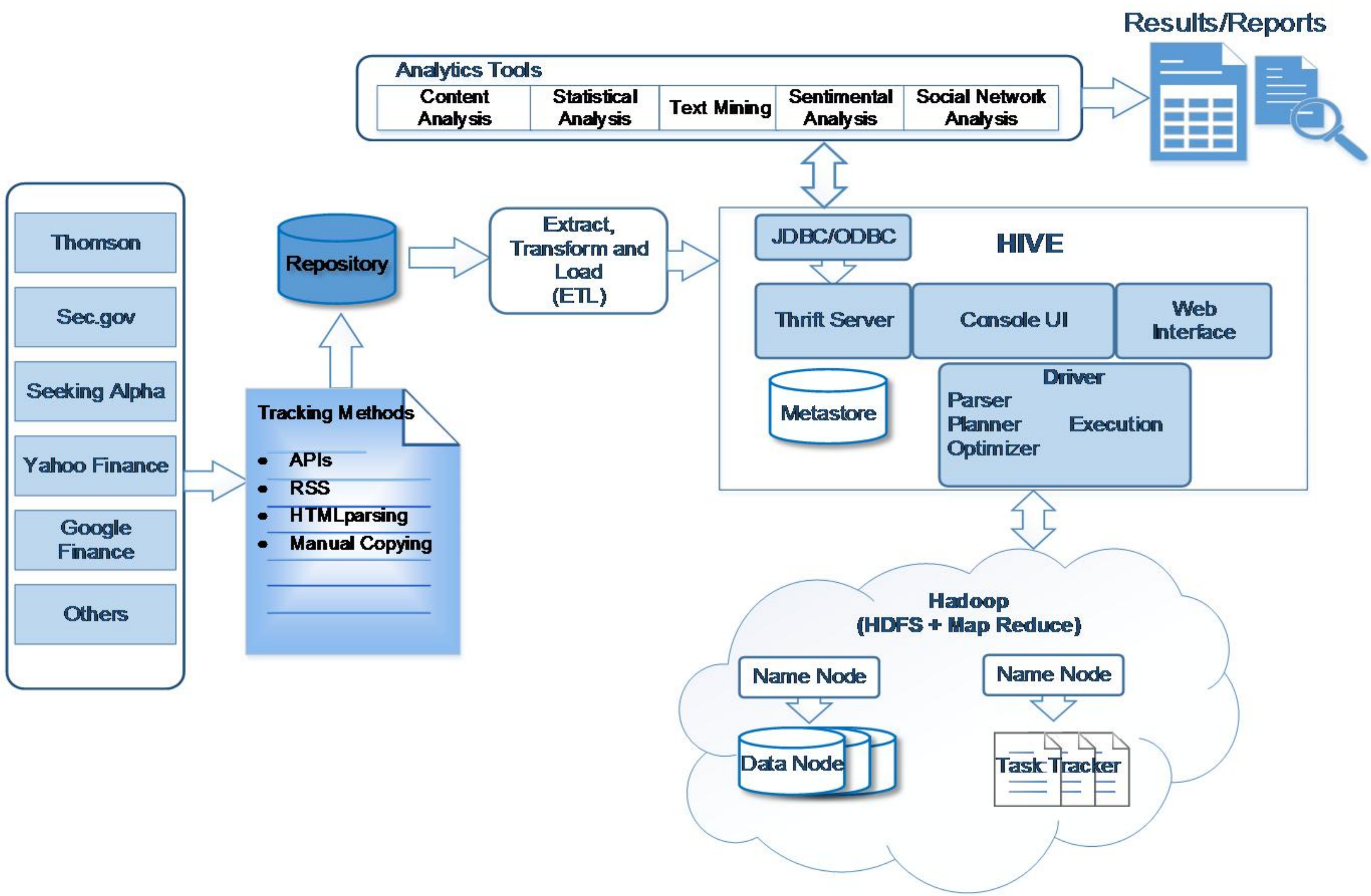

Figure 4 Framework of Bank Loan Pricing Analytics System

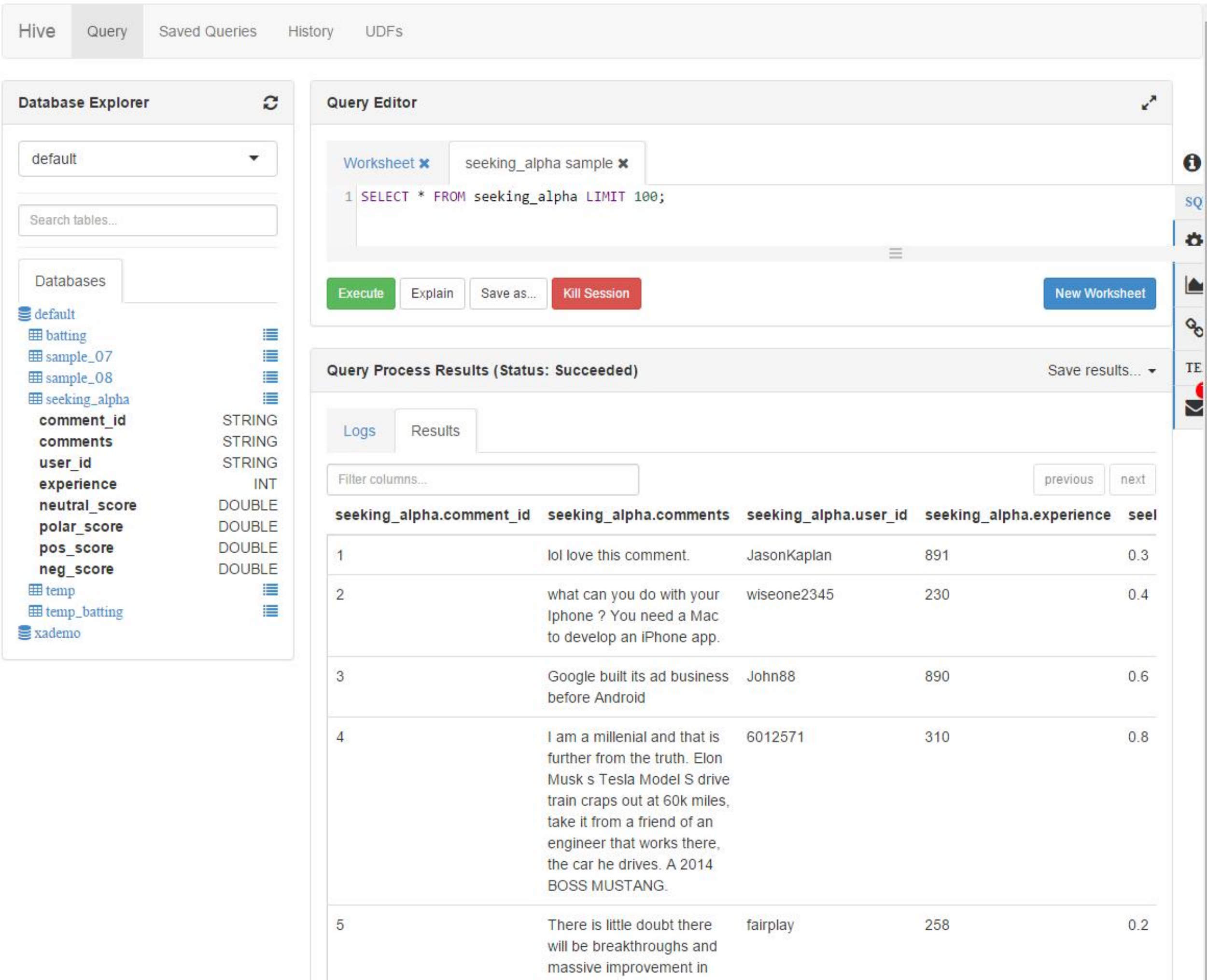

