JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 52, No. 2, Apr. 2017, pp. 811–836

doi:10.1017/S0022109017000187

Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing

Mine Ertugrul, Jin Lei, Jiaping Qiu, and Chi Wan*

Abstract

This paper investigates the impact of a firm's annual report readability and ambiguous tone on its borrowing costs. We find that firms with larger 10-K file sizes and a higher proportion of uncertain and weak modal words in 10-Ks have stricter loan contract terms and greater future stock price crash risk. Our results suggest that the readability and tone ambiguity of a firm's financial disclosures are related to managerial information hoarding. Shareholders of firms with less readable and more ambiguous annual reports not only suffer from less transparent information disclosure but also bear the increased cost of external financing.

I. Introduction

Annual reports filed by publicly traded firms pursuant to the Securities Exchange Act of 1934, Form 10-Ks, are the primary source of information for capital market participants such as shareholders, creditors, and financial analysts. The readability and other textual properties of financial disclosures thus have a great impact on the effective communication of valuation-relevant information between the firm and the market (Loughran and McDonald (LM) (2014)).

The literature has linked 10-K readability to earnings persistence (Li (2008)), analyst coverage and dispersion (Lehavy, Li, and Merkley (2011)), the investment decisions and trading behavior of individual investors who have limited information-processing expertise (Lawrence (2013)), and firm investment efficiency (Biddle, Hilary, and Verdi (2009)). More recently, LM (2014) propose the use of the file size of 10-K filings as a capable and robust measure of readability and demonstrate that a larger 10-K file size is significantly related to a poor corporate information environment, as evidenced by higher postfiling return volatility,

*Ertugrul, mine.ertugrul@umb.edu, Wan (corresponding author), chi.wan@umb.edu, College of Management, University of Massachusetts Boston; Lei, jlei@brocku.ca, Goodman School of Business, Brock University; and Qiu, qiu@mcmaster.ca, DeGroote School of Business, McMaster University. Qiu acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. Wan acknowledges financial support from the Program for Innovative Research Team of Shanghai University of Finance and Economics. We thank Tim Loughran (the referee) and Paul Malatesta (the editor) for constructive comments.

greater unexpected earnings, and more dispersed analyst forecasts. These studies provide significant insights into the impact of less readable annual reports on equity market participants.

Yet an important question remains unanswered: Does the readability of an annual report matter to creditors and affect a firm's borrowing cost? The answer to this question is crucial to understanding the cost of less readable annual reports and the wealth implications for shareholders. If the lack of readability is a source of informational risk that raises the cost of external capital, the shareholders of the firm will not only have difficulty extracting valuation-related information but also ultimately bear the increased cost of external financing.

In addition to readability, ambiguous text in annual reports can be a source of informational risk and interfere with investors' ability to comprehend reports. In particular, Loughran and McDonald (2011) document that the use of uncertain terms (e.g., *approximate, contingency, uncertain*, and *indefinite*) and weak modal words (e.g., *might*, *possible*, *approximate*, and *contingent*) in 10-K filings is significantly positively related to the stock return volatility in the year after the filing of the annual report. Loughran and McDonald (2013) establish a positive link between the ambiguity of language in S-1 initial public offering (IPO) filings and some key IPO performance metrics: First-day returns, absolute price revisions, and subsequent volatility. Their findings indicate that the ambiguous text of corporate disclosures increases valuation uncertainty. We thus posit that a high frequency of ambiguous words used in annual reports could make it more difficult for external capital providers such as banks to assess a firm's risk characteristics and its value properly. In other words, the ambiguous tone of annual reports could increase a firm's perceived information risk, which is then priced by creditors.

In this paper, we utilize the information on bank loan contracting and conduct the first systematic investigation into the impact of readability and the ambiguous tone of annual reports on the cost of corporate borrowing. We focus on bank loans for two reasons. First, bank loans are the primary source of external credit (Chava, Livdan, and Purnanandam (2009), Graham, Li, and Qiu (2008), and Li, Qiu, and Wan (2011)). The flow of funds data from the Federal Reserve System indicates that over the past decade, there was $780 billion in net debt security issuances and only $2 billion for equities. Among debt issues, bank loans play a significant role (approximately 54% of total debt since 1980).

Second, through active monitoring, banks have better access to firm-specific information than general shareholders and other market participants (e.g., financial analysts, mutual funds, and insurance companies). Moreover, banks may directly demand detailed firm-specific information during loan negotiations. The financial intermediation literature has accumulated strong evidence in support of banks' superior ability to process financial information compared with investors in the equity market (e.g., Bharath, Sunder, and Sunder (2008)). As such, despite evidence showing that the readability of annual report affects experienced equity market participants such as financial analysts, it is not clear whether readability matters to sophisticated creditors (i.e., banks) that are quasi-insiders of corporations and possess more information than external market participants. Investigating whether and to what extent hard-to-read reports and the ambiguous tone of disclosures affect banks' evaluation allows us to shed light on the nature

of readability and document tone to provide a complete picture of their effect on corporate stakeholders.

We start by examining how 10-K readability and ambiguous tone affect the loan spread, defined as the loan rate (including any annual fee paid to the bank group) minus the London Interbank Offered Rate (LIBOR). The spread represents the direct cost of a bank loan. Controlling for various firm and loan characteristics and industry and year fixed effects, we show that the effect of annual report readability (measured by 10-K file size) on loan spreads is statistically and economically significant. Our estimates indicate that a 1-standard-deviation increase in the logarithm of 10-K file size is associated with a 9.73% increase in the loan spread. Moreover, using Loughran and McDonald's (2011) sentiment word lists, we find that the high percentages of uncertain and weak modal words are linked to more stringent contractual terms in bank loans. Our estimates indicate that a 1-standard-deviation increase in weak modal (uncertain) words leads to a 1.95% (2.0%) increase in the average loan spread. These results suggest that low readability (whether it represents a lack of transparency or is a manifestation of structural complexity) and ambiguous tone heighten a firm's perceived information risk and directly weaken a firm's perceived creditworthiness (Barry and Brown (1984), Easley, Hvidkjaer, and O'Hara (2002), and Easley and O'Hara (2004)).

We then investigate the impact of these textual features on nonprice contract terms. The nonprice terms are important instruments for lenders to limit their exposure to borrowers' risks and mitigate agency costs. For instance, short-term loans can be useful in responding to information problems that lenders face by forcing frequent information disclosure and renegotiation of contract terms (Barclay and Smith (1995), Ortiz-Molina and Penas (2008)). In addition, banks might be more likely to ask for collateral from firms with a higher level of information risk. If low readability and the use of ambiguous language in the annual reports are related to information hoarding, lenders would be prompted to alter nonprice contract terms accordingly.

We find that a larger and thus less readable 10-K file size reduces loan maturity and increases the likelihood of collateral requirement. The finding that low readability is associated with more stringent nonprice loan terms indicates that banks increase their monitoring intensity for firms with hard-to-read annual reports. This is consistent with the argument made by LM (2014) that low readability is related to firms' intention to obfuscate mandated earnings-relevant information by burying it in longer documents. We also find that the frequencies of uncertain and weak modal words are significantly and positively related to the likelihood of security requirement. This result suggests that, beyond readability and other firm characteristics, the tone of 10-Ks contains useful information in assessing the riskiness of a firm and bears upon both price and nonprice loan terms.

Although our analysis strongly indicates that readability is a key determinant of the cost of borrowing and is linked to the attempt to conceal bad news, the impact of readability could partially be driven by firm complexity; that is, some firms might have more complex businesses and need longer, and thus potentially less readable, reports to disclose. Following LM (2014), we attempt to alleviate

this concern in our analysis by using a business segment index. Nevertheless, as LM note, one cannot completely separate readability from firm complexity.

To further gauge whether readability and ambiguous tone are related to information hoarding, we examine the relation between readability, the frequency of uncertain and weak modal words, and stock price crash risk. The prolonged hoarding of bad news through hard-to-read reports and stockpiling of ambiguous text can cause the stock price to become severely overvalued. However, there is a limit to the amount of bad information that a company can hide from the market. When the accumulated bad news reaches a tipping point, it will be suddenly released to the market all at once, causing the stock price to crash (Hutton, Marcus, and Tehranian (2009), Jin and Myers (2006)).[1] Hence, the impact of readability and the ambiguous tone of 10-K filings on crash risk offers unique insights into the effect of readability and tone of business disclosures on corporate information hoarding.

We show that low readability is positively related to a firm's future crash risk. This result is robust to alternative measures of proneness to stock price crash and controlling for the opacity of financial reports. The finding suggests that lack of readability is related to managers' hoarding of bad news, which leads to stock overvaluation and eventually to a plunge in the stock price. Turning to the tone of 10-K reports, we find that high frequencies of uncertain and weak modal words tend to increase the likelihood of a future crash. The increased crash proneness is also consistent with our earlier findings that 10-Ks with low readability and ambiguous tone are associated with higher costs of capital, suggesting that information hoarding decreases the firm's perceived creditworthiness and heightens its cost of capital. Overall, our results provide strong evidence that the readability and ambiguous tone of annual reports are related to a firm's information-hoarding activities that increase its informational risk and result in higher costs of borrowing.

Our paper makes four contributions to the growing literature on the causes and consequences of textual properties of business disclosures. First, we conduct a first systematic investigation of the relation between financial readability and borrowing cost. Our findings indicate that low readability exacerbates a firm's information risk and results in stringent loan contract terms. This finding implies that when shareholders face less readable reports, they suffer not only from less effective communication of valuation-relevant information but also from rising costs of external financing. Second, we provide new evidence that a firm's 10-K readability is related to managerial information-hoarding activities. The positive association between readability and stock price crash risk suggests that low readability goes beyond the concern of textual complexity and is related to information hoarding.

