Journal of Banking & Finance 68 (2016) 100–116

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

The impact of news articles and corporate disclosure on credit risk valuation

Feng-Tse Tsai [a,*], Hsin-Min Lu [b], Mao-Wei Hung [c]

[a] Department of Finance, Asia University, No. 500, Lioufeng Rd., Wufeng, Taichung City 41354, Taiwan
[b] Department of Information Management, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei City 106, Taiwan
[c] Department of International Business, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei City 106, Taiwan

A R T I C L E I N F O

Article history:
Received 12 August 2014
Accepted 30 March 2016
Available online 2 April 2016

JEL classification:
M41
G1
G12

Keywords:
Credit default swap
Credit risk
News coverage
News sentiment
Risk factor disclosure

A B S T R A C T

In this study, we investigate how qualitative information in newspapers and corporate filings affects credit risk valuation in the credit default swap (CDS) market. We adopted news coverage and news sentiment to quantify text information from news articles and quantified the qualitative risk disclosures of individual firms in their corporate filings (i.e., Form 10-K and 10-Q). Our empirical study, based on 13 years of CDS data, provides several conclusions. First, more news coverage and negative news sentiment increase credit risk. Second, a higher overall volume of risk factor disclosure in corporate public filings is linked to a higher credit risk for debt issuers. Moreover, financial risk has the strongest effect among the five types of risk disclosures we considered. Overall, our results suggest that text information from newspapers and corporate filings contains incremental informational content for firms' credit risk evaluations. These two information sources play distinctive roles in signaling issuers' future credit conditions.

1. Introduction

Modeling and managing credit risk is an important yet difficult task because creditworthiness is a complex function of both quantitative and qualitative information. Most credit risk models depend on quantitative information from financial reports and securities markets. These models lack the important qualitative information from corporations and third-party sources such as news articles. By including text information from newspapers and corporate filings, and then documenting their empirical significance, we form an extended information set for credit evaluation.

Recent studies on stock markets document the value of qualitative data in the news media and full-text corporate filings (Bodnaruk et al., 2015; Campbell et al., 2014; Chan, 2003; Loughran and McDonald, 2011; Tetlock, 2007; Tetlock et al., 2008), thereby shedding light on the usefulness of such qualitative data for credit risk modeling. The news media injects public information into the market, an act that can influence an investor's perception and attention. A recent study has shown that, after controlling for well-known risk factors, stocks with media coverage earn lower returns compared to those without media coverage (Fang and Peress, 2009). Ozick and Sadka (2013) report that media coverage predicts mutual fund performance. Chan (2003) documents that a stock's price changes differ depending on whether it is connected to a firm with recent media exposure. Previous studies have also documented that, in addition to media coverage, news tone (sentiment) in widely circulated newspapers, such as the Wall Street Journal (WSJ), influences short-term stock returns (Tetlock, 2007; Tetlock et al., 2008). Moreover, insider trading and differences in the quality of the news dissemination mechanism can explain cross-country differences in the reaction of a stock price to news announcements (Griffin et al., 2011).

Another source of qualitative information is the free-text disclosure in corporate filings, which contains valuable information for investors. Unlike the news media, which produces articles written by professional reporters in an effort to attract readership, corporate filings are prepared by management and communicate firm performance as well as potential future developments. The agency

* Corresponding author. Tel.: +886 4 2332 3456; fax: +886 4 2332 1181.
E-mail addresses: fttsai@gm.asia.edu.tw (F.-T. Tsai), luim@ntu.edu.tw (H.-M. Lu), mwhung@ntu.edu.tw (M.-W. Hung).

http://dx.doi.org/10.1016/j.jbankfin.2016.03.018
0378-4266/

problem might cause management to manipulate disclosure content. However, management may face lawsuits if they fail to disclose important risk factors. Previous studies investigate the value of corporate filings. Li (2008) documents that filing readability, a weighted average of sentence length and the proportion of complex words, is an indicator of future firm performance. Campbell et al. (2014) show that the free-text risk factor disclosures provided by corporate managers meaningfully convey a firm's risks to investors. Recently, Hoberg and Maksimovic (2015) conducted a text-based analysis to create measures of financial constraints using 10-K text extracted from the ''Management's Discussion and Analysis" (MD&A) section. Moreover, Bodnaruk et al. (2015) found that the frequency of financial constraining words predicts subsequent liquidity events.

Because qualitative information in the news media and corporate filings influences stock price and hence asset value, a change in firm asset value is likely to lead to a change in the credit quality of the firm (Odders-White and Ready, 2006), thereby influencing the credit market. From the perspective of credit risk valuation, such qualitative information can improve existing models by conveying risk that is difficult to capture using only quantitative information. The empirical study of Norden (2008), who documented that the credit default swaps (CDS) market response is stronger for firms with high amounts of media coverage, provides the initial support for our conjecture.

Despite the similarity between the stock market and the CDS market, the CDS market is different in at least three aspects. First, a CDS is traded on the OTC market, which is neither regulated nor supervised. The participants are exclusively institutional traders such as commercial and investment banks, insurance companies, and hedge funds (Norden, 2008). Institutional traders often have more relevant information than public investors. For example, some banks are CDS traders and lenders to the underlying firms; this allows these traders to have access to private information about their borrowers (Acharya and Johnson, 2007). Second, institutional investors are more skillful in analyzing qualitative information, such as news, and might be able to update their knowledge about the market more quickly than non-institutional traders. Third, stockholders and bondholders react differently to news in some cases. Because of the agency problem between stockholders and bondholders, certain events considered as good news by stock investors can have negative effects on bond returns (e.g. Goh and Ederington (1993)). It follows that the reaction to qualitative information in the CDS market might also differ from the stock market.

Owing to the distinct characteristics of the CDS market and the lack of studies on the effect of the news and corporate risk disclosures in the context of credit risk valuation, there is a need for a comprehensive study on the connection between the CDS market, credit risk valuation, and public qualitative data. To bridge this gap, we investigate how qualitative information in newspapers and corporate filings affect credit risk valuation in the CDS market. We included both text sources because corporate disclosures and news reports have different rationales (Kothari et al., 2009). Specifically, we empirically examine the impact of news coverage, news sentiment, and risk factor disclosures in corporate filings on CDS spreads. Following previous studies, we defined news coverage as the number of mentions a firm receives in the news during a quarter (Fang and Peress, 2009; Odders-White and Ready, 2006). News sentiment is the percentage of negative words based on the financial dictionary developed by Loughran and McDonald (2011). To quantify the risk disclosure in corporate filings, we adopted the risk disclosure dictionary (Campbell et al., 2014) and characterized risk disclosure as the risk word counts in systematic, idiosyncratic, financial, tax, and litigation risks normalized by the total number of words in a corporate filing. The overall volume of all risk words is taken as risk disclosure intensity.

Our study offers three main findings. First, our empirical analysis shows that negative news is associated with higher CDS spread, suggesting that investors in the credit market consider a negative news tone as a signal for degrading credit risk, and that firm-specific news is informative. Second, news media coverage, as measured by the number of news articles mentioning a firm in a quarter, is positively associated with CDS spreads. Third, the overall volume of risk factor disclosure in corporate filings, as measured by the logarithm of total number of risk keywords, has a positive effect on CDS spreads. Among the five risk types, financial risk has the strongest positive effect on CDS spreads.

The contributions of our research are as follows. First, we contribute to the emerging literature on the economic effect of news sentiment. Most previous studies on the subject have investigated this issue in the equity market alone. We extend the literature by documenting that the credit market also incorporates information quantified by negative news sentiment.

Second, we document that the news coverage effect also exists in the credit market. Fang and Peress (2009) documented the no-media premium that is likely caused by the smaller investor base of firms with no media coverage. The argument is less relevant to the CDS market because the buyers of CDS contracts are debt holders or lenders to a firm; these buyers want to hedge their credit exposure or transfer their credit risk to CDS sellers. News coverage is thus unlikely to influence the demand of CDS contracts. However, CDS sellers might decide to raise insurance fees (CDS spreads) when the underlying firm increasingly suffers from public information shocks, as these can lead to higher information uncertainty and thus a higher degree of uncertainty in the firm's valuation. The news coverage effect in the CDS market is not subsumed by its counterpart in the equity market.

Finally, we extend the literature on the information content of corporate disclosure by documenting that investors in the credit market incorporate corporate filings' risk disclosures into their assessments of firm risk. We found more risk factor disclosure to be associated with a higher credit risk. The result is consistent with Kravet and Muslu's argument that text risk disclosures sharpen retail investors' risk perceptions (Kravet and Muslu, 2013). In addition, our findings show that more financial and systematic risk factor disclosure is linked to higher future credit risk. Campbell et al. (2014) also found systematic risk factor disclosure to be positively related to future market risk (as measured by beta).

The remainder of this paper is organized as follows. Section 2 states our research hypotheses. Section 3 discusses the data sources, key variables, and empirical models. Section 4 presents and discusses empirical results, and checks their robustness. Finally, Section 5 concludes the paper.

2. Hypothesis development

Previous studies have offered several explanations for the effect of media coverage on stock prices. These explanations are based on alternative views of the media: as an information intermediary, a corporate governance tool, a biased transmitter of favorable information, and a generator of investor sentiment. The information intermediary hypothesis views media coverage as a critical source for communicating information about firms to the public, thereby effectively reducing the information superiority managers would have over outside investors (i.e., information asymmetry problems) (Healy and Palepu, 2001). Dyck et al. (2008) argue that the media can influence managerial decisions, an idea entangled with that of the corporate governance hypothesis, which posits that the media can balance the private benefits and costs associated with the managerial actions that expropriate shareholders' benefits.

These two hypotheses imply that greater levels of media coverage should enhance a firm's value, hence reducing its credit risk.

A third hypothesis, the biased media hypothesis, claims that the media tends to report information biased in favor of the companies covered (Gurun and Butler, 2012), leading to fewer negative words in news articles about favored firms. Alternatively, Tetlock (2007) adopted a qualitative measure of pessimistic news coverage to show that high media pessimism predicts downward price pressure (the media-induced sentiment hypothesis). Both hypotheses infer that more media coverage might deviate a firm's market value from its fundamental value, resulting in volatility in the firm's value and increasing its credit risk.

The main participants in CDS markets are financial institutions. Although they are less subject to limited attention than individual investors, they are still biased. For an individual, the media coverage of individual stocks can influence trading behavior (Barber and Odean, 2008; Da et al., 2011) because an individual is more likely to trade stocks featured in the media. For an institution, Fang et al. (2014) discovered that professional investors such as mutual fund managers also tend to buy media-covered stocks.

The literature on information intermediaries and corporate governance suggests that more news coverage should reduce a firm's credit risk (a negative relationship between a firm's news coverage and its credit spread should exist). Another possibility is that credit risk increases if an issuer has more news coverage based on media-biased and media-induced sentiment hypotheses. Prior research shows that more press coverage correlates with a higher information risk[1] and information uncertainty (Chen et al., 2013; Odders-White and Ready, 2006), thus increasing asset value uncertainty and credit risk. Direct evidence provided by Lu et al. (2010) shows that investors charge a significantly higher bond risk premium for both information uncertainty and information asymmetry. In addition, Hertzberg et al. (2011) found that public information about a firm in distress exacerbates lender coordination and increases the incidence of financial distress. Because media coverage can have a distinct impact on credit risk, we propose the following two competing hypotheses:

Hypothesis 1a. The news coverage of a debt issuer is negatively related to the issuer's credit risk.

Hypothesis 1b. The news coverage of a debt issuer is positively related to the issuer's credit risk.