Figure 5 User Interface of Hive Based Distributed Storage Platform

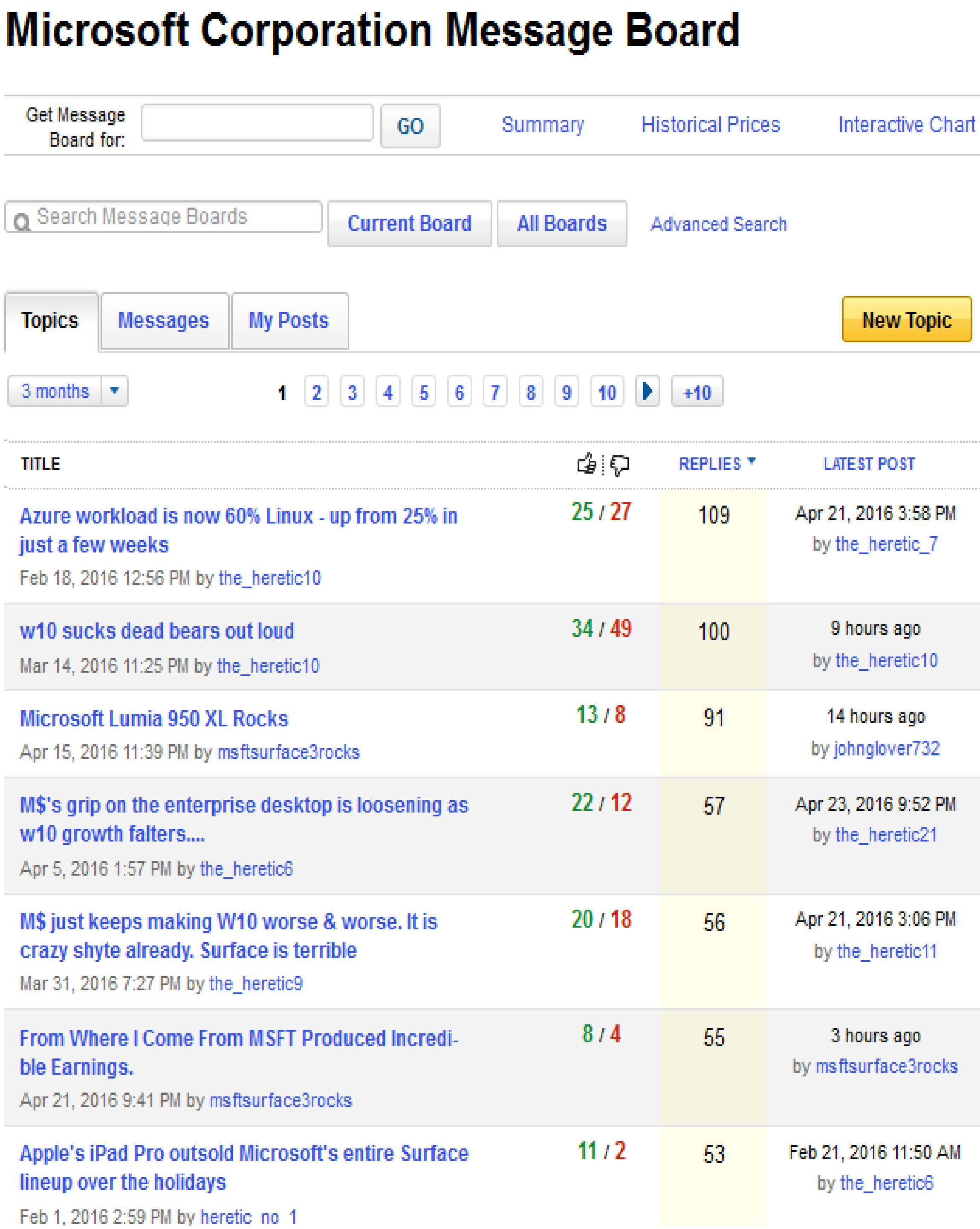

Figure 6 Message Board of Microsoft on Yahoo Finance

(Yahoo Finance, 2016)