[1] Such a concern was raised by U. S. Securities and Exchange Commission (SEC) chairman Mary Schapiro before the Financial Crisis Inquiry Commission on Jan. 14, 2010: "A central question … is whether investors received timely and accurate disclosure concerning deteriorating business conditions, increased risks, and downward pressure on asset values." Recent research shows that the lack of information transparency enables bad information hoarding, thereby increasing future crash risk (Hutton et al. (2009), Jin and Myers (2006)).

Third, our findings indicate that, in addition to readability, the ambiguous tone of 10-K reports has a significant bearing on banks' assessment of a firm. Fourth, our findings shed light, from the creditors' perspective, on the relevance of leading readability measures and the usefulness of the Loughran and McDonald's (2011) sentiment word lists. Our results support the notion that when firms try to obscure mandated earnings-relevant information, they are more likely to bury the results in longer and more ambiguous documents than to use complex words.

The rest of the paper is organized as follows: Section II outlines the data and key variables and discusses summary statistics. Section III examines the impact of annual report readability and uncertain document tone on bank loan contracting. Section IV investigates the relationship of readability and ambiguous tone with stock price crash risk. Section V concludes.

II. Main Variables, Data, and Summary Statistics

In this section, we describe the data used in our analysis, the construction of key variables, and summary statistics. Appendix A defines variables.

A. Main Variables of Interest

1. Financial Report Readability

Our primary measure of financial report readability is 10-K file size, as proposed by LM (2014). LM show that 10-K file size (in megabytes) is a relevant and robust measure of readability in financial disclosures. LM point out several issues related to the use of the Fog index in measuring the readability of financial documents.[2] In particular, the first part of the Fog index (average number of words per sentence) has reasonable correlations with other measures of readability. However, its second component (percentage of complex words) is a potentially misleading factor in the assessment of readability. The Fog index indicates that an increase in the number of multisyllable words decreases readability. However, many such words are frequently used in financial documents, such as *corporation, company, management*, and *operations*. These common financial terms are not difficult for an average investor to comprehend. Moreover, LM show that 52 "complex" words, out of 45,000 complex words in their 10-K sample, account for more than 25% of the complex word count, and almost all 52 of these complex words are common business terms. In light of the finding that the Fog index may fail to recognize the distinctiveness of financial terminology and spuriously deflate readability, LM propose the use of 10-K file size as a robust and capable proxy of readability in the context of financial reports. Furthermore, the file size

[2]The Fog index relies on the average number of words per sentence and the percentage of complex words in a document to quantify the degree to which text can be read and understood. Intuitively, the index estimates the years of formal education a reader needs to understand the text in first reading. Specifically,

$$\text{FOG} = (\text{AVG\_NO\_OF\_WORDS} + \text{COMPLEX\_WORDS}) \times 0.4,$$

where complex words are those with three or more syllables. A typical text with a Fog index of 18 or higher means it is unreadable, between 14 and 18 difficult to read, between 12 and 14 ideal, between 10 and 12 acceptable, and between 8 and 10 childish.

does not require arduous text parsing and thus is subject to fewer measurement errors.

2. Ambiguous Tone

Loughran and McDonald (2011) show that word classifications derived for nonbusiness disciplines frequently misclassify common words in the financial context. Hence, they compile categories of word lists including uncertain and weak modal words to accurately reflect the ambiguous tone of financial disclosures. Weak modal words such as *might*, *possible*, and *somewhat* indicate a lack of confidence. Words that indicate uncertainty, such as *approximate*, *assume*, *contingent*, *depend*, and *indefinite*, emphasize imprecision. We use the percentage of uncertain and weak modal words to examine the impact of the 10-K's tone on bank loan contracting and crash risk.

3. Cost of Bank Loans

The primary outcome variable in our bank loan analysis is the cost of bank debt. Much of the literature on the cost of bank debt (e.g., Graham et al. (2008), Bharath, Dahiya, Saunders, and Srinivasan (2011)) uses the loan spread over LIBOR at the time of the loan origination as a measure of the cost of bank debt. The AIS_DRAWN variable in the Loan Pricing Corporation's (LPC) DealScan database describes the amount a borrower pays in basis points over LIBOR for each dollar drawn down. It also adds the spread of the loan with any annual (or facility) fee paid to the bank group. We use the natural logarithm of the AIS_DRAWN variable as our measure of the cost of bank debt. We define ln(SPREAD) as the natural logarithm of the all-in-spread drawn over LIBOR in basis points. The unit of observation is a loan, also referred to as a *facility* or *tranche* in DealScan.

4. Firm-Level Crash Risk

Following Chen, Hong, and Stein (2001), Hutton et al. (2009), and Kim, Li, and Zhang (2011a), (2011b) and using firm-specific weekly returns, firm-specific crash risk is measured by three proxies: i) NCSKEW, the negative conditional skewness of future returns; ii) DUVOL, the down-to-up volatility; and iii) CRASH, the indicator of the occurrence of future extreme downside return movements. As an intuitive and rather crude indicator of crashes, CRASH is robust to potential measurement error. These three proxies provide a robust and broad assessment of crash likelihood. Appendix B details the calculation of the three crash risk proxies.

B. Data and Summary Statistics

We collect stock price data from Center for Research in Security Prices (CRSP) and financial data from Compustat for U.S. publicly traded firms from 1995 to 2013. Financial and utility firms are excluded. We merge these data sets with the set of firms for which readability indexes (10-K file size and the Fog index) are available. This leaves us with 32,207 firm-year observations for our crash risk analysis. In our bank loan analysis, we have 18,029 facility-year observations.

Panel A of Table 1 reports summary statistics for readability proxies, loan contract terms, and firm characteristics. The average 10-K file size

TABLE 1
Summary Statistics

Panel A of Table 1 reports summary statistics for the data used to analyze the impact of readability and the ambiguous tone of the annual reports on bank loans. Panel B presents summary statistics for the sample used to analyze the impact of annual report readability and ambiguous tone on firm future stock price crashes. Financial services firms and utility firms are excluded. The sample period is 1995–2013. The final column shows the results of univariate comparisons of means between firms with more readable 10-Ks (i.e., annual ln(FILE_SIZE) < annual sample median) and less readable 10-Ks (i.e., annual ln(FILE_SIZE) > annual sample median). The *t*-test statistics of the null hypothesis of an equal mean are reported in parentheses. All variables are defined in Appendix A. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Variable	No. of Obs.	Mean	Std. Dev.	Q1	Median	Q3	Univariate Comparison (more readable − less readable)
Panel A. Variables Used in the Bank Loan Analysis							
Proxies of 10-K Readability and Tone Ambiguity							
ln(FILE_SIZE)	18,029	−0.225	1.118	−1.149	−0.307	0.524	−1.091*** (60.8)
WEAK_MODAL (%)	18,029	0.445	0.191	0.294	0.419	0.571	0.000 (0.32)
UNCERTAIN (%)	18,029	1.247	0.334	0.999	1.237	1.490	0.012** (2.41)
FOG	18,029	19.55	1.519	18.50	19.37	20.36	−0.563*** (−25.9)
AVG_NO_OF_WORDS	18,029	22.07	4.442	18.83	21.26	24.57	−2.162*** (−34.4)
COMPLEX_WORDS (%)	18,029	25.45	6.204	25.26	26.71	28.09	0.786*** (5.75)
Loan Characteristics							
AIS_DRAWN (bps)	18,029	205.0	142.8	100	175	275	−13.61*** (−6.45)
ln(LOAN_SIZE)	18,029	4.804	1.614	3.881	5.010	5.941	−0.683*** (−29.2)
MATURITY (months)	18,029	45.92	22.48	30	52	60	−0.209 (−0.631)
SECURITY	18,029	0.427	0.494	0	0	1	0.060** (8.02)
Other Controls							
ln(ASSETS)	18,029	6.853	1.807	5.578	6.809	8.028	−1.075*** (−41.8)
MB	18,029	1.740	1.342	1.120	1.427	1.946	0.011 (0.55)
LEVERAGE	18,029	0.257	0.216	0.095	0.225	0.366	−0.058*** (−18.0)
PROFITABILITY	18,029	0.016	0.159	0.004	0.039	0.072	0.010*** (4.42)
TANGIBILITY	18,029	0.313	0.232	0.127	0.249	0.451	−0.021*** (−5.98)
EDF	18,029	1.841	3.312	0.000	0.015	1.988	−0.169*** (3.22)
Panel B. Variables Used in the Crash Risk Analysis							
Proxies of 10-K Readability and Tone Ambiguity							
ln(FILE_SIZE)	32,207	−0.361	1.044	−1.264	−0.329	0.399	−1.050*** (−139.0)
WEAK_MODAL (%)	32,207	0.524	0.226	0.345	0.499	0.674	−0.018*** (−7.23)
UNCERTAIN (%)	32,207	1.358	0.342	1.110	1.364	1.609	−0.017*** (−4.32)
FOG	32,207	19.42	1.946	18.53	19.38	20.33	−0.276*** (−12.8)
AVG_NO_OF_WORDS	32,207	21.80	4.560	18.93	21.11	24.07	−1.162*** (−23.1)
COMPLEX_WORDS (%)	32,207	26.99	2.012	25.67	27.02	28.22	0.505*** (22.7)
Crash Risk Measures							
NCSKEW	32,207	0.159	0.867	−0.317	0.094	0.537	−0.079*** (−8.21)
DUVOL	32,207	0.008	0.375	−0.237	−0.005	0.236	−0.036*** (−8.65)
CRASH	32,207	0.209	0.406	0	0	0	−0.012*** (−2.75)
Other Controls							
DTURN	32,207	−0.004	0.142	−0.033	0.001	0.035	−0.000 (−0.066)
SIGMA	32,207	0.078	0.043	0.047	0.067	0.097	0.001*** (2.74)
RET	32,207	0.004	0.011	−0.002	0.003	0.009	0.001*** (7.503)
ROA	32,207	−0.032	0.238	−0.033	0.034	0.076	0.012*** (4.34)
SIZE	32,207	5.838	2.097	4.343	5.816	7.191	−1.057*** (−46.8)
MB	32,207	1.505	1.585	0.547	1.009	1.816	0.087*** (4.94)
LEVERAGE	32,207	0.209	0.211	0.012	0.166	0.329	−0.063*** (−27.1)

is approximately 0.8 MB. The average percentage of weak modal words (WEAK_MODAL) is 0.445%, and that of uncertain words (UNCERTAIN) is 1.247%. The average Fog index is 19.55, which deems an average 10-K file unreadable.[3] Turning to loan characteristics, we observe that an average loan

[3] An average Fog index of 19.55 implies that the reader of an average 10-K needs about 20 years of formal education (roughly a PhD degree) to understand the text in a first reading. This requirement seems excessive. The use of long lists and/or bullet points (which appear as long sentences) could partly account for this enormous average Fog index value.