The findings of Tetlock (2007) and Tetlock et al. (2008) demonstrate that news sentiment, especially negative words, influences stock markets. Similarly, bad news about a firm or a country (e.g., Greece) will also result in increased yields of issued bonds. Additionally, previous studies found that CDS significantly increases around negative rating events (Hull et al., 2004; Norden and Weber, 2004). Therefore, we expect that news sentiment, especially that with a negative tone, influences credit risk.

Hypothesis 2. Relevant, negative sentiment in news articles is associated with an increase in the credit risk of a debt issuer.

A risk factor disclosure involves discussions of the conditions, trends, or issues that can affect a company's business, prospects, operating outcomes, or financial condition. Although some smaller companies are not required to include risk factors in their annual Form 10-K and quarterly Form 10-Q, these companies can still mitigate the risk of liability to their shareholders by providing proper disclosure of the material risks associated with investing in their securities. Risk factors thus act as an insurance policy as well as a strong defense against shareholder litigation. Using the content in Item 1A of Form 10-K mandated by the Securities and Exchange Commission (SEC) since 2005, Campbell et al. (2014) discovered that firms facing greater risk disclose more risk factors and that the information conveyed by the risk factor disclosures is reflected in systematic risk, idiosyncratic risk, information asymmetry, and firm value.

Nonetheless, the disclosure of risk factors is part of the incentive-based disclosure regime, according to the safe harbor provision of the Private Securities Litigation Reform Act (PSLRA), which was adopted in 1995, before the implementation of mandatory risk-factor disclosure in 2005. The PSLRA protects firms from the liability of forward-looking statements by adding to these statements the clause, "meaningful cautionary statements identifying important factors that could cause actual results to differ materially from those in the forward looking statement." Thus, the safe harbor provision provides an important incentive for public companies to disclose risk factors.

Risk factor disclosure shifted from a voluntary regime to a mandatory one in 2005, when the SEC added Item 1A to Form 10-K. Item 1A requires most public companies to disclose risk factors annually in addition to the quarterly updates as per Form 10-Q. Because corporations disclosed their risk factors either voluntarily (before 2005) or mandatorily (after 2005) in our sample period, we predict that the text information from risk factor disclosures communicates the creditworthiness of a debt issuer; an issuer bearing greater credit risk discloses more risk factors.

Hypothesis 3. Corporate risk factor disclosure intensity (volume) is positively linked to a corporation's credit risk.

The amount of risk disclosure has credit risk implications of a debt issuer. In addition, the type of risk a firm faces also has distinct implications. In this paper, we consider five subcategories (types) of risk factors: financial risk, litigation risk, tax risk, systematic risk, and idiosyncratic risk. Words were classified as indicators of (i) financial risk if they were generally related to liquidity, debt, covenants, or capital structure; (ii) litigation risk if they were associated with legal matters, lawsuits, intellectual property, or environmental issues; (iii) tax risk if they were related to the accounting of income taxes or general tax avoidance; (iv) systematic or idiosyncratic risks if they were most closely related to an economy-wide risk or a firm-specific risk, respectively. Of these types, we expected financial risk to be the most relevant to an issuer's credit risk.

Litigation risk disclosure, if value-relevant, should be positively related to an issuer's credit risk. Lin et al. (2013) analyzed the impact of Directors & Officers (D&O) insurance, which is purchased by a company to protect its directors and officers from personal liability in the event of litigation, on the charged bank loan spreads. They found that lenders consider D&O insurance coverage as an indicator of high credit risk (potentially, the link is via moral hazard or information asymmetry), and this relation is attenuated by monitoring mechanisms. Therefore, if companies unveil more litigation risk, they endanger their credit quality.

The relationship between tax risk and credit risk can be positive. Hasan et al. (2014) found that firms with greater tax avoidance incur higher spreads when obtaining bank loans. In addition, previous studies found that corporate income tax can affect a firm's capital structure (Heider and Ljungqvist, 2015). Therefore, we expect tax risk disclosure to affect credit risk, with more tax risk disclosures increasing credit risk.

More systematic risk mentioned in corporate filings is linked to a higher credit risk for the issuers. Economy-wide risk appears frequently in filings when the market is turbulent. Additionally,

[1] Chen et al. (2013) defined information risk as the risk that arises when some investors are better informed than others.

clustered defaults occur during a financial crisis. Therefore, the default risk is higher if more systematic risk emerges. High credit risk accompanies more firm-specific risk when idiosyncratic risk leads to worsening credit.

Hypothesis 3a. Corporate financial risk factor disclosure is positively associated with credit risk.

Hypothesis 3b. Corporate tax risk factor disclosure is positively associated with credit risk.

Hypothesis 3c. Corporate litigation risk factor disclosure is positively associated with credit risk.

Hypothesis 3d. Corporate systematic risk factor disclosure is positively associated with credit risk.

Hypothesis 3e. Corporate idiosyncratic risk factor disclosure is positively associated with credit risk.

3. Data and empirical models

3.1. Data sources

Our sample contains CDS data from 2001 to 2013, with matching news variables from the Wall Street Journal (WSJ), risk disclosure variables constructed from public corporate disclosure (i.e., 10-K and 10-Q forms), and firm characteristics constructed from accounting and stock market variables. We obtained the CDS data from the Markit Group, which is a financial information service provider created by major CDS dealers in 2001. The data have been widely adopted in recent academic research related to credit derivatives (e.g., Qiu and Yu, 2012). To deal with the lack of trade and quote information, Markit gathers daily closing prices from dealers' books. After filtering out stale prices and outliers, Markit uses pricing information from all data contributors to calculate a daily composite term structure of the CDS spread for each obligor.

We only considered the CDS contracts of U.S. companies denominated in U.S. dollars on senior unsecured debt with a maturity of five years and a Modified Restructing (MR) clause following Bongaerts et al. (2011). In practice, the five-year CDS is the most popular form of CDS and is less constrained by illiquidity. We focused on five-year contracts only for senior unsecured debts with an MR clause because such contracts are the most common ones of their type in North America.

We considered two types of textual data: newspapers and company public filings. We adopted the WSJ as our source for news articles. The WSJ is a widely circulated business newspaper, with a circulation of more than two million.[2] We developed a sequence of data processing procedures for identifying firms associated with each news article and the corresponding sentiment from the full text of the news. We obtained annual reports (10-K) and quarterly reports (10-Q) from SEC's Electronic Data Gathering and Retrieval (EDGAR) system.[3] We processed the full text in each document to compute the risk factor disclosure variables adopted in this study. The subsequent section discusses the construction of text variables used in our study.

[2] The figures include normal print editions, branded print editions (e.g., regional editions or editions tailored for commuters), and digital subscriptions. The data is compiled by the Alliance for Audited Media.

[3] We excluded 10-K/A, 10-K405, 10-K405/A, and 10-Q/A in this study.

We obtained accounting variables and credit ratings from the North American Compustat dataset and daily stock prices from the Center for Research in Security Prices (CRSP). The risk-free rate was obtained from the Federal Reserve Bank of St. Louis. We identified financial firms based on the Fama-French 17-industry classification (Fama and French, 1997) and excluded these firms from the sample. Further, we dropped firms without available 10-K and 10-Q reports on EDGAR. In addition, observations with any missing values were discarded; for example, firm-quarters with no news coverage in a quarter were excluded to remove data with missing news variables. The final dataset contained 8562 firm-quarters from 491 unique issuers.

3.2. Variable construction

Our study involved variables constructed from newspapers and corporate filings. In addition, we considered control variables, including accounting, market, and macroeconomic variables. We discuss the construction of these variables below.

3.2.1. News variables

We processed news articles to construct firm-specific news coverage and sentiment variables. For each news article, we identified the main company of the article by identifying the named entities in the article, mapping the identified named entity to a company identifier (PERMCO), and selecting the main company identifier for the article. A processed news article could be linked to either zero or one main company. Appendix A provides additional technical details of text processing algorithms.

News sentiment can gauge the level of pessimism in a news article. Drawing from previous studies, we adopted the Loughran and McDonald (LM) dictionary (Loughran and McDonald, 2011) to count the negative words for each firm within a quarter. For each of the words in the dictionary, we constructed a regular expression that searched through the full text (excluding the entity names identified during the main company mapping) for an exact match of the word. We recorded the number of matches for each word category (e.g., Negative) and the total word length of the document. The procedure was developed using Python 2.7. The built-in regular expression module, "re," was adopted for keyword matching.

Specifically, the negative tone measure (neg) was defined as follows:

$$neg = \frac{\text{negative word count}}{\text{total word count}}$$

We checked the correctness of our procedure by randomly sampling 100 news articles and manually verifying the output in each processing step. We also checked the articles with the largest negative sentiment scores (among the CDS issuers considered in our study; included in Appendix B) to confirm that our algorithms were working properly.

3.2.2. Risk disclosure variables

We constructed risk disclosure variables from the full-text mandatory corporate disclosures, i.e., the 10-K and 10-Q forms. To do so, we developed a sequence of automatic processing routings that downloaded the forms, extracted headers as well as the full text, cleaned up the text data, extracted the selected filing items, and computed the risk disclosure variables.

We downloaded the 10-K and 10-Q forms from the EDGAR FTP Server (ftp://ftp.sec.gov/edgar/) and stored the files in a relational database. For each downloaded file, we extracted meta-data, such as the period of report and filing date, from the header. We also extracted and preprocessed the main content in pure text or HTML format to correct for any encoding problems. We extracted Item 1A

(Risk Factors) from the 10-K filings following the procedure outlined by Campbell et al. (2014). All HTML tags were removed before the computation of risk disclosure variables.

Our study considers the overall risk disclosure intensity (*risk_all*) and five risk subcategories: systematic (*risk_sys*), idiosyncratic (*risk_idio*), financial (*risk_fin*), tax (*risk_tax*), and litigation (*risk_lr*) risks. Using the dictionary provided by Campbell et al. (2014), we computed the counts of the risk keywords under the five subcategories in corporate filings. We defined the overall risk disclosure intensity (*lrisk_all*) as the logarithm of one plus the total risk keyword count. In addition, we calculated the ratio of the risk keyword count to the total word count in each risk category to characterize disclosure in these categories. The risk disclosure variables for Item 1A (considered in robustness checks) were computed via a similar procedure. We included additional output for the 10-K filings with the highest scores in each risk subcategory in Appendix B. The risk disclosure measures are defined as follows.

$$lrisk\_all = \log(1 + \text{risk word count in all subcategories}),$$

$$risk\_sys = \frac{\text{risk word count in subcategory of systematic risk}}{\text{total word count}},$$

$$risk\_idio = \frac{\text{risk word count in subcategory of idiosyncratic risk}}{\text{total word count}},$$

$$risk\_fin = \frac{\text{risk word count in subcategory of financial risk}}{\text{total word count}},$$

$$risk\_tax = \frac{\text{risk word count in subcategory of tax risk}}{\text{total word count}},$$

$$risk\_lr = \frac{\text{risk word count in subcategory of litigation risk}}{\text{total word count}}.$$

3.2.3. Control variables

We constructed a list of control variables, mostly following Das et al. (2009). We adopted 10 accounting-based variables to represent firm size (*lta*), profitability (*roa*, *nig*, and *ic*), financial liquidity (*qr* and *cta*), trading account activity (*taa*), sales growth (*sg*), and capital structure (lev and *reta*). Further, we used market-based variables including the distance to default (*dtd*), annualized prior 100-trading day equity return (*sret*), and volatility (*svol*). Macroeconomic and other variables are the three-month T-bill rate (*trate*), previous 12-month value-weighted industry return (*iret*) and S&P 500 return (*mret*), investment grade dummy (*invgrd*), and CDS liquidity measure (*lqn*). Table 1 lists the variable definitions.