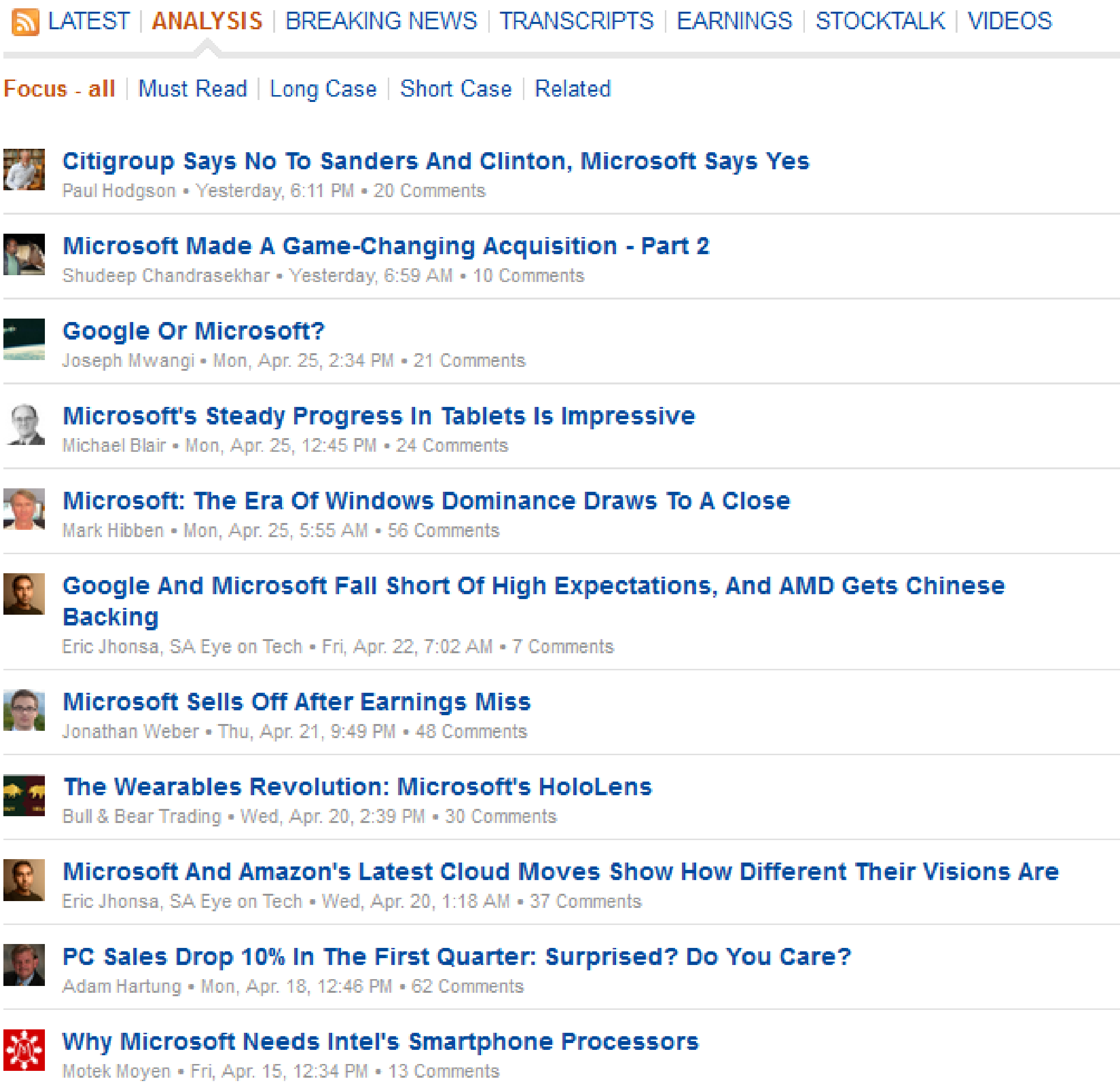

Figure 7 Sample Articles and Comments of Microsoft on Seeking Alpha

(Seeking Alpha, 2016b)

Table 1 Example of XBRL Specification

Name	Definition	Examples
Simple link	A link that points from one resource to another. It points to Linkbases from XBRL Instances and from Taxonomy Schemas or points to Taxonomy Schemas from XBRL Instance	<complexContent> <restriction base="anyType"> <attributeGroup ref="xlink:simpleType"/> <attribute ref="xlink:href" use="required"/> <attribute ref="xlink:arcrole" use="optional"/> <attribute ref="xlink:role" use="optional"/> <anyAttribute namespace="http://www.w3.org/XML/1998/namespace" processContents="lax"/> </restriction> </complexContent>
The "schemaRef" element in XBRL Instances	Every XBRL instance must contain at least one "schemaRef" element. It points to a Taxonomy Schema that becomes part of the DTS supporting that XBRL instance.	<element name="schemaRef" type="xl:simpleType" substitutionGroup="xl:simple"> <annotation> <documentation> Definition of the schemaRef element - used to link to XBRL taxonomy schemas from XBRL instances. </documentation> </annotation> </element>

Notes: Retrieved from http://www.xbrl.org/Specification/XBRL-2.1/REC-2003-12-31/XBRL-2.1-REC-2003-12-31+corrected-errata-2013-02-20.html#_3.5.1

Table 2 Example of Taxonomy Schemas

Name	Data Type	Definition
Interest Receivable	monetary	<xs:element id="us-gaap_InterestReceivable" name="InterestReceivable" nillable="true" substitutionGroup="xbrli:item" type="xbrli:monetaryItemType" xbrli:balance="debit" xbrli:periodType="instant"/>
InterestBearing ForeignDeposit MoneyMarket	monetary	<xs:element id="us-gaap_InterestBearingForeignDepositMoneyMarket" name="InterestBearingForeignDepositMoneyMarket" nillable="true" substitutionGroup="xbrli:item" type="xbrli:monetaryItemType" xbrli:balance="credit" xbrli:periodType="instant"/>
InterestBearing DepositLiabilities ByComponent Abstract	string	<xs:element abstract="true" id="us-gaap_InterestBearingDepositLiabilitiesByComponentAbstract" name="InterestBearingDepositLiabilitiesByComponentAbstract" nillable="true" substitutionGroup="xbrli:item" type="xbrli:stringItemType" xbrli:periodType="duration"/>
Investment OwnedValued ByTrusteesFlag	boolean	<xs:element id="us-gaap_InvestmentOwnedValuedByTrusteesFlag1" name="InvestmentOwnedValuedByTrusteesFlag1" nillable="true" substitutionGroup="xbrli:item" type="xbrli:booleanItemType" xbrli:periodType="duration"/>