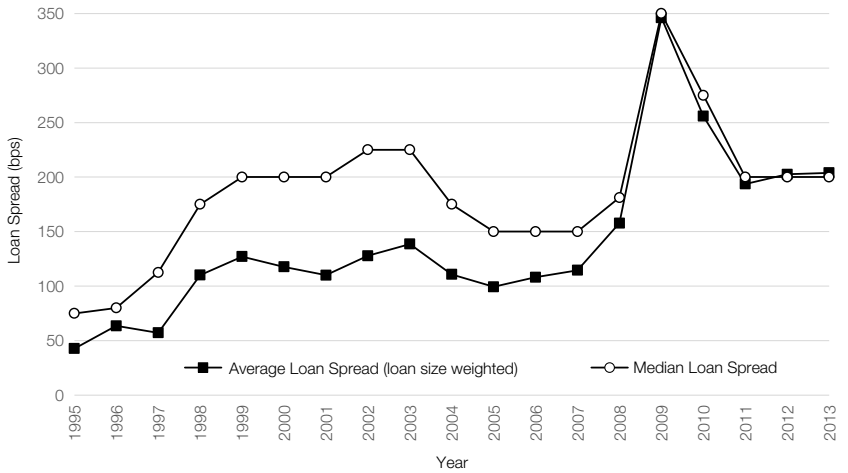
has a spread of 122 basis points (bps) and a maturity of 46 months. The average loan size in our sample is $297.5 million, and 43% of loans are secured. To shed light on the impact of readability, we split the sample based on the median of ln(FILE_SIZE) in a year and perform a univariate comparison of the means of the two subsamples (more vs. less readable filings). The results, presented in the last column of Panel A, show that firms with more readable annual reports have lower loan spreads and smaller loan size with a higher fraction of security requirement. However, given that loan contract terms are affected by many of a firm's characteristics, it is important to use multivariate analysis to control for related firm characteristics.

Panel B of Table 1 presents the summary statistics of the variables used in crash risk analysis. As evidenced by the means of NCSKEW and DUVOL, we find that the return distribution is often negatively skewed and featured with large downside deviations. The average of CRASH is 0.209, indicating that 20.9% of our firm-year observations experience at least one crash event. As in Panel A, we also conduct a univariate comparison between firms with more versus less readable reports. The results demonstrate that firms with more readable annual reports have lower negative skewness and down-to-up volatility, and they experience fewer crashes. This indicates the potential importance of readability in determining a firm's information-hoarding activities. Moreover, we observe that firms with more readable reports are smaller, with higher profitability and lower leverage.

Figure 1 shows the time trend of the average loan spread for our sample. The average borrowing cost stays in the 100–150 bps range from the late 1990s to 2007. The spreads start to increase sharply in 2008 and peak in 2009 at the height of the credit crunch. Although loan spreads start declining after 2009, they remain higher than the level before the financial crisis.

FIGURE 1
Time Trend of Loan Spreads

Figure 1 presents the loan-size-weighted average and median loan spreads from 1995 to 2013. The loan spread is measured as the all-in-spread drawn over the London Interbank Offered Rate (LIBOR) in basis points (bps). The sample includes publicly traded nonfinancial nonutility U.S. firms from the Loan Pricing Corporation's (LPC) DealScan database.



III. Readability, Ambiguous Tone, and Bank Loan Contracting

In this section, we examine whether and how the readability and ambiguous tone of financial disclosures are related to bank loan contracting, which reflects creditors' assessment of a firm's credit risk.

A. Readability and Loan Spreads

To analyze the effect of annual report readability on loan spreads, we use the following specification:

$$\ln(\text{SPREAD}) = f(\text{READABILITY}, \text{FIRM\_CHARACTERISTICS}, \text{LOAN\_CHARACTERISTICS}, \text{INDUSTRY\_FE}, \text{YEAR\_FE}, \text{LOAN\_TYPE\_FE}, \text{LOAN\_PURPOSE\_FE}), \quad (1)$$

where ln(SPREAD) is measured as the natural logarithm of the DealScan all-in-spread drawn, and READABILITY is a proxy for (lack of) 10-K readability. As controls in the pricing equation, we follow the conventions of bank-lending literature (e.g., Bharath et al. (2011), Graham et al. (2008), and Li et al. (2011)) to include relevant firm and loan characteristics.

Firm characteristics include several borrower-specific control variables. ln(ASSETS) is included because large firms might have greater stability and lower information asymmetry and therefore lower spreads. Market-to-book (MB) is a proxy for a firm's growth opportunities and measures the market value of a company's equity as a multiple of the book value of the company's equity. After controlling for characteristics such as leverage and tangibility of book assets, growth opportunity is expected to have a negative impact on loan cost because it represents the additional value over book assets that debt holders can access in the event of default. LEVERAGE reflects a firm's existing debt level. Firms with high leverage ratios, on average, have a high default risk and thus are expected to have high borrowing costs. We include PROFITABILITY to capture the fact that the default risk is relatively low for a profitable firm. TANGIBILITY is the ratio of total tangible assets (property, plant, and equipment (PPE)) to the book value of total assets (Campello and Giambona (2013)). A firm with high tangibility generally has high expected recovery rates in the event of default and thus should be able to borrow at a relatively low interest rate by pledging tangible assets to creditors. EDF stands for a firm's expected default frequency, which is a forward-looking measure of default probability. A small EDF indicates the firm's strong financial health and low default risk. The monthly EDF is calculated following Bharath and Shumway (2008). The annual average is included in our analysis to control for a company's default risk. All readability proxies and firm-level controls are lagged by 1 year.

Furthermore, we include nonprice loan terms that might affect the loan spread. Specifically, ln(MATURITY) is the natural logarithm of loan maturity in months. Long maturity might reflect a borrower's good credit quality and less information asymmetry (Wittenberg-Moerman (2008)). Therefore, a long-term debt may be charged a low interest rate. SECURITY is a binary variable indicating

whether the loan is secured or not.[4] Berger and Udell (1990) and Bharath et al. (2011) find that secured loans are more likely to be charged higher spreads. In addition, we control for ln(LOAN_SIZE), the natural logarithm of the amount of a single loan.[5] Loan size may reflect economies of scale in bank lending and is expected to be inversely related to the loan rate (Berger and Udell (1990)).

We also include loan type and loan purpose dummy variables in our analyses. Bank loans can be categorized into different types, including 364-day loans, term loans, and revolving loans, and can be used for different purposes, such as debt repayment, takeovers, and working-capital financing. Because loans with different types and purposes might have different risks, their pricing can also be different. In addition, in all specifications, we include industry, year, loan purpose, and loan-type dummies.

Table 2 shows the results of our baseline regressions specified in equation (1). The t-statistics, corrected for heteroskedasticity and firm-level clustering, are presented in parentheses. Column 1 reports the regression results, with ln(FILE_SIZE) as the leading explanatory variable. The coefficient of file size is found to be positive and significant at the 1% significance level, implying that the less readable the annual report of a firm, the higher its loan cost. The coefficient remains positive and statistically significant after the inclusion of loan characteristics in our analyses, as shown in column 2. Economically, the estimates in column 2 suggest that, ceteris paribus, a 1-standard-deviation increase in ln(FILE_SIZE) is associated with a 9.73% increase in the average loan spread.[6]

Some firms might have a more complex business and thus need longer reports to disclose information, which could reduce their 10-K readability. To control for the structural complexity of a firm's operations, we follow LM (2014) and reestimate our model with the business segment index as an additional control. The segment index, denoted as BSEG, is defined as the sum of the squared business segment proportions reported in the Compustat Segment data. A smaller value of BSEG suggests a higher degree of firm complexity because the firm operates across numerous business segments. As shown in column 3 of Table 2, our baseline result fully holds given the presence of BSEG. Further, we include ln(FIRM_AGE) as an additional variable to capture a firm's access to nonbank resources of lending. The results reported in column 4 show that the explanatory power of 10-K file size is largely unaffected by the inclusion of this control variable.

Columns 5–7 of Table 2 report the results for the Fog index and its two components. The index has a positive association with loan spreads. However, turning to its two components, columns 6 and 7 show that the statistical significance of the Fog index in column 5 is driven primarily by the explanatory power of the *average number of words per sentence*, whereas the *percentage of complex words* does not impinge on banks' assessment. The results are consistent with LM's (2014)

[4]Following Bharath et al. (2011), SECURITY is equal to 0 if a loan's security information is not recorded in DealScan. Similar results are obtained if we exclude SECURITY from the regression.

[5]For a line of credit, the loan size is measured by the total amount of the line of credit.

[6]Given that the coefficient of ln(FILE_SIZE) is 0.087 and ln(FILE_SIZE) has a standard deviation of 1.118 (shown in Table 1), a 1-standard-deviation increase of ln(FILE_SIZE) would increase the loan spread by 9.73% ($0.0973 = 0.087 \times 1.118$).