We also followed Das et al. (2009) to deal with several issues in accounting-based variables. Specifically, we treat semi-annual numbers of flow items in Compustat by setting the first and second quarter data to one-half of the semi-annual numbers in the second quarter and the third and fourth quarter data to one-half of the semi-annual numbers in the fourth quarter when the data reported in the second and fourth quarters are semi-annual values. To account for seasonal effects, we took the trailing four-quarter average of ROA, sales growth, interest coverage, and inventories over cost of goods sold. Further, we considered the property of interest coverage (*IC*) and hence dealt with *IC* before and during running the trailing average of *IC*. First, we set any quarterly *IC* to zero if it was negative because a negative *IC* is not meaningful. Second, we took the four-quarter trailing average of *IC* and censored any value above 100 based on the assumption that further increases in value convey no additional information. Finally, we followed Das et al. (2009) to set up a piecewise *IC* function, as shown in Table 1 Panel B; the change in its value allows the data to determine the shape of the nonlinearity. This specification allows the regression models to select different coefficient parameters on each increment of the *IC* ratio (*c1* to *c4* in our explanatory variables).

We constructed the distance to default (*dtd*) as a market-based measure, drawing from Merton's structural model (Merton, 1974). In computing *dtd*, we assumed that a firm's value evolves as a geometric Brownian process and that the equity value (E) of the firm is equivalent to a call option on the underlying value of the firm with a strike price equal to the face value of debt (D) and a time to maturity of T. We adopted Moody's KMV model, which solves for the firm value V and its volatility σ_v using a system of two simultaneous equations:

$$\begin{cases} E = VN(d_1) - \exp(-rT)DN(d_2) \\ \sigma_e = \left(\frac{V}{E}\right)\left(\frac{\partial E}{\partial V}\right)\sigma_v \end{cases}, \tag{1}$$

where σ_e, μ, E, D, and r are obtained exogenously and $d_1 = \frac{\ln\left(\frac{V}{D}\right) + \left(r + \frac{\sigma_V^2}{2}\right)T}{\sigma_V\sqrt{T}}$, $d_2 = d_1 - \sigma_V\sqrt{T}$. The distance-to-default (*dtd*) is a market-based default measure and is defined as follows:

$$dtd = \frac{\log\left(\frac{V}{D}\right) + \left(\mu - \frac{\sigma_v^2}{2}\right)T}{\sigma_v\sqrt{T}}. \tag{2}$$

We estimated the annualized standard deviation of equity returns (σ_e) from the previous 100 trading days of stock price returns obtained from CRSP; the annualized average equity returns (μ) was estimated on the basis of the previous 100 trading days. We required at least 50 trading days to be available in the estimation, a requisite we decided upon by following the procedure described in a previous study (Bharath and Shumway, 2008). The market value of equity (E) is the number of shares outstanding multiplied by the end-of-quarter closing stock price from Compustat. The face value of debt, F, was set to be the debt in current liabilities plus one-half of the long-term debt (Vassalou and Yuhang, 2004). The risk-free rate, r, was obtained by using the three-month treasury constant maturity rate from the Federal Reserve Bank of St. Louis (e.g. Duffie et al. (2007)). Using these inputs, we numerically solved the system of equations in Eq. (1) to obtain V and σ_v, and then calculated the distance to default using Eq. (2).

We constructed our dataset by merging variables from different sources by quarter. We calculated the news variables using news articles from the first day to the last day of a matching quarter. Risk disclosure variables were merged with accounting variables according to the "confirmed period of report" in the headers of the filings. Market variables, such as the annualized previous 100 trading day equity return and volatility, are computed using the most recent data of a quarter. Macroeconomic variables, such as the three-month T-bill rate, were linked to the last month of a quarter. Finally, we obtained the five-year CDS spreads and their quote numbers for the last trading day of the calendar quarter. Because of delays between the end of a quarter and the release of corporate filings, we also tested the robustness by an alternative data-merging approach that lagged accounting variables and risk disclosure variables for one quarter.

3.3. Summary statistics

Our final sample consists of 8562 quarterly CDS spreads of 491 unique issuers. Panel A in Table 2 lists the mean and median CDS spreads by year. The CDS spreads were stable during 2004 and 2007 but increased sharply in 2008, the year of the financial crisis. In contrast, the number of CDS contracts increased yearly and reached their peak at the end of 2006.

Panel B shows that the car industry had the highest credit risk, according to their median CDS spreads in our sample period. In

Table 1
Variable definitions.

Variable	Definition
Panel A. CDS, text, accounting-based, market-based, macroeconomic, and other variables	
(l)cs	The (log) value of a five-year CDS credit spread
Text variables	
(l)narticle	The (log) value of WSJ news articles in a quarter (plus one before taking logarithm)
neg	The ratio of negative words to total words in WSJ news in the quarter, based on the dictionary of Loughran and McDonald (2011)
(l)risk_all	The (log) value of the total risk word count from firms' Form 10-K or Form 10-Q, based on the dictionary of Campbell et al. (2014) (plus one before taking logarithm)
risk_sys	The ratio of the systematic risk word count to the total word count, based on the dictionary of Campbell et al. (2014)
risk_idio	The ratio of the idiosyncratic risk word count to the total word count, based on the dictionary of Campbell et al. (2014)
risk_fin	The ratio of the financial risk word count to the total word count, based on the dictionary of Campbell et al. (2014)
risk_tax	The ratio of the tax risk word count to the total word count, based on the dictionary of Campbell et al. (2014)
risk_lr	The ratio of the litigation risk word count to the total word count, based on the dictionary of Campbell et al. (2014)
Accounting-based variables	
(l)ta	The (log) value of the total assets divided by the consumer price index on all-urban consumers. All items use the period of 1982–1984 as the base
roa	Return on assets as constructed by net income divided by total assets
nig	Net income growth as calculated by net income minus the previous quarter's net income, divided by total assets
IC	Interest coverage as calculated by pretax income plus interest expenses, divided by interest expenses
c_1–c_4	Interest coverage ratio in the regression model, according to Panel B
qr	Quick ratio as constructed by current assets minus inventories, over current liabilities
cta	Cash to asset ratio as constructed by cash and equivalents over total assets
taa	Trading account activity as measured by the ratio of inventories to cost of goods sold
sg	Quarterly sales growth as measured by sales divided by the previous quarter's sales, minus one
lev	Book leverage as the ratio of total liabilities to total assets
reta	The ratio of retained earnings to total assets
Market-based variables	
dtd	Distance to default measure, a measure of default probability as per Merton (1974)
sret	Annualized prior 100-trading day equity return
svol	Annualized prior 100-trading day equity volatility
Macroeconomic & other variables	
trate	Three-month T-bill rate (Federal Reserve Bank)
iret	Previous 12-month value-weighted industry return (according to Fama-French 17 industries)
mret	Previous 12-month S&P 500 return
nrating	Numerical credit rating that assigns a number to each S&P long-term issuer's credit rating. The mapping follows: AAA(1), AA+(2), AA(3), AA−(4), A+(5), A(6), A−(7), BBB+(8), BBB(9), BBB−(10), BB+(11), BB(12), BB−(13), B+(14), B(15), B−(16), CCC+(17), CCC(18), CCC−(19), CC(20), C(21), D(22), and SD(23)
invgrd	Investment grade dummies: equal to 1 if an S&P long-term issuer's credit rating is above BBB- (i.e., nrating $\leqslant$10)
(l)qn	The (log) quote number of five-year CDS spreads (Qiu and Yu, 2012)

Panel B. Interest coverage categories: the interest coverage ratio in the regression models is defined as $IC_{it} = \sum_{j=1}^{4} \kappa_j c_{jit}$, *where the corresponding* c_{jit} *is listed in the table*

	c_{1it}	c_{2it}	c_{3it}	c_{4it}
$IC_{it} \in [0, 5]$	IC_{it}	0	0	0
$IC_{it} \in [5, 10]$	5	$IC_{it} - 5$	0	0
$IC_{it} \in [10, 20]$	5	5	$IC_{it} - 10$	0
$IC_{it} \in [20, 100]$	5	5	10	$IC_{it} - 20$

addition, the machinery industry had the most observations, followed by the retail and utilities industries (barring the "others" classes in our sample).

Panel C shows CDS spreads by S&P credit ratings. It indicates that CDS spreads generally increase with worsening credit ratings, both in mean and median values. In addition, most observations were clustered between A+ and B+, and the number of observations decreased when moving to both ends. Moreover, the CDS spread sharply increased from the BBB− group (investment grade firms) to the BB+ group (speculative grade firms). This large price gap signals dramatic changes in credit risk between these two groups.

Table 3 lists the descriptive statistics of the text and other control variables. The first quartile of news coverage is two news articles per quarter while the third quartile is nine news articles per quarter. The mean news coverage (9.44) is much larger than the median (4), suggesting a right-skewed distribution. In addition, the news tended to have a positive tone, as the positive word count was usually larger than the negative word count (not shown). However, a positive word count is a relatively noisy sentiment measure (Tetlock et al., 2008) because of the potential negation issues.[4] Therefore, we adopted only the negative word count as the sentiment measure. The sentiment, as measured by the ratio of the negative word count to the total word count, was 1.94% on average. The average number of risk words in Form 10-K and 10-Q was 1237.38, and the ratio of systematic risk were higher than those of the other risk subcategories.

3.4. Empirical model

Because a credit default swap (CDS) is an insurance contract in which the buyer periodically pays a premium to the seller in exchange for compensation in the event of a default, a fairly priced CDS must be characterized by having its expected present value of premium payments equal to the expected present value of default

[4] We thank the reviewer (Anonymous Referee #1) for pointing this out.

Table 2
Mean and median CDS spreads by year, industry classification, and credit rating. This table lists the means and medians of the CDS spreads for issuers in terms of year (Panel A); specific industry, as classified by Fama and French (1997) (Panel B); and S&P rating category (the corresponding numerical rating is in parentheses) (Panel C). The last column reports the number of issuers used to calculate the mean and median of the CDS spreads. Data contains all issuers (excluding financial firms) with available CDS spreads from 2001 to 2013.

Year	CDS mean	CDS median	Number of OBS.
Panel A: CDS spreads by year			
2001	165.75	88.43	283
2002	248.36	84.86	483
2003	146.95	49.72	690
2004	124.22	43.82	793
2005	120.99	45.84	849
2006	93.23	40.79	873
2007	100.20	40.25	657
2008	341.42	140.76	762
2009	346.02	115.30	755
2010	194.06	101.11	677
2011	195.03	106.06	628
2012	202.36	104.76	609
2013	166.83	80.88	503
Industry	CDS mean	CDS median	Number of OBS.
Panel B: CDS spreads by industry classification			
FOOD	75.62	46.69	581
MINES	209.04	160.33	125
OIL	118.89	58.05	589
CLTHS	118.35	56.25	201
DURBL	109.59	82.20	214
CHEMS	150.66	77.84	242
CNSUM	68.42	42.39	670
CNSTR	85.45	48.25	225
STEEL	205.53	112.11	217
FABPR	103.81	85.61	104
MACHN	146.04	73.07	997
CARS	455.09	247.69	160
TRANS	318.47	62.83	645
UTILS	171.92	86.54	751
RTAIL	255.76	97.64	785
OTHER	243.75	116.54	2056
S&P rating	CDS mean	CDS median	Number of OBS.
Panel C: CDS spreads by credit ratings			
AAA(1)	21.73	13.33	142
AA+(2)	30.83	30.98	3
AA(3)	30.45	23.98	234
AA−(4)	41.42	26.86	222
A+(5)	41.01	36.86	604
A(6)	51.04	40.00	1019
A−(7)	62.47	46.60	855
BBB+(8)	77.76	57.25	1013
BBB(9)	107.93	76.91	1419
BBB−(10)	150.67	115.47	992
BB+(11)	246.71	199.39	387
BB(12)	297.34	225.68	424
BB−(13)	436.13	332.62	421
B+(14)	467.97	390.00	321
B(15)	808.86	524.73	291
B−(16)	1146.53	811.07	152
CCC+(17)	1436.30	1039.40	33
CCC(18)	1621.01	1397.42	18
CCC−(19)	2606.27	2606.27	1
CC(20)	1888.27	1129.92	9
D(22)	4436.46	4436.46	1
SD(23)	2390.10	2390.10	1
Total			8562

loss payments under the risk-neutral probability measure. Specifically, the expected present value of premium payments at rate CS (credit spread) per annum for the premium leg is as follows:

$$E\left[\int_0^T \exp\left(-\int_0^\tau r_t dt\right) s(0,\ \tau) CS d\tau\right],$$

where $s(0,\ \tau) = exp(-\int_0^\tau \lambda_t dt)$ is the survival probability for the debt issuer from time 0 to τ, and λ_t denotes the default intensity. For the protection leg, the expected present value of loss payments for a notional value of \$1 is $\left[\int_0^T \exp\left(-\int_0^\tau r_t dt\right) s(0,\ \tau)\lambda_\tau(1-\rho_\tau)d\tau\right]$, where ρ_τ is the recovery rate at default time τ.