Notes: Retrieved from http://xbrl.fasb.org/us-gaap/2016/elts/us-gaap-2016-01-31.xsd

Table 3 Example of Taxonomy Linkbases

Element	Linked Elements	Weight	Definition
IncomeTaxesPaidRefund	IncomeTaxesPaidRefundClassifiedAsInvestingActivities	1	<link:calculationArc xlink:type="arc" xlink:arcrole="http://www.xbrl.org/2003/arcrole/summation-item" xlink:from="IncomeTaxesPaidRefund" xlink:to="IncomeTaxesPaidRefundClassifiedAsInvestingActivities" xlink:title="calculation: IncomeTaxesPaidRefund to IncomeTaxesPaidRefundClassifiedAsInvestingActivities" order="1.0" weight="1.0"/>
IncomeTaxesPaidRefund	IncomeTaxesPaidRefundClassifiedAsOperatingActivities	1	<link:calculationArc xlink:type="arc" xlink:arcrole="http://www.xbrl.org/2003/arcrole/summation-item" xlink:from="IncomeTaxesPaidRefund" xlink:to="IncomeTaxesPaidRefundClassifiedAsOperatingActivities" xlink:title="calculation: IncomeTaxesPaidRefund to IncomeTaxesPaidRefundClassifiedAsOperatingActivities" order="2.0" weight="1.0"/>

Notes: Retrieved from http://media.corporate-ir.net/media_files/irol/76/76540/xbrl/2013//trito-20131231_cal.xml

Table 4 Example of XBRL Instance Documents

Element	Context	Value	Definition
CashFlowsBefore ChangesWorking Capital	Current_ ForPeriod	57400	<iascf-pfs:CashFlowsBeforeChangesWorkingCapital numericContext="Current_ForPeriod"> 574000 </iascf-pfs:CashFlowsBeforeChangesWorkingCapital>
CashFlowsBefore ChangesWorking Capital	Prior_ ForPeriod	442000	<iascf-pfs:CashFlowsBeforeChangesWorkingCapital numericContext="Prior_ForPeriod"> 442000 </iascf-pfs:CashFlowsBeforeChangesWorkingCapital>

Notes: Retrieved from http://www.xbrl.org/taxonomy/int/fr/ias/ci/pfs/2002-11-15/SampleCompany-2002-11-15.xml

Table 5 Example of Firm-Tag Matrix

	$Firm_1$	$Firm_2$	...	$Firm_n$
Tag_1	0	1	…	1
Tag_2	1	1	…	1
...	…	…	…	…
Tag_n	1	0	…	1

Table 6 Evaluation of the Number of Clusters

Cluster number	Local density	Relative density	Total density
2	0.0171	0.7596	0.0116
3	0.0409	0.6735	0.0237
4	0.0309	0.5052	0.0179
5	0.0278	0.4755	0.0158
6	0.0279	0.4289	0.0150
7	0.0301	0.3806	0.0136
8	0.0344	0.3550	0.0152
9	0.0444	0.3696	0.0208
10	0.0495	0.3587	0.0215
11	0.0500	0.3417	0.0214
12	0.0510	0.3214	0.0209
13	0.0491	0.3052	0.0193
14	0.0670	0.3009	0.0258
15	0.0648	0.2870	0.0245
16	0.0666	0.2907	0.0250
17	0.0737	0.2899	0.0260
18	0.0730	0.2786	0.0251
19	0.0754	0.2586	0.0237
20	**0.1002**	**0.2540**	**0.0306**
21	0.0909	0.2470	0.0293
22	0.0983	0.2325	0.0289
23	0.0939	0.2215	0.0269
24	0.0894	0.2117	0.0255
25	0.0911	0.2097	0.0258
26	0.0904	0.2002	0.0246
27	0.0872	0.1908	0.0228
28	0.0901	0.1889	0.0228
29	0.1186	0.1834	0.0257
30	0.1186	0.1846	0.0263

Table 7 Firm Distribution among Industry Groups According to NAICS

NAICS	INDUSTRY	# Firms
11	Agriculture, Forestry, Fishing and Hunting	5
21	Mining, Quarrying, and Oil and Gas Extraction	110
22	Utilities	98
23	Construction	28
31	Manufacturing (Good, Beverage and Tobacco Product, Apparel and Leather and Allied Product Manufacturing, Textile, Textile Product Mills)	96
32	Manufacturing (Wood Product, Paper, Petroleum and Coal Products, Chemical, Plastics and Rubber Products and Nonmetallic Mineral Product Manufacturing, Printing and Related Support Activities)	195
33	Manufacturing (Primary Metal, Fabricated Metal Product, Machinery, Computer and Electronic Product, Electrical Equipment, Appliance, and Component , Transportation Equipment , Furniture and Related Product and Miscellaneous Manufacturing)	410
42	Wholesale Trade	54
44	Retail Trade (Motor Vehicle and Parts Dealers, Furniture and Home Furnishings Stores, Electronics and Appliance Stores, Building Material and Garden Equipment and Supplies Dealers, Food and Beverage Stores, Health and Personal Care Stores, Gasoline Stations, Clothing and Clothing Accessories Stores,	61
45	Retail Trade (Sporting Goods, Hobby, Book, and Music Stores, General Merchandise Stores, Miscellaneous Store Retailers, Nonstore Retailers)	29
48	Transportation and Warehousing(Air, Rail, Water, Truck , Transit and Ground Passenger , Pipeline, Scenic and Sightseeing and Support Activities for Transportation)	64
49	Transportation and Warehousing(Postal Service, Couriers and Messengers, Warehousing and Storage)	3
51	Information	82
52	Finance and Insurance	330
53	Real Estate and Rental and Leasing	23
54	Professional, Scientific, and Technical Services	80
55	Management of Companies and Enterprises	9
56	Administrative and Support and Waste Management and Remediation Services	31
61	Educational Services	12
62	Health Care and Social Assistance	23
71	Arts, Entertainment, and Recreation	24
72	Accommodation and Food Services	23
81	Other Services (except Public Administration)	8
92	Public Administration	1
Total		1799