TABLE 2

Annual Report Readability and Loan Spreads

Table 2 reports the results of OLS regressions of loan spread on 10-K readability and other control variables. The dependent variable is the natural logarithm of loan spreads (AIS_DRAWN). The regressors also include industry and year fixed effects and loan-type and loan-purpose fixed effects. All variables are defined in Appendix A. The *t*-statistics, corrected for heteroskedasticity and firm-level clustering, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: ln(SPREAD)						
Variable	1	2	3	4	5	6	7
$\ln(\text{FILE\_SIZE}_{t-1})$	0.094*** (8.67)	0.087*** (8.63)	0.088*** (8.61)	0.083*** (8.26)			
FOG_{t-1}					0.025*** (5.51)		
$\text{AVG\_NO\_OF\_WORDS}_{t-1}$						0.007*** (4.62)	
$\text{COMPLEX\_WORDS}_{t-1}$ (%)							−0.0004 (−0.10)
$\ln(\text{ASSETS}_{t-1})$	−0.211*** (−23.30)	−0.123*** (−10.06)	−0.129*** (−13.17)	−0.103*** (−8.56)	−0.109*** (−9.13)	−0.108*** (−9.09)	−0.106*** (−8.73)
MB_{t-1}	−0.079*** (−4.22)	−0.068*** (−4.09)	−0.067*** (−4.01)	−0.072*** (−4.17)	−0.067*** (−4.06)	−0.068*** (−4.07)	−0.068*** (−4.09)
LEVERAGE_{t-1}	0.564*** (13.11)	0.534*** (13.79)	0.552*** (14.07)	0.507*** (13.49)	0.543*** (13.79)	0.546*** (13.79)	0.554*** (14.03)
$\text{PROFITABILITY}_{t-1}$	−0.595*** (−6.58)	−0.449*** (−6.04)	−0.446*** (−5.98)	−0.429*** (−5.91)	−0.464*** (−6.15)	−0.465*** (−6.13)	−0.476*** (−6.21)
TANGIBILITY_{t-1}	−0.098** (−2.42)	−0.094** (−2.55)	−0.099*** (−2.73)	−0.089** (−2.48)	−0.101*** (−2.74)	−0.106*** (−2.89)	−0.106*** (−2.86)
EDF_{t-1}	0.337*** (8.16)	0.259*** (7.18)	0.253*** (6.87)	0.243*** (6.62)	0.266*** (7.26)	0.267*** (7.29)	0.266*** (7.25)
$\ln(\text{LOAN\_SIZE}_t)$		−0.098*** (−9.51)	−0.091*** (−11.10)	−0.101*** (−9.92)	−0.099*** (−9.75)	−0.099*** (−9.76)	−0.100*** (−9.64)
$\ln(\text{MATURITY}_t)$		−0.046*** (−3.34)	−0.049*** (−3.61)	−0.045*** (−3.32)	−0.048*** (−3.47)	−0.048*** (−3.45)	−0.049*** (−3.54)
SECURITY_t		0.318*** (21.89)	0.317*** (21.68)	0.306*** (21.34)	0.319*** (21.84)	0.319*** (21.79)	0.319*** (21.77)
BSEG			0.080** (2.40)				
$\ln(\text{FIRM\_AGE}_{t-1})$				−0.103*** (−9.11)			
Constant	6.458*** (57.24)	6.479*** (58.67)	6.317*** (54.74)	6.614*** (59.40)	5.780*** (39.48)	6.120*** (53.40)	7.080*** (72.71)
Industry and year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	18,029	18,029	17,570	18,029	18,029	18,029	18,029
Adj. R^2	0.573	0.611	0.612	0.620	0.609	0.608	0.607

findings that demonstrate the inappropriateness of using the percentage of complex words in gauging financial disclosures' readability.

As for other control variables, we find that larger firms and firms with higher market-to-book, profitability, and tangibility ratios have lower loan spreads. As expected, firms with higher leverage ratios and expected default frequency are charged higher loan spreads. Regarding loan-level controls, lower spreads are associated with larger loans and loans with longer maturities, whereas secured loans have higher spreads.

B. Uncertain and Weak Modal Words

Loughran and McDonald (2011) document that ambiguous tone, measured by the use of uncertain and weak modal words in 10-K filings, is significantly related to subsequent stock returns, abnormal trading volume, and return volatility. They also find evidence that ambiguous tone is linked to the likelihood of accounting fraud and self-reporting material weaknesses in internal controls. We posit that in addition to burying adverse news in long documents, the use of ambiguous language in mandated disclosures could also pertain to managers' intention to obscure value-relevant information. Therefore, the unclear tone of 10-Ks could increase information risk and, consequently, the loan costs. We thus examine the impact of ambiguous text in annual reports on loan spreads.

Columns 1 and 3 of Table 3 show that, ceteris paribus, the loan spreads are larger for firms that use a higher percentage of uncertain text in their 10-Ks. The statistical significance of WEAK_MODAL and UNCERTAIN is fully

TABLE 3
The Ambiguous Tone of Annual Reports and Loan Spreads

Table 3 reports the results of OLS regressions of loan spread on the percentage of weak modal and uncertain words and other control variables. The dependent variable is the natural logarithm of loan spreads (all-in-spread drawn). Industry and year fixed effects and loan-type and loan-purpose fixed effects are also included. All variables are defined in Appendix A. The *t*-statistics, corrected for heteroskedasticity and firm-level clustering, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: ln(SPREAD)			
Variable	1	2	3	4
$\ln(\text{FILE\_SIZE}_{t-1})$		0.085***		0.086***
		(8.41)		(8.49)
WEAK_MODAL_{t-1}	0.112**	0.102**		
	(2.35)	(2.15)		
UNCERTAIN_{t-1}			0.061**	0.060**
			(2.13)	(2.19)
$\ln(\text{ASSETS}_{t-1})$	−0.104***	−0.120***	−0.105***	−0.121***
	(−8.49)	(−9.81)	(−8.53)	(−9.86)
MB_{t-1}	−0.069***	−0.070***	−0.068***	−0.069***
	(−4.09)	(−4.08)	(−4.10)	(−4.10)
LEVERAGE_{t-1}	0.559***	0.540***	0.559***	0.540***
	(14.14)	(13.89)	(14.15)	(13.92)
$\text{PROFITABILITY}_{t-1}$	−0.462***	−0.437***	−0.469***	−0.443***
	(−5.97)	(−5.82)	(−6.07)	(−5.91)
TANGIBILITY_{t-1}	−0.102***	−0.091**	−0.099***	−0.088**
	(−2.77)	(−2.48)	(−2.68)	(−2.38)
EDF_{t-1}	0.270***	0.263***	0.267***	0.260***
	(7.37)	(7.29)	(7.22)	(7.15)
$\ln(\text{LOAN\_SIZE}_t)$	−0.100***	−0.099***	−0.100***	−0.098***
	(−9.52)	(−9.41)	(−9.56)	(−9.44)
$\ln(\text{MATURITY}_t)$	−0.051***	−0.049***	−0.051***	−0.048***
	(−3.71)	(−3.51)	(−3.68)	(−3.48)
SECURITY_t	0.317***	0.316***	0.318***	0.317***
	(21.56)	(21.69)	(21.58)	(21.69)
Constant	6.109***	6.303***	6.097***	6.287***
	(54.76)	(54.95)	(53.48)	(53.88)
Industry and year fixed effects	Yes	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes	Yes
No. of obs.	17,890	17,890	17,890	17,890
Adj. R^2	0.608	0.612	0.608	0.612

retained as we further control for ln(FILE_SIZE) in columns 2 and 4, respectively. Economically, the estimates suggest that a 1-standard-deviation increase in WEAK_MODAL (UNCERTAIN) is associated with a 1.95% (2.0%) increase in the average loan spread.[7] These results highlight that ambiguous text (classified according to Loughran and McDonald (2011) word lists) has a significant bearing on a firm's creditworthiness as perceived by lenders.

C. Readability, the Ambiguous Tone, and Nonprice Contractual Terms

Besides imposing higher loan spreads, lenders may limit their exposure to borrowers' risks and mitigate agency costs through the choice of nonprice loan terms. The strict nonprice terms, such as short maturity or collateral requirements, impose significant indirect costs on the borrowing firms (Graham et al. (2008), Smith and Warner (1979)). To assess the impact of annual report readability on creditors' overall valuation, we also examine how 10-K readability affects the two leading nonprice loan terms: Loan maturity and security requirement.

Short-term loans can force borrowers' frequent information disclosure in the renegotiation of contract terms (Barclay and Smith (1995), Ortiz-Molina and Penas (2008)). Hence, banks could be more likely to extend shorter-maturity loans to a borrower with less readable annual reports, which, as indicated by our previous findings, are related to the firm's perceived creditworthiness. To study the effect of 10-K readability on loan maturity, we regress the natural logarithm of debt maturity (in the unit of months) on readability proxies and various firm and loan characteristics and report the results in Panel A of Table 4. Columns 1 and 2 show that 10-K file size is negatively related to ln(MATURITY), indicating that firms with less readable 10-Ks have shorter-maturity loans. Regarding the economic impact, the estimates in column 2 suggest that a 1-standard-deviation increase in ln(FILE_SIZE) is associated with a 1.9% (≈0.87-month) decrease in the average loan maturity.[8] The impact of the Fog index on loan maturity, however, is not significant at conventional levels.

We now turn to the other key nonprice loan term: The requirement of collateral security. Posting collateral reduces lending risk and better aligns the interests of the bank and the firm by increasing the firm's opportunity cost. Banks might be more likely to ask for collateral from firms with higher information risk (Berger and Udell (1990)). We model the likelihood of security requirement with a probit model and present the results in Panel B of Table 4. The dependent variable, SECURITY, is an indicator variable that is equal to 1 if a loan is secured, and 0 otherwise. The findings corroborate what is reported in Panel A concerning the determination of ln(MATURITY). Specifically, we find that readability, measured by 10-K file size, has a significant bearing on the likelihood of pledging collateral, whereas the Fog index does not. Economically, column 2 suggests that a

[7]Given that the coefficient of WEAK_MODAL is 0.102 and WEAK_MODAL has a standard deviation of 0.191 (shown in Table 1), a 1-standard-deviation change in WEAK_MODAL would increase the loan spread by 1.9% ($0.019 = 0.191 \times 0.102$). The marginal impact of UNCERTAIN can be calculated in the same way.