Table 3
Descriptive statistics. This table lists the descriptive statistics of the text, accounting-based, market-based, and macroeconomic variables adopted in this study. The definitions of these variables are summarized in Table 1.

Variable	*N*	Mean	Median	1st quartile	3rd quartile
Text variables					
narticle	8562	9.44	4.00	2.00	9.00
neg(%)	8562	1.94	1.84	1.29	2.42
risk_all	8562	1237.38	829.00	478.00	1544.00
risk_idio(%)	8562	1.18	1.13	0.86	1.44
risk_sys(%)	8562	1.39	1.10	0.79	1.62
risk_fin(%)	8562	0.51	0.48	0.36	0.63
risk_tax(%)	8562	0.40	0.39	0.28	0.51
risk_lr(%)	8562	0.34	0.30	0.17	0.46
Accounting-based variables					
ta	8562	10833.27	5852.25	2759.6	12850.42
roa(%)	8562	1.25	1.30	0.54	2.25
nig(%)	8562	0.04	0.04	−0.09	0.16
ic	8562	12.21	6.11	2.87	12.72
qr(%)	8562	115.77	102.46	74.46	138.84
cta(%)	8562	9.10	5.99	2.45	12.62
taa(%)	8562	70.98	55.73	21.38	93.34
sg(%)	8562	2.66	2.09	0.19	4.30
lev(%)	8562	64.93	62.82	52.21	75.03
reta(%)	8562	21.88	23.86	7.62	41.07
Market-based variables					
dtd	8562	8.29	7.62	4.41	11.57
sret(%)	8562	11.69	15.52	−13.5	41.43
svol	8562	0.34	0.28	0.20	0.40
Macroeconomic variables					
trate(%)	8562	1.72	1.15	0.12	3.04
iret(%)	8562	7.48	11.71	−2.68	20.83
mret(%)	8562	4.13	5.48	−9.03	15.61

We obtain a fairly priced credit premium by equating the above two equations and making a rearrangement:

$$CS = \frac{\mathrm{E}\left[\int_0^T \exp\left(-\int_0^\tau r_t dt\right) s(0,\tau)\lambda_\tau(1-\rho_\tau)d\tau\right]}{\mathrm{E}\left[\int_0^T \exp\left(-\int_0^\tau r_t dt\right) s(0,\tau)d\tau\right]}. \tag{3}$$

Following Das et al. (2009), we assume that the default intensity is constant, conditioning on the given state vector X in each quarter; the recovery rate is constant over time ($\rho_\tau = \rho$); and the default intensity is of the following functional form:

$$\lambda = exp(B^T X),$$

where $B \equiv [\beta_1, \ldots, \beta_k]$ is a vector of coefficients and $X \equiv [x_1, \ldots, x_k]$ is a vector of explanatory variables, including text, accounting, market, macroeconomic, and other variables. Thus, the CDS spread in a discrete-time pricing model (time interval h; hereafter, a quarter) follows

$$\begin{aligned} CS &= \frac{E\left[\sum_{j=1}^n e^{-z_j jh}(1-\rho_j)e^{-\lambda_j(j-1)h}(1-e^{-\lambda_j h})\right]}{hE\left[\sum_{j=1}^n e^{-z_j jh} e^{-\lambda_j(j-1)h}\right]} \\ &= \frac{(1-\rho)(1-e^{-\lambda h})}{h}, \end{aligned} \tag{4}$$

where z_j is the zero-coupon discount rate for quarter j. Taking the logarithm of CS, we can obtain an approximate linear estimation equation in the following form:

$$\log(CS) \sim \log\left(\frac{1-\rho}{h}\right) + B^T X h.$$

This expression describes a linear relationship between the logarithm of credit spread and explanatory variables; hence, our empirical models ran panel regressions (Eq. (5)) as follows:[5]

$$\begin{aligned} \log(CS)_{it} &= \beta_1^T NEWS_{it} + \beta_2^T RISK_{it} + \beta_3^T ACC_{it} + \beta_4^T MKT_{it} \\ &\quad + \beta_5^T OTHERS_{it} + \alpha_i + d_t + \epsilon_{it}, \end{aligned} \tag{5}$$

where β_i are the coefficient vectors of the corresponding explanatory variables, including news variables ($NEWS \equiv [lnarticle, neg]^T$), corporate risk factor disclosure variables ($RISK \equiv [lall\_all, risk\_idio, risk\_sys, risk\_fin, risk\_tax, risk\_lr]^T$), accounting-based variables ($ACC \equiv [lta, roa, nig, c1, c2, c3, c4, qr, cta, taa, sg, lev, reta]^T$), market-based variables ($MKT \equiv [dtd, sret, svol]^T$), macroeconomic variables, and other control variables ($OTHERS \equiv [trate, mret, iret, invgrd, lqn]^T$). We also included a firm fixed effect α_i and a time fixed effect d_t. The variable ϵ_{it} denotes white noise. We present the empirical results of the above model in the following section.

4. Empirical results

We summarize the correlation and univariate analysis in Section 4.1. Section 4.2 presents the results of the multivariate analysis. We report additional robustness checks in Section 4.3.

4.1. Correlation and univariate analysis

Table 4 lists the pairwise correlations of the CDS spread (as a logarithm) and its explanatory variables. Credit spread was highly correlated with the distance to default (*dtd*) and the investment grade dummy (*invgrd*). In addition, a high credit risk (high CDS

[5] A Hausman test showed that the fixed-effect model was preferred to a random-effect model for our regression.

Table 4
Correlation matrix. This table lists the pairwise correlations between text variables and other explanatory variables. The definitions of these variables are given in Table 1. Asterisks denote statistical significance levels: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

	lcs	*lnarticle*	*neg*	*lrisk_all*	*risk_sys*	*risk_idio*	*risk_fin*	*risk_tax*	*risk_lr*
lcs	1								
lnarticle	−0.18***	1							
neg	0.14***	0.06***	1						
lrisk_all	0.06***	−0.17***	0.05***	1					
risk_idio	−0.05***	0.09***	0.04***	0.30***	1				
risk_sys	0.01	−0.16***	0.01	0.87***	−0.08***	1			
risk_fin	0.44***	−0.14***	−0.01	0.20***	−0.22***	0.10***	1		
risk_tax	−0.07***	−0.11***	−0.05***	0.14***	−0.07***	0.00	0.09***	1	
risk_lr	−0.07***	−0.06***	0.17***	0.27***	0.00	0.13***	−0.13***	−0.10***	1
lta	−0.29***	0.56***	0.08***	0.18***	0.03***	0.20***	−0.15***	−0.06***	0.18***
roa	−0.51***	0.09***	−0.12***	−0.07***	−0.01	−0.02**	−0.27***	0.06***	0.02**
nig	−0.05***	−0.02	−0.02*	0.00	−0.01	0.01	0.00	−0.02	0.01
c1	−0.63***	0.12***	−0.10***	−0.05***	0.11***	−0.03***	−0.39***	0.03***	0.01
c2	−0.51***	0.21***	−0.06***	−0.08***	0.13***	−0.08***	−0.34***	0.02**	−0.01
c3	−0.40***	0.24***	−0.04***	−0.02*	0.14***	−0.03***	−0.26***	0.04***	−0.01
c4	−0.25***	0.25***	−0.03***	0.01	0.05***	0.02**	−0.15***	0.03***	−0.04***
qr	−0.02*	−0.03***	0.04***	0.07***	0.25***	−0.02	−0.18***	−0.01	0.06***
cta	0.00	0.22***	0.04***	−0.04***	0.30***	−0.14***	−0.19***	−0.04***	−0.02*
taa	−0.13***	0.02*	0.01	−0.08***	0.16***	−0.13***	−0.22***	−0.05***	0.10***
sg	−0.07***	−0.03***	−0.04***	0.01	0.01	0.01	−0.04***	0.00	−0.01
lev	0.41***	−0.04***	0.05***	−0.06***	−0.16***	−0.07***	0.34***	0.00	−0.01
reta	−0.48***	0.09***	−0.09***	−0.05***	−0.01	0.02	−0.29***	0.05***	−0.03***
dtd	−0.73***	0.18***	−0.12***	−0.08***	0.06***	−0.08***	−0.33***	0.12***	0.06***
sret	−0.12***	−0.04***	−0.08***	0.00	−0.03***	0.01	0.03***	0.00	0.02
svol	0.63***	−0.03***	0.15***	0.04***	−0.01	0.03**	0.24***	−0.08***	−0.08***
trate	−0.31***	0.02	−0.10***	−0.16***	−0.10***	−0.14***	−0.07***	0.12***	−0.02*
iret	−0.11***	−0.03***	−0.08***	0.07***	−0.03***	0.06***	0.02**	0.10***	0.05***
mret	−0.12***	−0.03**	−0.07***	0.01	−0.01	−0.03**	0.03***	0.13***	0.04***
invgrd	−0.69***	0.14***	−0.07***	0.04***	0.08***	0.09***	−0.41***	0.04***	0.05***
lqn	−0.30***	0.21***	−0.03**	−0.09***	−0.09***	−0.06***	−0.10***	0.07***	0.05***

Table 5
Credit spread quartiles. This table presents the mean and median CDS spreads by the quartile of news variables and risk disclosure variables (definitions are in Table 1).

Variable		Quartile			
		1 (low)	2	3	4 (high)
News variables					
narticle	Mean	190.03	177.03	178.62	198.36
	Median	94.74	75.00	70.52	50.16
	N	3335	1475	2027	1725
neg	Mean	138.02	146.84	188.19	274.00
	Median	70.97	63.32	72.47	100.00
	N	2140	2141	2140	2141
Risk disclosure variables					
risk_all	Mean	166.28	194.31	206.06	180.43
	Median	65.07	69.67	79.33	81.88
	N	2140	2141	2140	2141
risk_idio	Mean	225.91	186.17	173.88	161.13
	Median	77.14	78.70	75.87	69.75
	N	2140	2141	2140	2141
risk_sys	Mean	202.17	170.21	180.42	194.27
	Median	92.65	63.00	70.35	80.88
	N	2140	2141	2140	2141
risk_fin	Mean	78.32	132.32	170.85	365.52
	Median	46.51	63.55	88.64	171.52
	N	2140	2141	2140	2141
risk_tax	Mean	221.32	181.11	174.16	170.50
	Median	77.67	79.35	75.87	69.22
	N	2140	2141	2140	2141
risk_lr	Mean	176.15	240.63	202.23	128.07
	Median	73.25	86.36	77.76	65.63
	N	2140	2141	2140	2141

Table 6
Multivariate analysis. This table displays the results of regressing the log of the CDS spread on news and risk disclosure variables (i.e., text variables), controlling for accounting-based, market-based, macroeconomic, and other variables. All models include control variables: firm and time-fixed effects. Model (1) includes news variables but excludes risk disclosure variables. Model (2) includes news variables and overall risk disclosure intensity variable. Model (3) includes news variables and risk-subcategory variables. Student's t-statistics are in parentheses and adjusted for clustering in firm and time. Asterisks denote statistical significance levels: **$p < 0.05$, ***$p < 0.01$.