Table 8 Frequent Elements in Cluster 1, Cluster 2, Cluster 5, and Cluster 6

Cluster	Frequently Used Elements
1	AccumulatedDeferredInvestmentTaxCredit,PublicUtilitiesDisclosureTextBlock,RegulatoryAssetsCurrent RegulatoryLiabilityCurrent, RegulatoryLiabilities, AdditionalCollateralAggregateFairValue, RegulatoryAssets, ScheduleOfJointlyOwnedUtilityPlantsTextBlock, JointlyOwnedUtilityPlantProportionateOwnershipShare, JointlyOwnedUtilityPlantOwnershipAmountOfPlantAccumulatedDepreciation. DebtInstrumentUnamortizedDiscountPremiumNet, PublicUtilitiesPolicyTextBlock
2	PartnersCapital, GeneralPartnersCapitalAccount, LimitedPartnersCapitalAccount, PartnersCapitalAccountDistributions, LimitedPartnersCapitalAccountUnitsOutstanding NetIncomeLossAllocatedToLimitedPartners , PartnersCapitalNotesDisclosureTextBlock, LiabilitiesAndPartnersCapital, NetIncomeLossPerOutstandingLimitedPartnershipUnit NetIncomeLossAllocatedToGeneralPartners , WeightedAverageLimitedPartnershipUnitsOutstanding
5	IncreaseDecreaseInUnearnedPremiums, IncreaseDecreaseInPremiumsReceivable, PrepaidReinsurancePremiums, ReinsurancePayable IncreaseDecreaseInDeferredPolicyAcquisitionCosts, NetInvestmentIncome SupplementalScheduleOfReinsurancePremiumsForInsuranceCompaniesTextBlock， SupplementaryInsuranceInformationForInsuranceCompaniesDisclosureTextBlock， IncreaseDecreaseInReinsuranceRecoverable , PremiumsReceivableAtCarryingValue
6	ScheduleOfProductWarrantyLiabilityTableTextBlock, FutureAmortizationExpenseAfterYearFive ScheduleOfAccruedLiabilitiesTableTextBlock, ScheduleOfDebtTableTextBlock BusinessAcquisitionsProFormaRevenue, LiabilitiesFairValueDisclosure BusinessAcquisitionsProFormaNetIncomeLoss, ShareBasedCompensationArrangementByShareBasedPayme. ShareBasedCompensationArrangementByShareBasedPayme, StandardProductWarrantyPolicy

Table 9 Results of Sample Selection

Stages	# of Firms	# of loans
Number of transactions in the Thompson one database	3835	5905
Exclude private and non-US companies	(3277)	(5149)
Exclude records without available LIBOR price	(55)	(80)
Exclude borrowers without submission of 10-k from 2009 to 2011	(45)	(88)
Exclude borrowers without posts on Yahoo message board and Seeking Alpha	(32)	(34)
Total number of samples	426	554

Table 10 Summary Statistics

Variable	Min	Median	Mean	Max	Std. Dev.
Loan Spread	5	175	199.621	1,150	123.194
XBRL	0	1	0.894	2	0.759
Social Media(Seeking Alpha)	-0.706	0.221	0.217	1.200	0.334
Social Media(Yahoo)	-1.161	-0.002	0.026	1.350	0.307
Log(Asset)	7.500	9.439	9.458	11.860	0.720
Leverage	0	0.260	0.280	1.450	0.202
Current Ratio	0.250	1.600	1.915	8.770	1.141
Cash to Debt Ratio	0	0.221	88.471	42,040.67	1.853
Profitability	-0.260	0.121	0.132	0.682	0.083
Interest Coverage	-23.281	5.174	31.839	1554.330	123.508
Tangibility	0	0.190	0.289	0.940	0.266
M/B	0.790	1.400	1.705	8.090	0.968
zscore	-10.600	1.779	1.774	6.274	1.436
Loan Characteristics					
Loan Size(M)	5	350	700.323	15,000	1,155
Maturity	3	60.120	53.262	84.720	14.601
Secured	0	0	0.128	1	0.335
Prior Relations	0	1	0.685	1	0.465

Table 11 Pearson Correlation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Spread	1.000													
2 Social Media	-.116 (.046)	1.000												
3 XBRL	-0.617 (.000)	.061 (.187)	1.000											
4 Asset	-0.607 (.000)	.011 (.438)	.809 (.000)	1.000										
5 Leverage	0.2 (.002)	-.024 (.364)	-.044 (.261)	.071 (.154)	1.000									
6 Current Ratio	0.112 (.05)	-.097 (.080)	-.150 (.015)	-.294 (.000)	-.249 (.000)	1.000								
7 Cash to Debt Ratio	0.0151 (.414)	.069 (.160)	-.110 (.055)	-.119 (.042)	-.105 (.064)	.138 (.023)	1.000							
8 Profitability	-0.291 (.000)	.025 (.359)	.078 (.128)	-.074 (.142)	.000 (.499)	.110 (.056)	.000 (.497)	1.000						

Table 11 (Continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
9 Intercoverage	.106 (.062)	.025 (.360)	-.121 (.040)	-.136 (.024)	-.231 (.000)	.254 (.000)	.301 (.000)	.143 (.019)	1.000					
10 Tangibility	-.034 (.311)	-.017 (.403)	.019 (.393)	.078 (.129)	.104 (.066)	-.270 (.000)	-.046 (.255)	.001 (.493)	-.061 (.188)	1.000				
11 MB	-.154 (.012)	.055 (.215)	-.088 (.101)	-.265 (.000)	-.044 (.260)	.210 (.001)	.090 (.096)	.647 (.000)	.108 (.059)	-.161 (.010)	1.000			
12 zscore	-.260 (.000)	.076 (.135)	-.023 (.368)	-.053 (.220)	-.259 (.000)	.093 (.090)	.008 (.455)	.365 (.000)	.147 (.016)	-.066 (.171)	.148 .(016)	1.000		
13 Loan Size	-.448 (.000)	-.035 (.305)	.416 (.000)	.590 (.000)	-.062 (.186)	-.152 (.014)	-.039 (.287)	.012 (.432)	-.024 (.367)	.001 (.493)	-.075 (.138)	-.003 (.482)	1.000	
14 Prior Relations	-.371 (.000)	-.005 (.473)	.401 (.000)	.472 (.000)	-.004 (.479)	-.227 (.000)	-.116 (.047)	-.011 (.437)	-.185 (.004)	.060 (.193)	-.252 (.000)	.102 (.040)	.242 (.000)	1.000

Notes: This table provides Spearman correlations of the main variables employed in our analysis. P-values are also reported.