[8]Given that the coefficient of ln(FILE_SIZE) is −0.017 and ln(FILE_SIZE) has a standard deviation of 1.118 (shown in Table 1), a 1-standard-deviation change in ln(FILE_SIZE) would reduce the loan maturity by 1.9% ($= 0.017 \times 1.118$).

1-standard-deviation increase in ln(FILE_SIZE) raises the likelihood of security demand by 1.4%.[9]

Overall, the results of nonprice terms are in line with those of loan spreads: Banks are more likely to extend shorter-maturity loans and require collateral from firms with long and less readable annual reports.

We also examine whether the impact of 10-K tone ambiguity goes beyond the price term. Panel A of Table 5 shows that the effects of weak modal or uncertain word frequencies on loan maturity are generally negative, albeit insignificant. Results from Panel B indicate that the presence of these types of words has a positive and statistically significant effect on the likelihood of security requirement. The estimates in column 2 (column 4) suggest that a 1% increase in the proportion of WEAK_MODAL (UNCERTAIN) is associated with a 0.21% (0.11%) rise in the probability of pledging collateral.

TABLE 4

Annual Report Readability and Nonprice Terms: Loan Maturity and Collateral Requirement

Table 4 investigates the impact of readability on key nonprice contractual terms. Panel A reports the results of OLS regressions of the natural logarithm of loan maturity (in months) on 10-K readability and control variables. Panel B reports the results of probit regressions of the loan collateral requirement on 10-K readability and control variables. The dependent variable, SECURITY, is an indicator variable that is equal to 1 if a loan is secured, and 0 otherwise. Industry and year fixed effects and loan-type and loan-purpose fixed effects are also included in all specifications. The *t*-statistics (Panel A) and *z*-statistics (Panel B), corrected for heteroskedasticity and firm-level clustering, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Loan Maturity

	Dependent Variable: ln(MATURITY)		
Variable	1	2	3
$\ln(\text{FILE\_SIZE}_{t-1})$	−0.022***	−0.017**	
	(−2.74)	(−2.16)	
FOG_{t-1}			−0.003
			(−0.85)
$\ln(\text{ASSETS}_{t-1})$	0.042***	−0.017**	−0.020***
	(9.88)	(−2.48)	(−3.03)
MB_{t-1}	−0.003	−0.006	−0.006*
	(−0.74)	(−1.63)	(−1.71)
LEVERAGE_{t-1}	0.231***	0.170***	0.168***
	(6.18)	(4.55)	(4.51)
$\text{PROFITABILITY}_{t-1}$	0.346***	0.298***	0.301***
	(6.04)	(5.34)	(5.39)
TANGIBILITY_{t-1}	0.027	0.016	0.018
	(0.94)	(0.59)	(0.64)
EDF_{t-1}	−0.327***	−0.290***	−0.291***
	(−7.85)	(−7.12)	(−7.18)
$\ln(\text{LOAN\_SIZE}_t)$		0.095***	0.095***
		(13.31)	(13.40)
SECURITY_t		0.067***	0.067***
		(5.39)	(5.40)
Constant	2.567***	2.565***	2.662***
	(34.32)	(34.04)	(26.57)
Industry and year fixed effects	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes
No. of obs.	18,029	18,029	18,029
Adj. R^2	0.490	0.508	0.508

(continued on next page)

[9]The marginal effect of a probit model is calculated as described by Greene (1997).

TABLE 4 (continued)
Annual Report Readability and Nonprice Terms: Loan Maturity and Collateral Requirement

Panel B. Collateral Requirement

	Dependent Variable: SECURITY		
Variable	1	2	3
$\ln(\text{FILE\_SIZE}_{t-1})$	0.104***	0.097***	
	(5.74)	(5.31)	
FOG_{t-1}			0.015
			(1.26)
$\ln(\text{ASSETS}_{t-1})$	−0.216***	−0.209***	−0.183***
	(−14.15)	(−10.31)	(−9.40)
MB_{t-1}	−0.102***	−0.101***	−0.100***
	(−5.69)	(−5.58)	(−5.56)
LEVERAGE_{t-1}	0.831***	0.802***	0.804***
	(7.97)	(7.88)	(7.98)
$\text{PROFITABILITY}_{t-1}$	−0.671***	−0.737***	−0.758***
	(−3.34)	(−3.40)	(−3.44)
TANGIBILITY_{t-1}	0.056	0.057	0.013
	(0.58)	(0.60)	(0.14)
EDF_{t-1}	0.136	0.205*	0.201*
	(1.29)	(1.96)	(1.90)
$\ln(\text{LOAN\_SIZE}_t)$		−0.022	−0.014
		(−1.22)	(−0.80)
$\ln(\text{MATURITY}_t)$		0.190***	0.205***
		(5.56)	(5.99)
Constant	2.567***	2.565***	2.662***
	(34.32)	(34.04)	(26.57)
Industry and year fixed effects	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes
No. of obs.	18,029	18,029	18,029
Pseudo-R^2	0.490	0.508	0.508

In sum, our results mark the important effect of 10-K readability and frequencies of uncertain and weak modal words on bank loan contracting and show that the two textual aspects of corporate disclosures greatly influence a creditor's informative assessment of a borrower's financial health.

IV. Stock Price Crash Risk

Our previous findings highlight that the lack of readability and the high usage of uncertain language in annual reports undermine the firm's creditworthiness as assessed by banks. To provide further evidence on whether poor readability and ambiguous tone are related to information hoarding, we examine the relation between readability, percentages of uncertain and weak modal words, and a firm's future vulnerability to crashes in stock price.

A. 10-K File Size and Future Stock Price Crashes

Bloomfield (2002) suggests that managers hide adverse information through complicated disclosures in the hope that the information will not be immediately reflected in the stock prices, if at all. Furthermore, LM (2014) point out that if firms are trying to obscure mandated valuation-relevant information, they are likely to bury the results in longer and thus less readable documents. This argument is supported by LM's survey of partners of major accounting firms. LM report that when accountants were asked how they would legally attempt to

TABLE 5

The Ambiguous Tone of Annual Reports and Nonprice Terms

Panel A of Table 5 reports the results of OLS regressions of the natural logarithm of loan maturity (in months) on the percentage of weak modal and uncertain words. Panel B reports the results of probit regressions of the loan collateral requirement on weak modal and uncertain words and control variables. The dependent variable, SECURITY, is an indicator variable that is equal to 1 if a loan is secured, and 0 otherwise. The regressors in Panel A include the following control variables: $\ln(\text{ASSETS}_{t-1})$, MB_{t-1}, LEVERAGE_{t-1}, $\text{PROFITABILITY}_{t-1}$, TANGIBILITY_{t-1}, EDF_{t-1}, $\ln(\text{LOAN\_SIZE}_t)$, and SECURITY_t. The regressors in Panel B include the following control variables: $\ln(\text{ASSETS}_{t-1})$, MB_{t-1}, LEVERAGE_{t-1}, $\text{PROFITABILITY}_{t-1}$, TANGIBILITY_{t-1}, EDF_{t-1}, $\ln(\text{LOAN\_SIZE}_t)$, and $\ln(\text{MATURITY}_t)$. Industry and year fixed effects and loan-type and loan-purpose fixed effects are also included in all specifications. All variables are defined in Appendix A. The *t*-statistics (Panel A) and *z*-statistics (Panel B), corrected for heteroskedasticity and firm-level clustering, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Loan Maturity

	Dependent Variable: ln(MATURITY)			
Variable	1	2	3	4
$\ln(\text{FILE\_SIZE}_{t-1})$		−0.017**		−0.017**
		(−2.14)		(−2.16)
WEAK_MODAL_{t-1}	−0.913	0.957		
	(−0.24)	(0.26)		
UNCERTAIN_{t-1}			−0.720	−0.869
			(−0.33)	(−0.38)
Other controls	Yes	Yes	Yes	Yes
Industry and year fixed effects	Yes	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes	Yes
No. of obs.	17,890	17,890	17,890	17,890
Adj. R^2	0.507	0.508	0.507	0.508

Panel B. Collateral Requirement

	Dependent Variable: SECURITY			
Variable	1	2	3	4
$\ln(\text{FILE\_SIZE}_{t-1})$		0.100***		0.106***
		(4.87)		(5.17)
WEAK_MODAL_{t-1}	0.709***	0.560***		
	(6.69)	(4.91)		
UNCERTAIN_{t-1}			0.375***	0.274***
			(6.06)	(4.09)
Other controls	Yes	Yes	Yes	Yes
Industry and year fixed effects	Yes	Yes	Yes	Yes
Loan-type and loan-purpose fixed effects	Yes	Yes	Yes	Yes
No. of obs.	17,890	17,890	17,890	17,890
Pseudo-R^2	0.228	0.230	0.226	0.229

conceal information that is required to be disclosed, they immediately identified the strategy of burying the awkward revelation in an overwhelming amount of uninformative text and data. Therefore, we expect that, ceteris paribus, hard-to-read 10-Ks are related to future crashes, which are caused by the sudden release of the accumulated unfavorable firm-specific information.

To examine the relation between annual report readability and future stock price crash risk, we run the following regression:

$$\text{CRASH RISK} = f(\text{READABILITY}, \text{CONTROL VARIABLES}, \text{INDUSTRY\_FE}, \text{YEAR\_FE}), \quad (2)$$

where all readability proxies and control variables are lagged. When the two continuous measures NCSKEW and DUVOL are employed as proxies for crash risk, we use ordinary least squares (OLS) regressions. Because CRASH is a

dichotomous variable indicating the occurrence of crashes, the probit regression is performed.