Independent variables	Dependent variable: *lcs*		
	(1)	(2)	(3)
lnarticle	0.0533***	0.0521***	0.0543***
	(3.18)	(3.16)	(3.35)
neg	0.0124**	0.0119**	0.0132***
	(2.48)	(2.39)	(2.64)
lrisk_all		0.0586***	
		(3.65)	
risk_idio			0.0285
			(0.96)
risk_sys			0.0642**
			(2.48)
risk_fin			0.1650**
			(2.39)
risk_tax			0.0027
			(0.04)
risk_lr			−0.0625
			(−0.79)
Control variables	Included	Included	Included
Firm fixed effect	Included	Included	Included
Time fixed effect	Included	Included	Included
Adjusted R-square	0.898	0.898	0.898
Observations	8562	8562	8562

spread) was associated with a high overall risk disclosure intensity (*lrisk_all*), high financial risk disclosure (*risk_fin*), high book leverage (*lev*), and high equity volatility (*svol*). Among the correlations between news variables and other explanatory variables, that between firm size (*lta*) and news coverage (*lnarticle*) was the highest. Additionally, news coverage (*lnarticle*) was positively correlated with negative sentiment (*neg*).

The positive correlation between CDS spreads and the overall risk disclosure intensity (*lall_risk*) in corporate filings provided initial evidence that corporate disclosures are informative. In addition, the relative amount of financial risk disclosure was positively correlated with the CDS spread, while the idiosyncratic risk, tax risk, and litigation risk were negatively correlated with the CDS spread. The signs of correlation coefficients imply how the market might interpret these risk subcategories.

Table 5 summarizes the univariate analysis. We divided our sample into quartiles for each news variable and risk disclosure variable (quartile 1 represents the lowest value). For each quartile, we computed the mean and median CDS spreads. The result shows that the median CDS spread gradually decreased as the news coverage increased. The mean CDS spread also has a decreasing trend with respect to news coverage, except in the last quartile, which has a slightly higher mean CDS spread compared to that of the third quartile. One possible explanation is that observations in this quartile tend to have more negative news sentiment, as implied by the positive correlation between news coverage and negative news sentiment, which drives up the CDS spreads of the upper tail in this quartile. As for the news sentiment variable, we found that a stronger negativity in news tone accompanies higher mean and median credit spreads from quartiles 2 to 4.

Among the risk disclosure variables, the mean CDS spread shows an upward trend along with increasing risk disclosure intensity (*lall_risk*) from the first to the third quartile. The mean CDS spread of the fourth quartile is slightly smaller compared to that of the third quartile. The median CDS spread increases with risk disclosure intensity. Financial risk (*risk_fin*) shows a strong upward trend along with the increasing mean and median CDS spreads from the first to the fourth quartile. The mean CDS spreads decrease from the first quartile to the fourth quartile of idiosyncratic risk (*risk_idio*) and tax risk (*risk_tax*). Systematic risk (*risk_sys*) and litigation risk (*risk_lr*) show a non-monotonic relation with the CDS spread. The univariate analysis provides initial evidence that news and some risk disclosure variables are informative. We report the multivariate analysis in the next section.

4.2. Multivariate analysis

To further investigate the relationships between text variables and the credit risk for debt issuers, we applied panel regressions considering firm and time-fixed effects. We controlled for accounting-based, market-based, macroeconomic, and other variables in all empirical models. The definitions of the control variables are provided in Table 1.

Table 6 shows the results of regressing the CDS spread on text and other control variables. Our sample contained 8562 data points from 491 issuers. Because of potential issues due to having repeated observations from the same issuers in the regression, this could have resulted in correlated residuals from the same issuers.[6] Thus, all reported *t*-values were adjusted for clustering in firm and time using two-way (firm and time) cluster-robust standard errors,[7] following the method of Thompson (2011) and Petersen (2009). We omitted the control variables owing to space constraints. Complete tables can be found in Appendix C.

Table 7
Multivariate analysis result of text variables on option-implied volatility and probability of informed trading (PIN). Student's *t*-statistics are in parentheses, adjusted for clustering in firm and time. Asterisks denote statistical significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Independent variables	Dependent variable: *impl_volatility* (1)	Dependent variable: *PIN* (2)
lta	−0.0050	−0.0236***
	(−0.31)	(−3.64)
lnarticle	0.0256***	−0.0035**
	(3.80)	(−2.46)
neg	0.0102***	−0.0003
	(3.75)	(−0.53)
risk_idio	0.0149	0.0019
	(1.57)	(0.71)
risk_sys	−0.0115	−0.0000
	(−1.53)	(−0.00)
risk_fin	0.0640***	0.0024
	(2.93)	(0.42)
risk_tax	−0.0225	0.0021
	(−1.07)	(0.44)
risk_lr	−0.0367*	−0.0014
	(−1.76)	(−0.25)
_cons	0.3450***	0.3000***
	(2.59)	(4.49)
Firm fixed effect	Included	Included
Time fixed effect	Included	Included
Adjusted *R*-square	0.768	0.526
Observations	6058	6465

Model (1) of Table 6 includes news coverage and negative sentiment in addition to control variables. News coverage had a significantly positive coefficient (0.0533; *p*-value <0.01). The positive relationship between news coverage and credit risk serves as evidence supporting Hypothesis 1b. This implies that more news increases firm credit risk because of media-biased and media-induced sentiment, which cause uncertain asset value and hence lead to higher credit risk. We will examine this hypothesis further in the following subsection.

We note that the correlation coefficient between news coverage and CDS spreads is negative (−0.18), whereas the regression coefficient of news coverage is significantly positive. One reason for sign reversal might be the large correlation between firm size and news coverage (0.56 in Table 4). Large firms generally have more news coverage. We controlled for firm size in Model (1) of Table 6, causing the sign of news coverage to flip. To further confirm our argument, we conducted a double sort by firm size and news coverage. The mean CDS spread for small firms increased from 215.76 to 320.23 from the first news quartile (lowest) to the fourth quartile (highest). Similarly, for large firms, the mean CDS spread increased from 129.18 to 181.38 from the first quartile to the fourth quartile. We included the double sort results in Table C.2 of Appendix C.

The negative news sentiment was also significantly positive (0.0124; *p*-value <0.05). Our results show that pessimism in news articles can be linked to a worsening of creditworthiness in debt issues, supporting Hypothesis 2.

Models (2) and (3) of Table 6 include risk disclosure variables in the empirical model. The overall volume of risk disclosure (*lrisk_all*) (see Model (2) of Table 6) was significantly positive (0.0586; *p*-value <0.01), suggesting that more risk disclosure increases CDS spreads. This result is consistent with the value-relevant story in Hypothesis 3, which states that more risk disclosure from a firm

[6] We thank the reviewer (Anonymous Referee #2) for pointing this out.
[7] We thank Professor Petersen for kindly providing the codes on his website.

Table 8
Regression results using sub-samples and alternative specifications. This table presents four additional tests for robustness. Model (1) displays results using a subsample after the SEC started to mandate risk factor disclosures in 2005. Text variables were computed from all content in corporate filings (both 10-Q and 10-K). Model (2) shows the results of computing risk disclosure variables on a subsample containing text from Item 1A (risk factors) in the 10-Ks mandated by SEC since 2005. Model (3) lagged the accounting-based and risk disclosure variables by one quarter to accommodate for the delayed announcement of corporate filings. Model (4) adopts the same specification as Model (3) but instead uses the same subsample used by Model (2). Student's t-statistics are in parentheses, adjusted for clustering in firm and time. Asterisks denote statistical significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Independent variables	Dependent variable: *lcs*			
	(1)	(2)	(3)	(4)
lnarticle	0.0258	0.0340	0.0566***	0.0446
	(1.63)	(0.74)	(3.40)	(1.39)
neg	0.0115**	0.0149	0.0106**	0.0130
	(2.03)	(0.87)	(2.14)	(0.74)
risk_idio	0.0104	−0.0429	0.0414	−0.0623
	(0.34)	(−0.83)	(1.32)	(−1.45)
risk_sys	0.0429	0.0081	0.0616**	−0.0086
	(1.63)	(0.18)	(2.34)	(−0.26)
risk_fin	0.1810**	0.1530*	0.1430**	0.1750**
	(2.52)	(1.75)	(2.06)	(2.20)
risk_tax	−0.0123	−0.0228	0.0405	−0.2010
	(−0.18)	(−0.25)	(0.72)	(−1.33)
risk_lr	0.0589	0.0786	−0.0571	0.1560***
	(0.88)	(1.09)	(−0.79)	(2.66)
Control variables	Included	Included	Included	Included
Firm fixed effect	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Adjusted R-square	0.909	0.898	0.898	0.907
Observations	6313	1435	8071	1231

is associated with the credit risk of that firm. To discover which types of risk affect an issuer's credit risk, we included five subcategories of risk factors in Model (1). Remarkably, we found financial risk (*risk_fin*) to have a significantly positive impact on CDS spreads (0.165; p-value <0.05; see Model (3)). This result suggests that, after controlling for accounting and market-based variables, self-disclosure regarding financial uncertainty is valued by credit market investors, hence validating Hypothesis 3a. Additionally, systematic risk (*risk_sys*) is also positively linked to the credit risk of issuers. Specifically, the market also values the amount of self-disclosed systematic risks, supporting Hypothesis 3d. We find no evidence supporting Hypotheses 3b, 3c, and 3e because the coefficients of tax, litigation, and idiosyncratic risk factors lack statistical significance. In summary, these results suggest that the overall risk disclosure intensity and risk subcategory disclosures, including financial risk and systematic risk, are valuable to the credit market.

4.3. Influence channels of news coverage

Our multivariate analysis supports Hypothesis 1b, and suggests that, for news coverage, its influence on credit risk through increased information uncertainty dominates its effect through reduced information asymmetry. We further examined these conjectures by investigating how news coverage influences option-implied volatility and the probability of informed trading (PIN).

Option-implied volatility indicates equity uncertainty in the future, and it is a proxy for information uncertainty. We adopted 30-day deep out-of-the-money put options (delta = −20%) in the last trading days of a quarter obtained from OptionMetrics. The value of such options is most sensitive to the left tail of risk-neutral stock return distribution, as is the CDS spread (Cao et al., 2010).

Table 9
Regression results before, during, and after the financial crisis period. This table presents regression results in three sub-periods. The financial crisis period is defined by NBER from December 2007 to June 2009 (Aboody et al., 2014). Student's t-statistics are in parentheses and adjusted for clustering in firm and time. Asterisks denote statistical significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$.