Table 12 XBRL Adoption, Social Media Sentiment, and Bank Loan Price

Dependent variable:	Log(spread)	Log(spread)	Log(spread)	Log(spread)
	(1)	(2)	(3)	(4)
XBRL	-.132***			-.125*
	(-3.242)			(-1.806)
Social Media (Yahoo)		-.047		
		(-.701)		
Social Media (Seeking alpha)			-.157**	.204
			(-1.894)	(1.224)
XBRL*Social Media(Seeking alpha)				-.284**
				(-2.337)
Log(asset)	-.185***	-.315***	-.412***	-.251***
	(-3.416)	(-7.284)	(-6.997)	(-2.794)
Leverage	.407***	.556***	.573***	.525***
	(3.609)	(4.464)	(3.149)	(2.934)
Current Ratio	.030*	.019	.007	.010
	(1.670)	(.929)	(.213)	(.335)
Cash to Debt Ratio	-.00001	-.00003***	-.00001	-.00001
	(-.769)	(-2.725)	(-.743)	(-1.275)
Profitability	-1.270***	-1.654***	-1.472**	-1.470***
	(-3.460)	(-3.633)	(-2.570)	(-2.571)
Interest Coverage	.0003	.002***	.0003*	.0003*
	(1.494)	(3.777)	(1.582)	(1.553)
Tangibility	-.104	.046	-.218	-.269*
	(-1.193)	(.447)	(-1.490)	(-1.867)
MB	-.088***	-.120***	-.139***	-.135***
	(-3.301)	(-3.785)	(-3.310)	(-3.280)
zscore	-.068***	-.055***	-.051**	-.052**
	(-3.869)	(-2.903)	(-2.089)	(-2.180)
Loan Size	.0001***	.0001***	-.0001***	.0001***
	(-5.220)	(-4.755)	(-3.606)	(-4.211)
Prior Relations	-.045	-.061	-.047	-.041
	(-1.089)	(-1.313)	(-.638)	(-.572)
Control For				

Table 12 (Continued)

Industry Effect	Y	Y	Y	Y
Loan-Type effect	Y	Y	Y	Y
Loan-Purpose Effect	Y	Y	Y	Y
Observations	411	321	211	211
Adjusted R-Squared	0.632	0.618	0.667	0.684

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 13 Paired Sample Test

	Mean	Std. Deviation	t	df	Sig.
Seeking Alpha - Yahoo	0.260	0.461	8.082	205	.000

Table 14 Examples of Yahoo Postings and Seeking Alpha Postings

Yahoo postings	Seeking Alpha postings
Me 4! Mikey D's is the best performing stock in my portfolio.	Thanks for the analysis on MCD, DGI. This is one of my favorite companies. Wonderful product, excellent recognition, and who really thinks McDonald's won't be around in 50 years? 100 years? I keep wanting to pick more MCD up on dips, but it just seems to dip much less than others on my watchlist, so it is still a very small position for me. I hope to change that on the next dip.
Dont fight the trend. The trend is your friend.	As always, first class article. I am long MCD. Bought my first 100 shares back in 1988. I wish I would have kept those, but I was a "trader" back in the day. There is no telling what my yield on cost for those shares would be today. (I KNOW there is, I'm just to lazy to look it up and I don't want to have to kick the crap out of myself
HOMEMADE HAMBURGER RISING NOW!	If you bought MCD, you must know something! Thanks for the heads up on the dividend increase. I'm sure my granddaughter will have it by then and maybe I will too
summertime is for beer not coffee.	I consider MCD to be more of a growth stock than a dividend stock. It has taken me a long time to compromise my yield on this stock down to 3%, but I never could get in. The same thing happened when I started investing in PG many years ago.

Table 15 Robustness Check- Control for Clusters

Dependent Variable	Log(spread)	Log(spread)	Log(spread)
	(1)	(2)	(3)
XBRL	-.160***		-.032*
	(-2.871)		(-1.233)
Social Media		-.231**	.022
		(-2.452)	(.856)
XBRL*Social Media			-.162***
			(-2.573)
Log(asset)	-.135*	-.359***	-.229***
	(-1.795)	(-3.851)	(-2.213)
Leverage	.452***	.634***	.854***
	(2.791)	(2.862)	(1.032)
Current Ratio	.032	-.008	-.001
	(1.306)	(-.209)	(-.01)
Cash to Debt Ratio	-.0001	.018**	.008***
	(-.156)	(2.449)	(1.522)
Profitability	-.704	-1.678**	-1.349**
	(-1.301)	(-2.381)	(-1.522)
Interest Coverage	.0001	.001	.0001
	(.970)	(.907)	(.325)
Tangibility	-.108	.004	-.024
	(-.980)	(.022)	(-1.153)
MB	-.165***	-.148***	-.102***
	(-4.201)	(-2.994)	(-.1.832)
zscore	-.071***	-.042*	-.023*
	(-3.656)	(-1.629)	(-1.325)
Loan Size	-.0001***	-.0001***	-.0001***
	(-5.485)	(-4.193)	(-3.514)
Prior Relations	-.093*	-.009	-.018
	(-1.591)	(-.1)	(-.216)
Control For			
Cluster Effect	Y	Y	Y
Industry Effect	N	N	N

Table 15 (Continued)

Loan-Type effect	Y	Y	Y
Loan-Purpose Effect	Y	Y	Y
Observations	259	163	163
Adjusted R-Squared	0.567	0.609	0.652

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 16 Robustness Check- Exclude Financial and Utility Firms