Control variables include those suggested by previous studies on crash risk (e.g., Chen et al. (2001), Hutton et al. (2009), and Kim et al. (2011a), (2011b)). In particular, we control for average monthly share turnover ($DTURN_{t-1}$), negative skewness in the past fiscal year ($NCSKEW_{t-1}$), stock return volatility ($SIGMA_{t-1}$), average stock return (RET_{t-1}), operating performance, firm size, market-to-book ratio, and financial leverage. Industry and year fixed effects are also included.

Table 6 reports our regression results for equation (2). All statistical inferences are drawn based on standard errors that are robust to heteroscedasticity and firm-level clustering. Panel A shows the OLS results with NCSKEW as the measure of crash risk. Columns 1 and 2 indicate that 10-K file size is positively associated with the negative skewness of future returns, suggesting that firms with longer and thus less readable 10-Ks are more likely to experience price crashes. Moreover, the impact of file size is economically significant because a 1-standard-deviation increase in ln(FILE_SIZE) would lead to an 18.2% increase in the average value of NCSKEW.[10] The coefficient of the Fog index is not statistically significant. When its two components are analyzed separately, we find that, although the first component (average number of words per sentence) is positively related to crash risk, as expected (shown in column 4), the impact of the second component (percentage of complex words) is not significant.

As for control variables, consistent with prior studies, we find that average monthly share turnover ($DTURN_{t-1}$), negative skewness in the past fiscal year ($NCSKEW_{t-1}$), stock return volatility ($SIGMA_{t-1}$), and average stock return (RET_{t-1}) are positively related to crash risk as measured by the negative conditional skewness of future returns. Similar to Chen et al. (2001) and Hutton et al. (2009), we find that a firm's size and growth opportunities (i.e., the market-to-book ratio) are also positively related to its crash risk.

Turning to other crash proxies, Panel B of Table 6 presents the OLS results with our second measure, the down-to-up volatility (DUVOL). Our previous findings regarding the relationship between readability and crash risk as well as the superiority of file size are broadly retained. In particular, Panel B confirms that, ceteris paribus, firms with longer 10-K filings are more prone to crashes. Concerning the Fog index in column 3, its statistical insignificance is a result of the conflicting signs of its two components (column 4 vs. 5). In Panel C, we report the results of probit regressions, in which the dependent variable, CRASH, is a dummy variable that takes the value of 1 if a firm experiences one or more crash weeks (defined in Appendix B) in a fiscal year, and 0 otherwise. These results substantiate our finding that firms with longer and thus less readable 10-Ks are more likely to experience price crashes. However, neither the Fog index nor its components demonstrate a consistent impact on ex ante crash risk.

[10] Given that the coefficient of ln(FILE_SIZE) is 0.026 and ln(FILE_SIZE) has a standard deviation of 1.118 (shown in Table 1), a 1-standard-deviation change in ln(FILE_SIZE) would increase the average of NCSKEW by 0.029 ($=0.026 \times 1.118$), an 18.2% jump, because the average of NCSKEW is 0.159.

TABLE 6
Annual Report Readability and Future Crash Risk

Table 6 examines the relation of stock price crash risk and 10-K readability. Panels A and B show OLS results, and Panel C reports the results of probit regressions. The dependent variables are NCSKEW (Panel A), DUVOL (Panel B), and CRASH (Panel C). All regressions also include industry and year fixed effects. The *t*-statistics (Panels A and B) and *z*-statistics (Panel C), corrected for heteroscedasticity and firm-level clustering, are reported in parentheses. All variables are defined in Appendix A. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. NCSKEW, the Negative Conditional Skewness of Future Returns

	Dependent Variable: NCSKEW				
Variable	1	2	3	4	5
ln(FILE_SIZE$_{t-1}$)	0.063*** (8.13)	0.026*** (3.14)			
FOG$_{t-1}$			0.004 (1.39)		
AVG_NO_OF_WORDS$_{t-1}$				0.002** (2.17)	
COMPLEX_WORDS$_{t-1}$ (%)					−0.004 (−1.63)
DTURN$_{t-1}$		0.082** (2.29)	0.082** (2.31)	0.082** (2.30)	0.082** (2.32)
NCSKEW$_{t-1}$		0.051*** (7.02)	0.052*** (7.31)	0.052*** (7.31)	0.052*** (7.32)
SIGMA$_{t-1}$		0.430** (2.42)	0.522*** (2.98)	0.514*** (2.94)	0.522*** (2.99)
RET$_{t-1}$		4.307*** (7.77)	4.409*** (8.08)	4.416*** (8.09)	4.415*** (8.09)
ROA$_{t-1}$		0.097*** (3.37)	0.097*** (3.34)	0.098*** (3.41)	0.096*** (3.36)
SIZE$_{t-1}$		0.042*** (12.14)	0.046*** (14.32)	0.046*** (14.25)	0.046*** (14.27)
MB$_{t-1}$		0.034*** (8.02)	0.032*** (7.71)	0.032*** (7.69)	0.032*** (7.67)
LEVERAGE$_{t-1}$		0.071** (2.49)	0.076*** (2.69)	0.075*** (2.68)	0.075*** (2.68)
Constant	2.227*** (50.48)	1.849*** (38.28)	−0.510*** (−7.84)	−0.482*** (−10.33)	−0.323*** (−3.88)
Industry and year fixed effects	Yes	Yes	Yes	Yes	Yes
No. of obs.	31,436	31,436	32,207	32,207	32,207
Adj. R^2	0.019	0.042	0.042	0.042	0.042

Panel B. DUVOL, the Down-to-Up Volatility

	Dependent Variable: DUVOL				
Variable	1	2	3	4	5
ln(FILE_SIZE$_{t-1}$)	0.029*** (8.78)	0.009** (2.46)			
FOG$_{t-1}$			0.001 (1.01)		
AVG_NO_OF_WORDS$_{t-1}$				0.001* (1.79)	
COMPLEX_WORDS$_{t-1}$ (%)					−0.002 (−1.60)
DTURN$_{t-1}$		0.037** (2.43)	0.036** (2.39)	0.036** (2.38)	0.036** (2.39)
NCSKEW$_{t-1}$		0.024*** (8.33)	0.025*** (8.62)	0.025*** (8.62)	0.025*** (8.63)
SIGMA$_{t-1}$		−0.197*** (−2.63)	−0.158** (−2.14)	−0.161** (−2.18)	−0.158** (−2.15)

(continued on next page)

TABLE 6 (continued)
Annual Report Readability and Future Crash Risk

Panel B. DUVOL, the Down-to-Up Volatility (continued)

	Dependent Variable: DUVOL				
Variable	1	2	3	4	5
RET_{t-1}		2.616*** (11.12)	2.647*** (11.42)	2.649*** (11.43)	2.649*** (11.44)
ROA_{t-1}		0.067*** (5.62)	0.068*** (5.68)	0.068*** (5.74)	0.068*** (5.71)
$SIZE_{t-1}$		0.024*** (15.75)	0.025*** (17.97)	0.025*** (17.90)	0.025*** (17.90)
MB_{t-1}		0.014*** (7.60)	0.013*** (7.34)	0.013*** (7.33)	0.013*** (7.30)
$LEVERAGE_{t-1}$		0.022* (1.78)	0.023* (1.92)	0.023* (1.91)	0.023* (1.90)
Constant	0.959*** (49.23)	0.774*** (36.74)	−0.317*** (−11.04)	−0.310*** (−15.21)	−0.246*** (−6.80)
Industry and year fixed effects	Yes	Yes	Yes	Yes	Yes
No. of obs.	31,436	31,436	32,207	32,207	32,207
Adj. R^2	0.020	0.058	0.059	0.059	0.059

Panel C. CRASH, the Likelihood of the Occurrence of Future Extreme Downside Return Movements

	Dependent Variable: CRASH				
Variable	1	2	3	4	5
$\ln(FILE\_SIZE_{t-1})$	0.037*** (2.99)	0.032** (2.36)			
FOG_{t-1}			−0.001 (−0.17)		
$AVG\_NO\_OF\_WORDS_{t-1}$				0.0001 (0.04)	
$COMPLEX\_WORDS_{t-1}$ (%)					−0.001 (−0.20)
$DTURN_{t-1}$		−0.060 (−0.98)	−0.064 (−1.07)	−0.064 (−1.07)	−0.064 (−1.07)
$NCSKEW_{t-1}$		0.010 (0.96)	0.010 (0.95)	0.010 (0.95)	0.010 (0.95)
$SIGMA_{t-1}$		−4.034*** (−13.32)	−3.927*** (−13.15)	−3.930*** (−13.15)	−3.931*** (−13.18)
RET_{t-1}		5.941*** (6.34)	6.043*** (6.53)	6.043*** (6.53)	6.045*** (6.53)
ROA_{t-1}		−0.093** (−2.03)	−0.094** (−2.08)	−0.094** (−2.07)	−0.094** (−2.07)
$SIZE_{t-1}$		−0.003 (−0.57)	0.002 (0.29)	0.001 (0.28)	0.001 (0.27)
MB_{t-1}		0.047*** (7.53)	0.046*** (7.49)	0.046*** (7.49)	0.046*** (7.48)
$LEVERAGE_{t-1}$		0.204*** (4.38)	0.218*** (4.79)	0.218*** (4.79)	0.218*** (4.79)
Constant	−0.616 (−1.30)	−0.301 (−0.59)	−0.288 (−0.56)	−0.307 (−0.61)	−0.282 (−0.55)
Industry and year fixed effects	Yes	Yes	Yes	Yes	Yes
No. of obs.	31,421	31,421	32,191	32,191	32,191
Pseudo-R^2	0.018	0.028	0.028	0.028	0.028

Hutton et al. (2009) document that firms with more opaque financial statements are more vulnerable to crash risk. As a robustness check for the result in Table 6, we further control for the opacity of financial statements, as defined by Hutton et al. (2009). Specifically, opacity (denoted as OPAQUE) is calculated as the absolute value of discretionary accruals, according to the modified Jones model (Dechow, Sloan, and Sweeney (1995)). The idea behind this measure is that firms with large absolute values of discretionary accruals are more likely to manage reported earnings, thus revealing less firm-specific information to investors. In untabulated results, our key findings are unaffected by the inclusion of OPAQUE. Specifically, 10-K file size continues to be positively related to various crash risk measures. However, the coefficient of the Fog index remains insignificant.