Independent variables	Dependent variable: *lcs*		
	(1) before	(2) during	(3) after
lnarticle	0.0327*	0.0569**	−0.0073
	(1.89)	(2.37)	(−0.34)
neg	0.0087	−0.0240	0.0146**
	(1.26)	(−1.43)	(2.32)
risk_idio	−0.0386	−0.0832	0.0197
	(−1.16)	(−1.32)	(0.50)
risk_sys	0.0449	0.0740	0.0340
	(1.34)	(1.40)	(0.79)
risk_fin	0.1200	0.0457	0.2650**
	(1.49)	(0.33)	(2.42)
risk_tax	0.0705	−0.1580	−0.0633
	(0.92)	(−0.80)	(−0.95)
risk_lr	−0.0451	−0.1690	0.1190
	(−0.55)	(−1.39)	(1.15)
Control variables	Included	Included	Included
Firm fixed effect	Included	Included	Included
Time fixed effect	Included	Included	Included
Adjusted R-square	0.918	0.928	0.914
Observations	4628	1175	2759

We regressed implied volatility on news coverage, news sentiment, and risk disclosure variables, and controlled for firm size. As reported in Table 7, the news coverage coefficient is significantly positive (0.0256; p-value <0.1), supporting our conjecture that news coverage increases CDS spreads by increasing information uncertainty, which leads to higher anticipated asset volatilities, and hence high credit risk. In addition, we found that negative news sentiment and financial risk factor disclosure have positive effects on option-implied volatility as well. In summary, the regression result suggests that more news coverage, more negative news sentiment, and more disclosure about financial risk factor are related to higher forward-looking uncertainty of equity value and hence cause increasing firm credit risk.

Another related question is the effect of news coverage on PIN,[8] an information asymmetry measure developed by Easley et al. (1996) to estimate the probability of (private) information-based trading on listed stocks. We regressed PIN on news and risk disclosure variables, controlled for firm size. As reported in Table 7, we found that news coverage reduced information asymmetric problems for an issuer. On the other hand, news sentiment and risk disclosure variables had no effect on PIN.

In summary, news coverage could transmit to firm credit risk through information uncertainty and information asymmetry. News coverage induces information uncertainty and reduces information asymmetry at the same time. The net effect of news coverage on issuer credit risk is positive. Negative news sentiment and disclosure of financial risk factor could increase information uncertainty. None of the evidence indicates that negative sentiment in news articles and risk factor disclosures in corporate filings will reduce information asymmetry.

4.4. Robustness checks

We conducted several robustness checks to assess issues related to SEC risk disclosure regularization changes, alternative risk disclosure variable definitions, and the delay of filings

[8] We adopted the pre-computed PIN values that Dr. Stephen Brown kindly provided on his website.

(filing available after the end of a quarter). Additionally, we also evaluated the impacts of text information on credit risk before, during, and after the financial crisis.

SEC has mandated a risk-factor item (Item 1A) since 2005. We assessed the effect of the regularization change by the two following checks. We first restricted our sample to observations from 2005 to 2013. In this check, the risk disclosure variables were to be computed from the whole filings of 10-K and 10-Q. Model (1) of Table 8 reports the results. The disclosure of financial risk factor remains significantly positive, while systematic risk becomes insignificant. In the second check, we computed risk disclosure variables using only the text in Item 1A of the 10-K filings. The number of observations fell to 1435. As reported in Model (2) of Table 7, the financial risk factor disclosure remained significantly positive but the significance is reduced.

To account for the delays in corporate filings, we lagged accounting variables and risk disclosure variables by one quarter so that all variables were known to the market. This problem is relevant to trading strategy implementation, but is not an issue in uncovering the determinants of CDS spreads per se (Das et al., 2009). As reported in Model (3) of Table 8, the result remains qualitatively the same as our main result (Model (3) of Table 6). We repeated the same check with the risk disclosure variables computed from the text of Item 1A in 10-K filings. As reported in Model (4) of Table 8, the result is similar to Model (2) of Table 8, except that the coefficient of litigation risk factor disclosure becomes significantly positive.

In addition, we assessed the effect of financial crisis by sub-period analysis. We adopted the NBER definition of financial crisis period (December 2007 to June 2009), and divided our sample into three sub-periods: before the financial crisis, during the financial crisis, and after the financial crisis. We are interested to know (1) whether our main result was driven by a specific sub-period, and (2) how the results change over time.

As reported in Table 9, the results showed that news coverage has significant positive impacts on firm credit risk before and during the financial crisis. Negative news sentiment affects the credit risk of an issuer, especially after the financial crisis. We noted that the average negative word ratio (in an unreported table) was 1.88% before the crisis, 2.09% during the crisis, and 1.97% after the crisis. In addition, financial risk factors disclosed by a debt issuer only matter to its creditworthiness after the financial crisis.

In summary, we find that financial risk factor disclosure continues to significantly and positively affect issuer credit risk after the 2005 SEC risk disclosure regularization change, and in the subsample where only Item 1A in 10-K filings are adopted. Financial risk factor disclosure is also robust to the alternative specification that accommodates delayed publication of corporate filings. News sentiment remains strong for two out of the four alternative specifications. Further, news coverage is robust to the delayed announcement specification, but lacks significance when using other subsamples. On the other hand, sub-period analysis shows that news coverage increases firm credit risk before and during the financial crisis, while negative news sentiment and financial risk factor disclosure level up the firm credit risk after the financial crisis.

5. Conclusion

Qualitative information, including the text from newspapers and corporate filings, play a distinct role in determining issuer credit risk, as compared to quantitative information adopted previously. We captured qualitative information using variables including news coverage, negative news sentiment, risk disclosure intensity, and five risk subcategories. Further, following the literature, we included accounting- and market-based variables as control variables.

Our study offers several prominent findings. First, higher amounts of news coverage lead to higher credit risk of debt issuers. This implies that more news enhances information uncertainty through media-induced investor sentiment. Second, negative news sentiment is associated with high issuer credit risks. Third, risk factor disclosures in SEC corporate filings are informative to the CDS market. An increase in corporate risk disclosure intensity as measured by the amount of risk keywords is linked to higher credit risks. Finally, among the five types of risk disclosure considered, financial risk has the strongest effect in driving up credit risk.

We further investigated how news coverage influences issuer credit risks. We adopted option-implied volatility as a proxy for information uncertainty, and PIN as a proxy for information asymmetry. Empirical analysis showed that news coverage is positively associated with option-implied volatility but negatively associated with PIN. In other words, news coverage increases information uncertainty and decreases information asymmetry simultaneously. Our main finding displayed the net effect that is dominated by information uncertainty.

In conclusion, our results show that text information from news articles and SEC corporate filings is useful for financial decision makers in credit markets. Moreover, CDS spreads depend not only on traditional variables, such as accounting and market variables, but also on qualitative information obtained from public information sources. We also find that the business press plays the role of an information intermediary and corporate filings are important information transmission channels in capital markets. To evaluate a borrower's credit risk in the real world, lenders can benefit from exploiting such text information.

Acknowledgements

We are grateful to Asia University Taiwan (grant number: 102-asia-35), the National Science Council of Taiwan (grant numbers: NSC99-2410-H-468-032, NSC99-2410-H-002-261, and NSC100-2410-H-002-025-MY3), and the Ministry of Science and Technology of Taiwan (grant numbers: MOST103-2410-H-468-005 and MOST103-2410-H-002-110-MY3) for financial support.

Appendix A. News variable construction procedure and key modules and tools for text analysis

The first step was to identify the named entity in an article. We adopted the Stanford Named Entity Recognizer (NER) (http://nlp.stanford.edu/software/CRF-NER.shtml) to process the title and full text of an article. Because we only needed to identify company names, we adopted the three-class model trained on both the Conference on Natural Language Learning (CoNLL) and Message Understanding Conference (MUC) datasets. This step recorded all the "Organizations" identified by the NER module.

The second step mapped the identified company names to company identifiers. We adopted PERMCO from the Center for Research in Security Prices (CRSP) as the company identifier. PERMCO's advantage lies in its ability to be easily linked to existing financial market data. We obtained a comprehensive list of public company names, their PERMCOs and corresponding valid date ranges from the STOCKNAMES tables in the CRSP dataset. Each of the company names identified in the first step was matched against the company names in the STOCKNAMES table valid on the article's publication date. The company names in the STOCKNAMES tables did not contain punctuation and included extra spaces for acronyms. We developed a set of rules to accommodate these systematic differences and computed similarities between the identified company names and the candidates in STOCKNAMES. An identified company name was mapped to a PERMCO if the similarity was high and the strings used to compute similarity were more than two words in length.

The last step was to identify the main company associated with an article. We assigned an article to a PERMCO corresponding with the company name that appeared most frequently in the article. We dropped articles that were less than 50 words in length. Moreover, articles that were mapped to more than ten PERMCOs were excluded in the subsequent analyses because these articles often contained short updates from many companies and should not be mapped to a single company. After mapping an article to its main company, we computed the media coverage of a firm by counting the number of WSJ news articles that were mapped to the firm.

The following table summarizes the key modules and tools used in our text processing programs

Key modules and tools	Purpose	Internet resource
Python 2.7	Programming language for text analysis	https://www.python.org/
Java 1.7	Programming language for text analysis	http://www.oracle.com/technetwork/java/javase/downloads/index.html
Library re (Python)	Sentiment and risk disclosure variable computation	https://docs.python.org/2/library/re.html
Library ClientCookie (Python)	Company filing download	http://wwwsearch.sourceforge.net/old/ClientCookie/
Library BeautifulSoup (Python)	HTML filing processing	http://www.crummy.com/software/BeautifulSoup/
Library NLTK (Python)	Natural language processing functions	http://www.nltk.org/
PostgreSQL 9.1	Data management	http://www.postgresql.org/
Stanford NER	Company name identification	http://nlp.stanford.edu/software/CRF-NER.shtml

Appendix B. Additional text analysis results

See Tables B.1–B.6.

Appendix C. Complete tables and additional empirical results

See Tables C.1 and C.2.

Table B.1
List of top ten news articles with the highest NEG score.

Date	Company Name (PERMCO)	NEG (%)	Title	Full Text (First Sentence)
3/12/2008	Constellation Energy Group (20258)	15.8	Constellation Energy Group: Unit Fined $7 Million On Release Violations	A Constellation Energy Group unit was fined nearly $7 million in civil and disgorgement penalties for violations of natural-gas capacity-release laws...
8/22/2003	Safeway Inc. (22765)	15.4	Business Brief – Safeway Inc.: Yucaipa Cos. Is Countersued, Alleging Interference in Sale	Supermarket operator Safeway Inc. filed a countersuit against private-investment company Yucaipa Cos. of Los Angeles, claiming breach of contract, fraud and intentional interference...
8/18/2008	AMR Corp. (20010)	14.1	Southwest Fails to Get FAA Fine Reduced	Southwest Airlines Co. failed to convince the U.S. government to reduce a $10.2 million fine proposed for alleged safety violations...
2/22/2006	Micron Technology Inc. (7065)	13.8	Micron Technology Inc.: Latest Suit Against Rambus Cites Racketeering Laws	Micron Technology Inc. fired another salvo in a legal war against Rambus Inc., accusing the Silicon Valley company of violating racketeering laws as part of a plan to enforce patents on memory-chip technology
11/4/2003	Newmont Mining Corp. (21286)	13.5	Industrial Brief – Newmont Mining Corp.: Subpoena Asks for Documents Tied to a Shareholder Dispute	Newmont Mining Corp. said it received a subpoena from the Justice Department requiring it to produce documents related to a shareholder dispute that took place from 1994 to 1999...
9/27/2001	Intel Corp. (2367)	13.3	Business Brief – Intel Corp.: Four Patent Suits Are Filed Against Via Technologies	Intel Corp., of Santa Clara, Calif., filed four patent-infringement lawsuits in three countries against Taiwan-based Via Technologies Inc.
7/8/2003	Microsoft Corp. (8048)	13.2	Technology Brief: Microsoft Corp.	Massachusetts is investigating whether Microsoft Corp. retaliated against a computer maker for promoting a rival operating system, a possible violation of the company's settlement with the Bush administration and 18 other states
8/20/2004	Best Buy Co. (7506)	13.1	Best Buy Co.: Ohio Files Lawsuit Claiming Unfair, Deceptive Practices	The state of Ohio sued Best Buy Co., accusing the consumer- electronics retailer of engaging in unfair and deceptive business practices
9/13/2006	Microsoft Corp. (8048)	13.0	Paltalk Holdings Inc.	Paltalk Holdings Inc., a video-chat company, filed a patent- infringement lawsuit against Microsoft Corp., claiming that videogame playing on Microsoft's Xbox console infringes Paltalk's patents
7/11/2003	Rite Aid Corp. (21515)	12.4	Retail Brief – Rite Aid Corp.: Former Executive Pleads Guilty To Obstructing-Justice Charge	Another former Rite Aid Corp. executive pleaded guilty to conspiracy to obstruct justice, a charge related to the government's investigation of an accounting-fraud scandal at the drugstore chain

Table B.2
List of representative sentences from 10-K filings with the highest *risk_sys*. The representative sentences were selected by locating three consecutive sentences with the highest average percentage of words marked as systematic risk. Words classified to systematic risk, financial risk, idiosyncratic risk, tax risk, and litigation risk were marked in superscripts as SysRisk, FinRisk, IdioRisk, TaxRisk, and LRRisk. Multiple filings from the same company were only listed once.