Dependent Variable	Log(spread)	Log(spread)	Log(spread)
	(1)	(2)	(3)
XBRL	-0.128***		-.083*
	(-3.069)		(-1.203)
Social Media		-.087*	.102
		(-.955)	(2.230)
XBRL*Social Media			-.087**
			(-1.321)
Log(asset)	-.153***	-.413***	-.189***
	(-2.739)	(-6.662)	(-3.822)
Leverage	.381***	.526***	.228**
	(3.359)	(2.842)	(1.351)
Current Ratio	.019	.00001	.0002
	(1.030)	(.0004)	(.011)
Cash to Debt Ratio	-.00001	-.00001	-.00001
	(-.797)	(-.837)	(-.765)
Profitability	-1.636***	-1.973***	-1.601***
	(-4.350)	(-3.292)	(-3.112)
Interest Coverage	-.0003*	-.0001*	.0002*
	(1.828)	(1.805)	(1.236)
Tangibility	-.037	-.049	.004
	(-.412)	(-.302)	(.032)
MB	-.067**	-.110**	-.057**
	(-2.432)	(-2.501)	(-1.869)
zscore	-.078***	-.058**	-.071***
	(-4.371)	(-2.365)	(-3.456)
Loan Size	-.0001***	-.0001***	.0001***
	(-6.043)	(-4.156)	(-4.632)
Prior Relations	-.028	.021	-.028
	(-.661)	(.281)	(-.393)
Control For			
Industry Effect	Y	Y	Y
Loan-Type effect	Y	Y	Y

Table 16 (Continued)

Loan-Purpose Effect	Y	Y	Y
Observations	370	187	187
Adjusted R-Squared	.657	.698	.715

Notes: Significance at the 10%, 5%, and 1% levels is indicated by *, ** and *** respectively.

Table 17 Robustness Check- Median Regression

Dependent Variable	Log(spread)	Log(spread)	Log(spread)
	(1)	(2)	(3)
XBRL	-.108***		-.098**
	(-2.696)		(-1.854)
Social Media		-.001**	0.032
		(-2.439)	(1.223)
XBRL*Social Media			-.0004*
			(-1.559)
Prior Relations	-.034*	-.06	-.053*
	(-.761)	(-1.364)	(-1.228)
Log(asset)	-.285***	.281***	-.283***
	(-6.237)	(5.641)	(-8.320)
Leverage	.001	.002*	.001
	(.348)	(1.487)	(.235)
Current Ratio	.001*	.001*	.001
	(1.507)	(1.891)	(2.002)
Cash to Debt Ratio	-.0003	-.0001	-.0001
	(-1.308)	(-.317)	(-.282)
Profitability	.0004	.0003	.0001
	(.665)	(.568)	(0.352)
Interest Coverage	-.0003*	-.001**	-0.0001*
	(-1.412)	(-2.503)	(-.885)
Tangibility	-.002*	-.0004	-.001
	(-2.031)	(-0.373)	(-.562)
MB	-.003***	-.003***	-0.002***
	(-5.204)	(-5.321)	(-4.385)
zscore	-.001***	-.001***	-.001***
	(-3.827)	(-3.537)	(-4.215)
Loan Size	-.0004	-.001***	-.001*
	(-.767)	(-.409)	(-1.125)
Control For			
Industry Effect	Y	Y	Y
Loan-Type effect	Y	Y	Y

Table 17 (Continued)

Loan-Purpose Effect	Y	Y	Y
Observations	259	183	183
Adjusted R-Squared	0.602	0.592	0.657

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 18 Firm Size, XBRL Adoption, Social Media Sentiment, and Loan Price

Dependent Variable	Log(spread)	Log(spread)
	(1)	(2)
XBRL	-.164***	
	(-3.797)	
XBRL*Small Firms	-.052	
	(-.806)	
Social Media		-.312***
		(-2.879)
Social Media*Small Firms		.370**
		(2.165)
Small Firms	.170**	.362***
	(2.350)	(4.627)
Leverage	.399***	.582***
	(3.489)	(3.130)
Current Ratio	.033**	.010
	(1.818)	(.3)
Cash to Debt Ratio	-.00001	-.00001
	(-.763)	(-.567)
Profitability	-1.180***	-1.785***
	(-3.210)	(-3.051)
Interest Coverage	.0003	.0003
	(1.465)	(1.419)
Tangibility	-.104	-.128
	(-1.181)	(-.853)
MB	-.082***	-.110***
	(-3.033)	(-2.602)
zscore	-.065***	-.035
	(-3.627)	(-1.416)
Loan Size	-.0001***	-.00015***
	(-7.595)	(-7.680)
Prior Relations	-.044	-.055
	(-1.041)	(-.718)
Control For		
Industry Effect	Y	Y

Table 18 (Continued)

Loan-Type effect	Y	Y
Loan-Purpose Effect	Y	Y
Observations	411	211
Adjusted R-Squared	0.627	0.656

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 19 New Relationship, XBRL Adoption, Social Media Sentiment, and Loan Price

Dependent Variable	Log(spread)	Log(spread)
	(1)	(2)
XBRL	-.118***	
	(-2.737)	
XBRL* New Loans	-.050	
	(-.869)	
Social Media		-.243***
		(-2.400)
Social Media* New Loans		.270
		(1.464)
New Loans	.080	-.011
	(1.386)	(-.127)
Log(asset)	-.183***	-.411***
	(-3.379)	(-6.997)
Leverage	.410***	.552***
	(3.631)	(3.036)
Current Ratio	.030*	.008
	(1.705)	(.268)
Cash to Debt Ratio	-.00001	-.00001
	(-.832)	(-.886)
Profitability	-1.260***	-1.644***
	(-3.429)	(-2.821)
Interest Coverage	.0003	.0003
	(1.479)	(1.495)
Tangibility	-.105	-.213
	(-1.205)	(-1.455)
MB	-.089***	-.133***
	(-3.319)	(-3.139)
zscore	-.069***	-.052**
	(-3.907)	(-2.122)
Loan Size	-.0001***	-.0001***
	(-5.284)	(-3.605)
Control For		
Industry Effect	Y	Y

Table 19 (Continued)

Loan-Type effect	Y	Y
Loan-Purpose Effect	Y	Y
Observations	411	211
Adjusted R-Squared	0.632	0.669

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 20 Syndication, XBRL Adoption, Social Media Sentiment, and Bank Loan Price