In summary, the positive relation between low readability and risk of stock price crash provides additional evidence that low readability is related to managerial intention to hoard information. The heightened crash risk is also consistent with the finding that the lack of readability increases a firm's borrowing cost. Taken together, the results suggest that a firm could bury its valuation-relevant information in less readable annual reports. Such actions, however, could affect the firm's creditworthiness as perceived by lenders and increase its borrowing cost.[11]

B. The Impact of Uncertain and Weak Modal Words on Future Crash Risk

As shown in Section III.B, after controlling for 10-K readability, uncertain and weak modal word frequencies have an incremental impact on creditors' assessment of borrower risk. We now examine whether the ambiguous tone of 10-K filings is directly related to crash risk. We run regressions with the same set of control variables used in Table 6 and present the results in Table 7. Across all three panels, corresponding to different crash risk proxies, columns 1 and 3 consistently show that a large fraction of uncertain and weak modal words intensifies future crash risk. Specifically, the estimates in Panel A indicate that a 1-standard-deviation increase in the percentage of UNCERTAIN (WEAK_MODAL) would increase the probability of crash risk by 1.24% (1.07%), given the corresponding sample standard deviation of 0.342% (0.226%) shown in Panel B of Table 1. Columns 2 and 4 accentuate that 10-K readability and its unclear tone have distinct explanatory power for future crash proneness. The results suggest that managers who are trying to conceal information about the bleak prospects of their firms are also more likely to use ambiguous language in their disclosures.

V. Conclusion

Annual reports are a primary source of information for all market participants. Their lack of readability has raised concerns from regulators and the popular press regarding the effectiveness of management communication and its

[11] To address potential endogeneity issues, we employ instrumental variable analysis using two instrumental variables: i) the average 10-K file size within an industry in a given year and ii) the average readability of the reports of firms that are located within the same area (sharing the first three digits of a 5-digit zip code) in a given year. We find that the coefficient of 10-K file size remains positive and significant in both loan spread and crash risk regressions.

TABLE 7

The Ambiguous Tone of Annual Reports and Future Crash Risk

Table 7 presents the regression analysis results of stock price crash risk on the percentage of weak modal and uncertain words. Panels A and B show OLS results, and Panel C reports the results of probit regressions. The dependent variables are NCSKEW (Panel A), DUVOL (Panel B), and CRASH (Panel C). The regressors include the following control variables: $DTURN_{t-1}$, $NCSKEW_{t-1}$, $SIGMA_{t-1}$, RET_{t-1}, ROA_{t-1}, $SIZE_{t-1}$, MB_{t-1}, and $LEVERAGE_{t-1}$. All regressions also include industry and year fixed effects. The *t*-statistics (Panels A and B) and *z*-statistics (Panel C), corrected for heteroscedasticity and firm-level clustering, are reported in parentheses. All variables are defined in Appendix A. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. NCSKEW, the Negative Conditional Skewness of Future Returns

	Dependent Variable: NCSKEW			
Variable	1	2	3	4
$\ln(FILE\_SIZE_{t-1})$		0.021**		0.024***
		(2.54)		(2.88)
$WEAK\_MODAL_{t-1}$	0.194***	0.189***		
	(6.91)	(6.69)		
$UNCERTAIN_{t-1}$			0.106***	0.105***
			(5.81)	(5.75)
Other controls	Yes	Yes	Yes	Yes
Industry and year fixed effects	Yes	Yes	Yes	Yes
No. of obs.	31,627	31,349	31,627	31,349
Adj. R^2	0.043	0.043	0.043	0.043

Panel B. DUVOL, the Down-to-Up Volatility

	Dependent Variable: DUVOL			
Variable	1	2	3	4
$\ln(FILE\_SIZE_{t-1})$		0.007*		0.008**
		(1.89)		(2.20)
$WEAK\_MODAL_{t-1}$	0.080***	0.078***		
	(6.67)	(6.51)		
$UNCERTAIN_{t-1}$			0.046***	0.045***
			(5.84)	(5.80)
Other controls	Yes	Yes	Yes	Yes
Industry and year fixed effects	Yes	Yes	Yes	Yes
No. of obs.	31,627	31,349	31,627	31,349
Adj. R^2	0.060	0.059	0.059	0.059

Panel C. CRASH, the Likelihood of the Occurrence of Future Extreme Downside Return Movements

	Dependent Variable: CRASH			
Variable	1	2	3	4
$\ln(FILE\_SIZE_{t-1})$		0.028**		0.030**
		(2.00)		(2.19)
$WEAK\_MODAL_{t-1}$	0.200***	0.193***		
	(4.43)	(4.24)		
$UNCERTAIN_{t-1}$			0.133***	0.130***
			(4.37)	(4.25)
Other controls	Yes	Yes	Yes	Yes
Industry and year fixed effects	Yes	Yes	Yes	Yes
No. of obs.	31,611	31,334	31,611	31,334
Pseudo-R^2	0.029	0.029	0.029	0.029

impact on investors. This paper extends a growing body of literature on the textual analysis of corporate disclosures and investigates important consequences of less readable reports and ambiguous tone in bank lending.

We find that firms with less readable annual reports, as measured by the size of their 10-Ks, face higher loan spreads and more restrictive nonprice con-

tract terms, namely, shorter maturity and a greater likelihood of collateral requirements. Furthermore, our analysis shows that limited readability is associated with a greater likelihood of future stock price crashes. Our study thus provides the first empirical evidence that poor annual report readability increases the information risk perceived by creditors and leads to higher external financing costs. Moreover, our paper accentuates the importance of using proper readability measures and word-classification schemes that are carefully derived for business usage (e.g., Loughran and McDonald (2011), Loughran and McDonald (2014)).

Our results indicate that, in addition to readability, the ambiguous tone of the annual reports is closely related to managers' information-concealing behavior and results in less favorable loan terms. Specifically, we find that a high percentage of uncertain and weak modal text in business disclosures increases loan spreads and the likelihood of security requirement. The frequency of uncertain and weak modal words is also positively associated with stock price crash risk, suggesting that managers might use ambiguous language to obscure adverse news from investors.

Given that banks are a group of informationally privileged investors that routinely screen borrowers and are experienced in processing less readable financial disclosures, our findings suggest that low readability might not just be an issue of falling short in "plain English." Lack of readability and ambiguous tone in annual reports are related to a distorted corporate information environment and impair a firm's perceived creditworthiness. Importantly, the impact of 10-K readability and tone could be prevalent across stakeholders of a firm. Ultimately, corporate stakeholders are not only afflicted with a less effective communication of valuation-relevant information but also incur a higher cost of external financing.

Appendix A. Variable Definitions

1. Proxies of Readability and Tone Ambiguity

ln(FILE_SIZE): The natural logarithm of the file size in megabytes of the SEC EDGAR "complete submission text file" for the 10-K filing. We thank Tim Loughran and Bill McDonald for making the file size data available at http://www3.nd.edu/~mcdonald/Data/LoughranMcDonald_10-X_FileSize.csv.

FOG: Defined as (AVG_NO_OF_WORDS + COMPLEX_WORDS) × 0.4. The index is interpreted as the number of years of formal education required for a person of average intelligence to read the document once and understand it. A high value of the Fog index implies less readable text.

AVG_NO_OF_WORDS: Average number of words per sentence, calculated as the number of words in the 10-K divided by the total number of sentences.

COMPLEX_WORDS: The percentage of 10-K complex words. A complex word is defined as one with three or more syllables.

WEAK_MODAL: The percentage of weak modal words as defined by Loughran and McDonald (2011).

UNCERTAIN: The percentage of words conveying uncertainty as defined by Loughran and McDonald (2011).

2. Bank Loan Characteristics

AIS_DRAWN: The DealScan all-in-spread drawn measuring the all-inclusive cost of a drawn loan measured in basis points.

ln(SPREAD): The natural logarithm of the DealScan all-in-spread drawn.

LOAN_SIZE: The amount of a loan in millions of dollars.

MATURITY: The loan maturity measured in months.

SECURITY: The dummy variable indicating the collateral requirement.

3. Controls for the Analysis of Bank Loans

ln(ASSETS): The natural logarithm of total assets.

MB: The ratio of the market value of equity to the book value of equity at the end of the fiscal year.

LEVERAGE: The book value of long-term debt scaled by total assets.

PROFITABILITY: The ratio of earnings before interest, taxes, depreciation, and amortization (EBITDA) to total assets.

TANGIBILITY: The ratio of total tangible assets (property, plant, and equipment (PPE)) to book value of total assets, following Campello and Giambona (2013).

EDF: A firm's expected default frequency, following Bharath and Shumway (2008).

Loan-type dummies: Indicating different loan types (i.e., revolving loans, term loans, and 364-day loans).

Loan-purpose dummies: Indicating different loan purposes (i.e., acquisition, commercial paper (CP) backup, general corp. purpose, debt repayment, takeovers, and working capital).

BSEG: The sum of the squared business segment proportions as reported for the firm in the Compustat Segment data.

ln(FIRM_AGE): The natural logarithm of the number of years since the firm first appeared in the CRSP database.