Company name: NUCOR CORP	CIK: 73309
Type: 10-K	Period of report: 2008-12-31
Date filed: 2009-02-26	Document score [SysRisk]: 71.18%

- This excess capacity[SysRisk] results in manufacturers in certain countries exporting significant amounts of steel[SysRisk] and steel[SysRisk] products at prices[SysRisk] below their cost of production[IdioRisk]
- These imports, which are also affected by demand[SysRisk] in the domestic market[SysRisk], international currency conversion rates and domestic and international government actions, can result in downward pressure on steel[SysRisk] prices[SysRisk], which could materially adversely affect our business, results of operations, financial condition[FinRisk] and cash flows
- Overcapacity in China, the world's largest producer and consumer of steel[SysRisk], has the potential to result in a further increase in imports of low-priced, unfairly traded steel[SysRisk] and steel[SysRisk] products to the United States

Company name: SEMPRA ENERGY	CIK: 1032208
Type: 10-K	Period of report: 2010-12-31
Date filed: 2011-02-24	Document score [SysRisk]: 69.73%

- External factors such as weather[IdioRisk], the price of electricity[SysRisk], electric deregulation[LRRisk], the use of hydroelectric power, development of renewable energy[SysRisk] resources, development of new natural gas[SysRisk] supply sources, and general economic[SysRisk] conditions can also result in significant shifts in demand[SysRisk] and market[SysRisk] price
- The Sempra Utilities face competition[SysRisk] in the residential and commercial customer markets[SysRisk] based on customers' preferences for natural gas[SysRisk] compared with other energy[SysRisk] products
- In the noncore industrial market[SysRisk], some customers are capable of securing alternate fuel[SysRisk] supplies from other suppliers[IdioRisk] which can affect the demand[SysRisk] for natural gas[SysRisk]

Company name: NOBLE ENERGY INC	CIK: 72207
Type: 10-K	Period of report: 2002-12-31
Date filed: 2003-03-11	Document score [SysRisk]: 68.52%

- Volatility and level of hydrocarbon commodity prices
- Historically, natural gas[SysRisk] and crude oil[SysRisk] prices[SysRisk] have been volatile
- These prices[SysRisk] rise and fall based on changes in market[SysRisk] supply and demand[SysRisk] fundamentals and changes in the political, regulatory[LRRisk] and economic[SysRisk] climates and other factors that affect commodities[SysRisk] markets[SysRisk] generally and are outside of Noble Energy's control

Table B.3
List of representative sentences from 10-K filings with the highest *risk_fin*. The representative sentences were selected by locating three consecutive sentences with the highest average percentage of words marked as financial risk. Words classified to systematic risk, financial risk, idiosyncratic risk, tax risk, and litigation risk were marked in superscripts as SysRisk, FinRisk, IdioRisk, TaxRisk, and LRRisk. Multiple filings from the same company were only listed once.

Company name: RITE AID CORP	CIK: 84129
Type: 10-K	Period of report: 2003-03-01
Date filed: 2003-05-02	Document score [FinRisk]: 39.91%

- This action resulted in a reduction of outstanding capital lease[FinRisk] obligations[FinRisk] of $850,792
- Accordingly, the Company recognized a loss on lease[FinRisk] modifications of $21,882 in fiscal 2002, and recorded a net deferred gain of $168,483, which will be amortized over the remaining noncancellable lease[FinRisk] terms
- In addition, the Company repaid certain obligations[FinRisk] totaling $16,467 related to leasehold improvements[FinRisk]

Company name: AMERICAN TOWER CORP/MA/	CIK: 1053507
Type: 10-K	Period of report: 2004-12-31
Date filed: 2005-03-30	Document score [FinRisk]: 37.75%

- We plan to continue to reduce our overall indebtedness[FinRisk] in 2005 with cash flow from operations, and may opportunistically further reduce indebtedness[FinRisk] and interest expense through future[SysRisk] capital market[SysRisk] or strategic transactions
- In 2004, we generated sufficient cash flow from operations to fund our capital expenditures[FinRisk] and cash interest obligations[FinRisk]
- We believe our cash generated from operations for the year ending December 31, 2005 also will be sufficient to fund our capital expenditures[FinRisk] and our cash debt service (interest and principal repayments) obligations[FinRisk] for 2005

Table B.4
List of representative sentences from 10-K filings with the highest risk_idio. The representative sentences were selected by locating three consecutive sentences with the highest average percentage of words marked as idiosyncratic risk. Words classified to systematic risk, financial risk, idiosyncratic risk, tax risk, and litigation risk were marked in superscripts as SysRisk, FinRisk, IdioRisk, TaxRisk, and LRRisk. Multiple filings from the same company were only listed once.

Company name: BEST BUY CO INC	CIK: 764478
Type: 10-K	Period of report: 2001-03-03
Date filed: 2001-06-01	Document score [IdioRisk]: 75.29%

- In addition, all Best Buy stores feature a configure-to-order process for personal computers that enables computer buyers to custom-order a computer system from such vendors[IdioRisk] as Compaq and Hewlett–Packard
- Best Buy spends approximately 3% of store sales on advertising[IdioRisk], including the weekly distribution[IdioRisk] of approximately 48 million newspaper inserts
- The Company is reimbursed by vendors[IdioRisk] for a substantial portion of advertising[IdioRisk] expenditures through cooperative advertising[IdioRisk] arrangements

(*continued on next page*)

Company name: UNISYS CORP — CIK: 746838
Type: 10-K — Period of report: 2004-12-31
Date filed: 2005-02-18 — Document score [IdioRisk]: 74.68%

- Principal competitors[SysRisk] are systems[IdioRisk] integrators, consulting and other professional services firms, outsourcing providers, infrastructure services providers, computer hardware manufacturers and software[IdioRisk] providers
- Unisys competes primarily on the basis of service, product[IdioRisk] performance, technological innovation[IdioRisk], and price
- Unisys believes that its continued investment in engineering and research and development[IdioRisk], coupled with its marketing[IdioRisk] capabilities, will have a favorable impact on its competitive position

Company name: SAKS INC — CIK: 812900
Type: 10-K — Period of report: 2003-02-01
Date filed: 2003-04-30 — Document score [IdioRisk]: 71.94%

- Management monitors profitability and sales history with each vendor[IdioRisk] and believes it has alternative sources available for each category of merchandise it purchases
- Management believes it maintains good relationships with its vendors[IdioRisk]
- The Company has six distribution[IdioRisk] facilities[IdioRisk] serving its stores

Table B.5
List of representative sentences from 10-K filings with the highest *risk_tax*. The representative sentences were selected by locating three consecutive sentences with the highest average percentage of words marked as tax risk. Words classified to systematic risk, financial risk, idiosyncratic risk, tax risk, and litigation risk were marked in superscripts as SysRisk, FinRisk, IdioRisk, TaxRisk, and LRRisk. Multiple filings from the same company were only listed once.

Company name: INTERNATIONAL PAPER CO/NEW/ — CIK: 51434
Type: 10-K — Period of report: 2007-12-31
Date filed: 2008-02-29 — Document score [TaxRisk]: 35.61%

- Net deferred tax[TaxRisk] liability[TaxRisk] $ (2321) $ (1378) Deferred tax[TaxRisk] assets[TaxRisk] and liabilities are recorded in the accompanying consolidated balance sheet under the captions Deferred income tax[TaxRisk] assets, Deferred charges and other assets, Other accrued liabilities and Deferred income taxes[TaxRisk]
- The increase in 2007 in net deferred tax[TaxRisk] liabilities[TaxRisk] principally relates to the Company's use of net operating loss carryforwards[TaxRisk]
- The valuation allowance for deferred tax[TaxRisk] assets[TaxRisk] as of January 1, 2007, was $111 million

Company name: Alberto-Culver CO — CIK: 1368457
Type: 10-K — Period of report: 2007-09-30
Date filed: 2007-11-28 — Document score [TaxRisk]: 26.97%

- Significant components of the company's deferred tax[TaxRisk] assets[TaxRisk] and liabilities related to continuing operations at September 30, 2007 and 2006 are as follows: Total gross deferred tax[TaxRisk] assets[TaxRisk] 53,871 51,399
- Total net deferred tax[TaxRisk] assets[TaxRisk] (liabilities) $ 16,093 (4876)
  Other current assets at September 30, 2007 and 2006 include $18.4 million and $12.2 million, respectively, of net deferred tax[TaxRisk] assets[TaxRisk]
- At September 30, 2007, other assets include $17.6 million of net deferred tax[TaxRisk] assets[TaxRisk] and income taxes[TaxRisk] payable includes deferred tax[TaxRisk] liabilities[TaxRisk] of $160,000
  Management believes that it is more likely than not that results of future[SysRisk] operations will generate sufficient taxable[TaxRisk] income to realize the net deferred tax[TaxRisk] assets[TaxRisk]

Company name: CAMPBELL SOUP CO — CIK: 16732
Type: 10-K — Period of report: 2008-08-03
Date filed: 2008-10-01 — Document score [TaxRisk]: 25.12%

- Earnings from continuing operations before income taxes[TaxRisk]
- The following is a reconciliation of the effective income tax[TaxRisk] rate on continuing operations with the U.S. federal[LRRisk] statutory income tax[TaxRisk] rate
- State income taxes[TaxRisk] (net of federal[LRRisk] tax[TaxRisk] benefit)
  (1) See Note 7 for information on the divestiture of certain Australian salty snack foods brands
  In the second quarter of 2008, the company recorded a tax[TaxRisk] benefit of $13 resulting from the resolution of a state tax[TaxRisk] contingency

Table B.6
List of representative sentences from 10-K filings with the highest *risk_lr*. The representative sentences were selected by locating three consecutive sentences with the highest average percentage of words marked as litigation risk. Words classified to systematic risk, financial risk, idiosyncratic risk, tax risk, and litigation risk were marked in superscripts as SysRisk, FinRisk, IdioRisk, TaxRisk, and LRRisk. Multiple filings from the same company were only listed once.