Dependent Variable	Log(spread)	Log(spread)
	(1)	(2)
XBRL	-.577***	
	(-3.219)	
XBRL*Syndication	.444**	
	(2.529)	
Social Media		-1.178**
		(-1.978)
Social Media* Syndication		1.040*
		(1.737)
Syndication	.532***	.185
	(2.865)	(1.043)
Log(asset)	-.164***	-.388***
	(-3.029)	(-6.446)
Leverage	.427***	.643***
	(3.772)	(3.341)
Current Ratio	.033**	.010
	(1.843)	(.325)
Cash to Debt Ratio	-.00001	-.00001
	(-.745)	(-.706)
Profitability	-1.172***	-1.336**
	(-3.199)	(-2.316)
Interest Coverage	.0003	.0004*
	(1.522)	(1.640)
Tangibility	-.102	-.219
	(-1.180)	(-1.498)
MB	-.089***	-.141***
	(-3.355)	(-3.354)
zscore	-.067***	-.049**
	(-3.841)	(-2.029)
Loan Size	-.0001***	-.0001***
	(-5.548)	(-3.785)
Prior Relations	-.037	-.048
	(-.877)	(-.638)
Control For		

Table 20 (Continued)

Industry Effect	Y	Y
Loan-Type effect	Y	Y
Loan-Purpose Effect	Y	Y
Observations	411	211
Adjusted R-Squared	0.638	0.670

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 21 Loan Maturities, XBRL Adoption and Social Media Sentiment

Dependent Variable	Log(Maturity)	Log(Maturity)
	(1)	(2)
XBRL	-.007	
	(-.161)	
Social Media		-.029
		(-.362)
Log(asset)	.074	.011
	(1.336)	(.186)
Leverage	.077	.162
	(.668)	(.910)
Current Ratio	.037**	.053*
	(2.048)	(1.725)
Cash to Debt Ratio	.000004	.000002
	(.547)	(.276)
Profitability	.001	.304
	(002)	(.543)
Interest Coverage	-.0001	-.0001
	(-.602)	(-.474)
Tangibility	.063	.001
	(.710)	(.009)
MB	.013	-.027
	(.475)	(-.654)
zscore	.049***	.025
	(2.739)	(1.051)
Loan Size	.00001	.00002
	(.409)	(.896)
Prior Relations	-.037	-.030
	(-.872)	(-.417)
Control For		
Industry Effect	Y	Y
Loan-Type effect	Y	Y
Loan-Purpose Effect	Y	Y

Table 21 (Continued)

Observations	411	211
Adjusted R-Squared	0.348	0.449

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

Table 22 Collateral, XBRL Adoption, and Social Media Sentiment

Dependent Variable	Secured	Secured
	(1)	(2)
XBRL	-.062*	
	(-1.635)	
Social Media		-.079
		(-1.275)
Log(asset)	-.075	-.131***
	(-1.484)	(-2.976)
Leverage	.261**	.176
	(2.462)	(1.298)
Current Ratio	.030*	-.009
	(1.775)	(-.372)
Cash to Debt Ratio	-.00001*	-.00001
	(-1.719)	(-1.433)
Profitability	-1.410***	-.753*
	(-4.092)	(-1.765)
Interest Coverage	.001***	.001***
	(2.693)	(2.995)
Tangibility	.008	-.107
	(.104)	(-.975)
MB	.036	-.010
	(1.415)	(-.306)
zscore	.033**	.025
	(2.015)	(1.382)
Loan Size	.00002	.00002
	(1.198)	(1.323)
Prior Relations	-.084**	-.129**
	(-2.146)	(-2.348)
Control For		
Industry Effect	Y	Y
Loan-Type effect	Y	Y
Loan-Purpose Effect	Y	Y
Observations	411	211

Table 22 (Continued)

Adjusted R-Squared	0.128	0.136

Notes: Significance at the 10%, 5%, and 1% levels is indicated by $*$, $**$ and $***$ respectively.

APPENDICES

APPENDIX A: DEFINITION OF STUDY VARIABLES

Variables	Definitions
Log(spread)	The natural logarithm of spread, where spread is the initial interest rate spread over London Interbank Offered Rate (LIBOR)
XBRL	Times of borrowers included XBRL in its financial reporting from 2009 to 2011
Social Media	Sum of financial terms sentiment and general sentiment, where Financial terms sentiment is the finance-related sentiment score of postings, general sentiment is the general sentiment score of postings.
Log(Asset)	The natural logarithm of the total assets of borrowers
Leverage	Total debts including long-term debt and short term debt divided by firm book assets
Current Ratio	Current assets divided by current liability
Cash to Debt Ratio	Total cash divided by total debt
Profitability	Net income over total sales
Interest Coverage	EBIT divided by total interest expense
Tangibility	Net property, plant, and equipment divided by total assets
MB	Market to book ratio
Loan Size	Total amount of bank loan
Prior Relations	Dummy variable which is equal to one when there is a previous lending relationship between lenders and borrowers, it equals zero otherwise
zscore	(1.2*Working capital+1.48Retained earnings + 3.3*EBIT + 0.999*Sales)/Total assets
Maturity	Loan maturity
Secured	Dummy variable, which equals one if a firm's assets are less than the sample median of total assets, and zero otherwise
New Loans	Dummy variable, which equals one if when there is no a previous lending relationship between lenders and borrowers, and zero otherwise
Syndication	Dummy variable, which equals one if a loan is offered by more than one lender, and zero otherwise

APPENDIX A (CONTINUED)

Variables	Definitions
Small Firms	Dummy variable, which equals one if a firm's assets are less than the sample median of total assets, and zero otherwise

VITA

Dazhi Chong
Department of Information Technology & Decision Sciences
Strome College of Business
Old Dominion University, Norfolk, VA 23529

Education

Ph.D., Information Technology, Old Dominion University, Norfolk, VA, 2016

M.S., Computer Application Technique, Hefei University of Technology, China, 2006

B.S., Accounting, Anhui Institute of Finance and Trade, China, 1998

Dissertation

The Effect of XBRL and Social Media on Information Asymmetry: Evidence from Bank Loan Contracts

Research Interests:

Business Analytics, Social Network Analysis, Computer Supported Cooperative Work, Data Mining, Financial Analysis

Teaching Experiences:

August 2012 - May 2016: Adjunct Professor, Strome College of Business, Old Dominion University, Norfolk, VA

Award

Outstanding Doctoral Student in Information Technology, Strome College of Business, Old Dominion University, 2014-2015