4. Crash Risk Measures

NCSKEW: The negative conditional skewness of future returns. As in the related literature, NCSKEW is defined as the standardized negative value of the third central moment of firm-specific weekly return scaled by its sample variance raised to the power of 3/2. We follow the literature by employing the negative value of the skewness to ease interpretation so that a larger value of NCSKEW indicates more negatively skewed returns and thus greater crash risk.

DUVOL: For each stock i over fiscal year t, we first group firm-specific weekly returns into "up" weeks, in which the returns are greater than the stock's annual average return, and "down" weeks, in which the returns are below the average. Then, DUVOL is calculated as the logarithm of the ratio of the standard deviation of firm-specific weekly returns in "down" weeks to that of the "up" weeks. A large value of DUVOL suggests that the stock has large downside price deviations (i.e., great crash risk).

CRASH: A dummy variable that takes the value of 1 if a firm experiences 1 or more crash weeks in a fiscal year, and 0 otherwise. A firm's crash weeks are defined as those during which the firm-specific return is below 3.2 standard deviations of the average firm-specific weekly returns over the entire fiscal year.

5. Controls for the Analysis of Crash Risk

DTURN: The average monthly share turnover over the current fiscal year minus the average monthly share turnover over the previous fiscal year, where monthly share turnover

is calculated as the monthly trading volume divided by the total number of shares outstanding during the month.

SIGMA: The standard deviation of firm-specific weekly returns over the fiscal year.

RET: The mean of firm-specific weekly returns over the fiscal year.

ROA: Income before extraordinary items divided by lagged total assets.

SIZE: The natural logarithm of the market value of equity at the end of the fiscal year.

MB: The ratio of the market value of equity to the book value of equity at the end of the fiscal year.

LEVERAGE: The book value of long-term debt scaled by total assets.

OPAQUE: The absolute value of discretionary accruals, which are measured using the modified Jones model (Dechow et al. (1995)).

Appendix B. Measures of Firm-Level Crash Risk

Specifically, for firm i during its fiscal year t, we first estimate firm-specific weekly residual returns from the expanded market model shown as follows:

$$\text{(B-1)} \qquad r_{i,\tau} = \alpha_i + \beta_{1,i} r_{m,\tau-2} + \beta_{2,i} r_{m,\tau-1} + \beta_{3,i} r_{m,\tau} + \beta_{4,i} r_{i,\tau+1} + \beta_{5,i} r_{m,\tau+2} + \varepsilon_{i,\tau},$$

where $r_{i,\tau}$ is the return on stock i in week τ, and $r_{m,\tau}$ is the return on the CRSP value-weighted market index in week τ. The lead and lag terms of the market return are included to account for nonsynchronous trading (Dimson (1979)). The weekly returns are used partly to avoid the concern caused by thinly traded stocks. To exclude firms that went public, were delisted, or experienced trading halts, we drop a firm that was traded for less than 26 weeks over a fiscal year. Similar practices are adopted by Morck, Yeung, and Yu (2000) and An and Zhang (2013). The firm-specific weekly return (denoted as $W_{i,\tau}$) is defined as the natural logarithm of 1 plus the residuals from equation (B-1).

Return asymmetry is generally appraised by negative skewness. Our first measure of crash risk is NCSKEW, the negative conditional skewness of future returns. As in the related literature, NCSKEW is defined as the standardized negative value of the third central moment of firm-specific weekly return scaled by its sample variance raised to the power of 3/2. More specifically, NCSKEW of stock i in its fiscal year t is calculated as

$$\text{(B-2)} \qquad \text{NCSKEW}_{i,t} = \frac{-n(n-1)^{3/2} \sum_{\tau \in t} W_{i,\tau}^3}{(n-1)(n-2)(\sum W_{i,\tau}^2)^{3/2}},$$

where n is the number of weekly observations in year t. We follow the literature by employing the negative value of the skewness to ease interpretation so that a larger value of NCSKEW indicates more negatively skewed returns and thus greater crash risk.

To construct the second crash risk proxy, DUVOL, for each stock i over fiscal year t, we first group firm-specific weekly returns into "up" weeks, in which the returns are greater than the stock's annual average return, and "down" weeks, in which the returns are below the average. Then, we calculate DUVOL as the logarithm of the ratio of the standard deviation of firm-specific weekly returns in "down" weeks to that of the "up" weeks. A large value of DUVOL suggests that the stock has large downside price deviations (i.e., great crash risk).

The third measure of crash risk is a dummy variable that takes the value of 1 if a firm experiences one or more crash weeks in a fiscal year, and 0 otherwise. Consistent with the literature on crash risk, a firm's crash weeks are defined as those during which the firm-specific return is below 3.2 standard deviations of the average firm-specific weekly returns over the entire fiscal year.

References

An, H., and T. Zhang. "Stock Price Synchronicity, Crash Risk, and Institutional Investors." *Journal of Corporate Finance*, 21 (2013), 1–15.

Barclay, M. J., and C. W. Smith Jr. "The Maturity Structure of Corporate Debt." *Journal of Finance*, 50 (1995), 609–631.

Barry, C. B., and S. J. Brown. "Differential Information and the Small Firm Effect." *Journal of Financial Economics*, 13 (1984), 283–294.

Berger, A. N., and G. F. Udell. "Collateral, Loan Quality, and Bank Risk." *Journal of Monetary Economics*, 25 (1990), 21–42.

Bharath, S. T.; S. Dahiya; A. Saunders; and A. Srinivasan. "Lending Relationships and Loan Contract Terms." *Review of Financial Studies*, 24 (2011), 1141–1203.

Bharath, S. T., and T. Shumway. "Forecasting Default with the Merton Distance to Default Model." *Review of Financial Studies*, 21 (2008), 1339–1369.

Bharath, S. T.; J. Sunder; and S. V. Sunder. "Accounting Quality and Debt Contracting." *Accounting Review*, 83 (2008), 1–28.

Biddle, G. C.; G. Hilary; and R. S. Verdi. "How Does Financial Reporting Quality Relate to Investment Efficiency?" *Journal of Accounting and Economics*, 48 (2009), 112–131.

Bloomfield, R. J. "The 'Incomplete Revelation Hypothesis' and Financial Reporting." *Accounting Horizons*, 16 (2002), 233–243.

Campello, M., and E. Giambona. "Real Assets and Capital Structure." *Journal of Financial and Quantitative Analysis*, 48 (2013), 1333–1370.

Chava, S.; D. Livdan; and A. Purnanandam. "Do Shareholder Rights Affect the Cost of Bank Loans?" *Review of Financial Studies*, 22 (2009), 2973–3004.

Chen, J.; H. Hong; and J. C. Stein. "Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices." *Journal of Financial Economics*, 61 (2001), 345–381.

Dechow, P. M.; R. G. Sloan; and A. P. Sweeney. "Detecting Earnings Management." *Accounting Review*, 70 (1995), 193–225.

Dimson, E. "Risk Measurement When Shares Are Subject to Infrequent Trading." *Journal of Financial Economics*, 7 (1979), 197–226.

Easley, D.; S. Hvidkjaer; and M. O'Hara. "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*, 57 (2002), 2185–2221.

Easley, D., and M. O'Hara. "Information and the Cost of Capital." *Journal of Finance*, 59 (2004), 1553–1583.

Graham, J. R.; S. Li; and J. Qiu. "Corporate Misreporting and Bank Loan Contracting." *Journal of Financial Economics*, 89 (2008), 44–61.

Greene, W. H. *Econometric Analysis*, 3rd ed. Upper Saddle River, NJ: Prentice Hall (1997).

Hutton, A. P.; A. J. Marcus; and H. Tehranian. "Opaque Financial Reports, R^2, and Crash Risk." *Journal of Financial Economics*, 94 (2009), 67–86.

Jin, L., and S. C. Myers. "R^2 around the World: New Theory and New Tests." *Journal of Financial Economics*, 79 (2006), 257–292.

Kim, J.; Y. Li; and L. Zhang. "CFOs versus CEOs: Equity Incentives and Crashes." *Journal of Financial Economics*, 101 (2011a), 713–730.

Kim, J.; Y. Li; and L. Zhang. "Corporate Tax Avoidance and Stock Price Crash Risk: Firm-Level Analysis." *Journal of Financial Economics*, 100 (2011b), 639–662.

Lawrence, A. "Individual Investors and Financial Disclosure." *Journal of Accounting and Economics*, 56 (2013), 130–147.

Lehavy, R.; F. Li; and K. Merkley. "The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts." *Accounting Review*, 86 (2011), 1087–1115.

Li, F. "Annual Report Readability, Current Earnings, and Earnings Persistence." *Journal of Accounting and Economics*, 45 (2008), 221–247.

Li, S.; J. Qiu; and C. Wan. "Corporate Globalization and Bank Lending." *Journal of International Business Studies*, 42 (2011), 1016–1042.

Loughran, T., and B. McDonald. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66 (2011), 35–65.

Loughran, T., and B. McDonald. "IPO First-Day Returns, Offer Price Revisions, Volatility, and Form S-1 Language." *Journal of Financial Economics*, 109 (2013), 307–326.

Loughran, T., and B. McDonald. "Measuring Readability in Financial Disclosures." *Journal of Finance*, 69 (2014), 1643–1671.

Morck, R.; B. Yeung; and W. Yu. "The Information Content of Stock Markets: Why Do Emerging Markets Have Synchronous Stock Price Movements?" *Journal of Financial Economics*, 58 (2000), 215–260.

Ortiz-Molina, H., and M. F. Penas. "Lending to Small Businesses: The Role of Loan Maturity in Addressing Information Problems." *Small Business Economics*, 30 (2008), 361–383.

Smith, C. W., Jr., and J. B. Warner. "On Financial Contracting: An Analysis of Bond Covenants." *Journal of Financial Economics*, 7 (1979), 117–161.

Wittenberg-Moerman, R. "The Role of Information Asymmetry and Financial Reporting Quality in Debt Trading: Evidence from the Secondary Loan Market." *Journal of Accounting and Economics*, 46 (2008), 240–260.