Company name: ALTRIA GROUP, INC — CIK: 764180
Type: 10-K — Period of report: 2010-12-31
Date filed: 2011-02-25 — Document score [LRRisk]: 39.64%

- Subsidiaries (and former subsidiaries) of Altria Group, Inc. are involved in several matters subjecting them to potential costs of remediation[LRRisk] and natural resource damages under Superfund[LRRisk] or other laws and regulations[LRRisk]
- Altria Group, Inc.'s subsidiaries expect to continue to make capital and other expenditures in connection with environmental[LRRisk] laws and regulations[LRRisk]
- Summary of Significant Accounting Policies, Altria Group, Inc. provides for expenses associated with environmental[LRRisk] remediation[LRRisk] obligations[FinRisk] on an undiscounted basis when such amounts are probable and can be reasonably estimated

Company name: TENET HEALTHCARE CORP — CIK: 70318
Type: 10-K — Period of report: 2002-05-31
Date filed: 2002-08-14 — Document score [LRRisk]: 36.89%

- Based on the existing and proposed HIPAA regulations[LRRisk], the Company believes that the cost of its compliance[LRRisk] with HIPAA will not have a material[SysRisk] adverse effect on its business, financial position or results of operations
- The Company's operations, as well as the Company's purchases and sales of facilities[IdioRisk], also are subject to compliance[LRRisk] with various other environmental[LRRisk] laws, rules and regulations[LRRisk]
- The Company believes that the cost of such compliance[LRRisk] will not have a material[SysRisk] adverse effect on its business, financial position or results of operations

Company name: MEADWESTVACO CORP	CIK: 1159297
Type: 10-K	Period of report: 2001-12-31
Date filed: 2002-03-18	Document score [LRRisk]: 33.92%

- Environmental[LRRisk] organizations are challenging the Cluster Rules in the U.S. Court of Appeals
- MeadWestvaco and other companies are participating in the litigation[LRRisk] which could result in additional compliance[LRRisk] costs in excess of $150 million over several years if the legal challenge by these environmental[LRRisk] organizations is successful
- In 1998 and 1999, the EPA issued Notices of Violation to eight paper industry facilities[IdioRisk], including Westvaco's Luke, MD, mill, alleging violation of EPA's Prevention of Significant Deterioration (PSD) regulations[LRRisk] under the Clean Air Act requiring permitting and emissions evaluation prior to industrial expansion[IdioRisk]

Table C.1
Multivariate analysis. This table displays the results of regressing the log of the CDS spread on news and risk disclosure variables (i.e., text variables), controlling for accounting-based, market-based, macroeconomic, and other variables. All models include control variables; firm and time-fixed effects. Model (1) includes news variables but excludes risk disclosure variables. Model (2) includes news variables and risk-subcategory variables. Student *t*-statistics are in parentheses and adjusted for clustering in firm and time. Asterisks denote statistical significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Independent variables	Dependent variable: *lcs*		
	(1)	(2)	(3)
News			
lnarticle	0.0533***	0.0521***	0.0543***
	(3.18)	(3.16)	(3.35)
neg	0.0124**	0.0119**	0.0132***
	(2.48)	(2.39)	(2.64)
Corporate filings			
lrisk_all		0.0586***	
		(3.65)	
risk_idio			0.0285
			(0.96)
risk_sys			0.0642**
			(2.48)
risk_fin			0.1650**
			(2.39)
risk_tax			0.0027
			(0.04)
risk_lr			−0.0625
			(−0.79)
Accounting			
lta	0.0770	0.0679	0.0694
	(1.38)	(1.21)	(1.27)
roa	−0.0149**	−0.0149**	−0.0141**
	(−2.31)	(−2.36)	(−2.20)
nig	0.0132**	0.0132**	0.0127**
	(2.23)	(2.22)	(2.18)
c1	−0.0889***	−0.0872***	−0.0879***
	(−6.10)	(−6.08)	(−6.18)
c2	−0.0142	−0.0141	−0.0144
	(−1.54)	(−1.53)	(−1.57)
c3	−0.0110**	−0.0107**	−0.0114**
	(−2.12)	(−2.07)	(−2.18)
c4	0.0000	0.0000	−0.0001
	(0.02)	(0.02)	(−0.05)
qr	−0.0001	−0.0001	−0.0001
	(−0.52)	(−0.43)	(−0.50)
cta	0.0005	0.0004	0.0007
	(0.24)	(0.16)	(0.29)
taa	−0.0003	−0.0003	−0.0003
	(−1.09)	(−1.05)	(−1.06)
sg	−0.0001	−0.0002	−0.0001
	(−0.15)	(−0.20)	(−0.06)
lev	0.0010	0.0010	0.0011
	(0.57)	(0.57)	(0.57)
reta	−0.0002	−0.0002	−0.0002
	(−0.30)	(−0.29)	(−0.28)
Market			
dtd	−0.0453***	−0.0449***	−0.0446***
	(−9.92)	(−9.81)	(−9.94)
sret	−0.0267*	−0.0259*	−0.0271**
	(−1.93)	(−1.88)	(−2.00)
svol	0.7130***	0.7050***	0.7020***
	(6.25)	(6.20)	(6.23)

(*continued on next page*)

Table C.1 (*continued*)

Independent variables	Dependent variable: *lcs*		
	(1)	(2)	(3)
Macroeconomic & others			
trate	0.0692**	0.0895***	0.0834***
	(2.28)	(2.98)	(2.81)
iret	−0.0018***	−0.0018***	−0.0018***
	(−3.10)	(−3.16)	(−3.00)
mret	0.0086***	0.0087***	0.0083***
	(11.93)	(10.83)	(9.57)
invgrd	−0.4460***	−0.4390***	−0.4440***
	(−7.52)	(−7.47)	(−7.55)
lqn	0.0471*	0.0458*	0.0470*
	(1.73)	(1.68)	(1.73)
_cons	3.6080***	3.2970***	3.4580***
	(5.58)	(5.07)	(5.49)
Firm fixed effect	Included	Included	Included
Time fixed effect	Included	Included	Included
Adjusted R-square	0.898	0.898	0.898
Observations	8562	8562	8562

Table C.2
Mean CDS spread by news coverage and firm size.

Firm size		News coverage quartile			
		1 (low)	2	3	4 (high)
1 (small)	Mean	215.76	219.47	223.39	320.23
	N	2344	835	891	211
2 (big)	Mean	129.18	121.67	143.50	181.38
	N	991	640	1136	1514

References

Aboody, D., Hughes, J.S., Bugra Ozel, N., 2014. Corporate bond returns and the financial crisis. Journal of Banking & Finance 40, 42–53.
Acharya, V.V., Johnson, T.C., 2007. Insider trading in credit derivatives. Journal of Financial Economics 84, 110–141.
Barber, B.M., Odean, T., 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785–818.
Bharath, S.T., Shumway, T., 2008. Forecasting default with the merton distance to default model. Review of Financial Studies 21, 1339–1369.
Bodnaruk, A., Loughran, T., McDonald, B., 2015. Using 10-K text to gauge financial constraints. Journal of Financial and Quantitative Analysis 50, 623–646.
Bongaerts, D., De Jong, F., Driessen, J., 2011. Derivative pricing with liquidity risk: theory and evidence from the credit default swap market. The Journal of Finance 66, 203–240.
Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H.-M., Steele, L., 2014. The information content of mandatory risk factor disclosures in corporate filings. Review of Accounting Studies 19, 396–455.
Cao, C., Yu, F., Zhong, Z., 2010. The information content of option-implied volatility for credit default swap valuation. Journal of Financial Markets 13, 321–343.
Chan, W.S., 2003. Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics 70, 223–260.
Chen, C.-W., Pantzalis, C., Park, J.C., 2013. Press coverage and stock price deviation from fundamental value. Journal of Financial Research 36, 175–214.
Da, Z.H.I., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.
Das, S.R., Hanouna, P., Sarin, A., 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. Journal of Banking & Finance 33, 719–730.
Duffie, D., Saita, L., Wang, K., 2007. Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics 83, 635–665.
Dyck, A., Volchkova, N., Zingales, L., 2008. The corporate governance role of the media: evidence from Russia. The Journal of Finance 63, 1093–1135.
Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity. Information, and infrequently traded stocks. The Journal of Finance 51, 1405–1436.
Fama, E.F., French, K.R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. The Journal of Finance 64, 2023–2052.
Fang, L.H., Peress, J., Zheng, L., 2014. Does media coverage of stocks affect mutual funds' trading and performance? Review of Financial Studies 27, 3441–3466.
Goh, J.C., Ederington, L.H., 1993. Is a bond rating downgrade bad-news, good-news, or no news for stockholders. The Journal of Finance 48, 2001–2008.
Griffin, J.M., Hirschey, N.H., Kelly, P.J., 2011. How important is the financial media in global markets? Review of Financial Studies 24, 3941–3992.
Gurun, U.G., Butler, A.W., 2012. Don't believe the hype: local media slant, local advertising, and firm value. The Journal of Finance 67, 561–597.
Hasan, I., Hoi, C.K., Wu, Q., Zhang, H., 2014. Beauty is in the eye of the beholder: the effect of corporate tax avoidance on the cost of bank loans. Journal of Financial Economics 113, 109–130.
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, 405–440.
Heider, F., Ljungqvist, A., 2015. As certain as debt and taxes: estimating the tax sensitivity of leverage from state tax changes. Journal of Financial Economics 118, 684–712.
Hertzberg, A., Liberti, J.M., Paravisini, D., 2011. Public information and coordination: evidence from a credit registry expansion. The Journal of Finance 66, 379–412.
Hoberg, G., Maksimovic, V., 2015. Redefining financial constraints: a text-based analysis. Review of Financial Studies 28, 1312–1352.
Hull, J., Predescu, M., White, A., 2004. The relationship between credit default swap spreads, bond yields, and credit rating announcements. Journal of Banking & Finance 28, 2789–2811.
Kothari, S.P., Li, X., Short, J.E., 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. The Accounting Review 84, 1639–1670.
Kravet, T., Muslu, V., 2013. Textual risk disclosures and investors' risk perceptions. Review of Accounting Studies 18, 1088–1122.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics 45, 221–247.
Lin, C., Officer, M.S., Wang, R., Zou, H., 2013. Directors' and officers' liability insurance and loan spreads. Journal of Financial Economics 110, 37–60.
Loughran, T.I.M., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance 66, 35–65.
Lu, C.-W., Chen, T.-K., Liao, H.-H., 2010. Information uncertainty, information asymmetry and corporate bond yield spreads. Journal of Banking & Finance 34, 2265–2279.
Merton, R.C., 1974. On the pricing of corporate debt: the risk structure of interest rates. The Journal of Finance 29, 449–470.
Norden, L., 2008. Credit Derivatives, Corporate News, and Credit Ratings. SSRN eLibrary.
Norden, L., Weber, M., 2004. Informational efficiency of credit default swap and stock markets: the impact of credit rating announcements. Journal of Banking & Finance 28, 2813–2843.
Odders-White, E.R., Ready, M.J., 2006. Credit ratings and stock liquidity. Review of Financial Studies 19, 119–157.
Ozick, G., Sadka, R., 2013. Media coverage and hedge fund returns. Financial Analysts Journal 69, 57–75.
Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Review of Financial Studies 22, 435–480.
Qiu, J., Yu, F., 2012. Endogenous liquidity in credit derivatives. Journal of Financial Economics 103, 611–631.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. The Journal of Finance 62, 1139–1168.
Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. The Journal of Finance 63, 1437–1467.
Thompson, S.B., 2011. Simple formulas for standard errors that cluster by both firm and time. Journal of Financial Economics 99, 1–10.
Vassalou, M., Yuhang, X., 2004. Default risk in equity returns. The Journal of Finance 59, 831–868.