Review of Accounting Studies (2021) 26:815–863
https://doi.org/10.1007/s11142-020-09575-4

Measuring credit risk using qualitative disclosure

John Donovan[1] · Jared Jennings[2] · Kevin Koharki[3] · Joshua Lee[4]

Accepted: 11 December 2020/ Published online: 11 January 2021

Abstract

We use machine learning methods to create a comprehensive measure of credit risk based on qualitative information disclosed in conference calls and in management's discussion and analysis section of the 10-K. In out-of-sample tests, we find that our measure improves the ability to predict credit events (bankruptcies, interest spreads, and credit rating downgrades), relative to credit risk measures developed by prior research (e.g., z-score). We also find our measure based on conference calls explains *within-firm* variation in future credit events; however, we find little evidence that the measures of credit risk developed by prior research explain *within-firm* variation in credit risk. Our measure has utility for both academics and practitioners, as the majority of firms do not have readily available measures of credit risk, such as actively-traded CDS or credit ratings. Our study also adds to the growing body of research using machine-learning methods to gather information from conference calls and MD&A to explain key outcomes.

Keywords Credit risk · Disclosure · Machine-learning · Textual analysis

JEL classifications G20 · G23 · G30 · G32 · G33 · M40 · M41

✉ Jared Jennings
jaredjennings@wustl.edu

John Donovan
jdonova4@nd.edu

Kevin Koharki
kkoharki@purdue.edu

Joshua Lee
joshlee84@byu.edu

Extended author information available on the last page of the article

1 Introduction

We introduce a comprehensive and distinct measure of credit risk derived from the qualitative information in firms' public disclosures. Banks, credit rating agencies, and underwriters consider both quantitative and qualitative inputs when assessing a firm's credit risk (e.g., Ganguin and Bilardello 2005; Agarwal and Hauswald 2010; Bozanic and Kraft 2015). A large body of research in economics, finance, and accounting examines the determinants and consequences of credit risk. Research primarily focuses on quantitative signals, such as Altman's (1968) Z-score, Ohlson's (1980) O-score, and expected default frequency (EDF), to assess credit risk and does not directly consider qualitative information from firm disclosures, largely due to the difficulty in identifying, measuring, and properly weighting the latter.[1] Recent advances in computing technology have allowed researchers to begin to explore the usefulness of qualitative information from firm disclosures. For example, Mayew et al. (2015) use a dictionary approach to provide evidence that language questioning the firm's ability to continue as a going-concern in Management's Discussion and Analysis in the 10-K (MD&A) predicts bankruptcy.

We introduce supervised machine learning technology to this stream of research as an alternative means of extracting credit risk information from qualitative disclosures. Supervised machine learning models can add significant explanatory power to credit risk measures, because they use computer algorithms to identify patterns in disclosures that researchers might overlook. We use these models to create comprehensive measures of credit risk using the qualitative information contained in two firm-level disclosures: the earnings conference call and MD&A section of the annual 10-K report. We find that our measures add significant explanatory power in measuring credit risk, relative to and incremental to other measures.[2] For example, when predicting bankruptcy, we find that the incremental explanatory power of our measure is at least six times greater than that of other credit risk measures. We also compare the relative usefulness of the language in the MD&A and conference call for assessing firm credit risk. Our measures derived from conference calls and MD&As both explain *across-firm* variation in credit risk, but only the measures derived from conference calls explain *within-firm* variation. Furthermore, measures of credit risk do not consistently explain *within-firm* variation in credit risk. We expect that our measures can be particularly useful when other market-based measures of credit risk (e.g. credit ratings or credit default swap spreads) are not available.

We analyze both conference calls and MD&As, because these disclosures potentially contain information relevant for assessing firms' credit risk. We expect conference calls to be particularly useful, as they are held quarterly and represent a timely source of information. Ganguin and Bilardello (2005) suggest that credit rating agencies regularly gather and process quantitative and qualitative information from

[1] We define quantitative signals as signals that are based on numerical data (e.g., returns). EDF, which is derived from Merton (1974), uses market-based numerical information (e.g., stock price volatility) to estimate the likelihood of firm default. EDF is likely a function of quantitative and qualitative information, given that market-based information is a function of both qualitative and quantitative information.

[2] Barth et al. (2008) and Beatty et al. (2008) estimate the association between credit ratings and firm fundamentals and then use the coefficients and firm fundamentals to calculate an estimated credit rating for all firms, regardless of whether they have credit ratings.

conference calls. Moreover, the interactive nature of the Q&A session with analysts creates a dynamic setting, which can increase the likelihood of spontaneous and informative disclosure about the firm's creditworthiness. We also analyze the content of the MD&A, because research suggests that it contains information relevant for assessing credit risk (Mayew et al. 2015).

We recognize, however, that these disclosures may contain little credit-risk relevant information for several reasons. Disclosures may be primarily focused on providing information to equity investors, whose payoffs differ from those of debtholders (e.g. Watts 2003). In addition, managers may use their discretion to strategically avoid providing information about credit risk in their disclosures. For example, managers may avoid revealing credit-risk-relevant information during conference calls by scripting their responses to analysts' questions (Lee 2016), by refusing to respond to specific questions (Hollander et al. 2010), or by refusing to call on certain analysts (Mayew 2008). Managers and lawyers also carefully vet the MD&A, potentially omitting relevant information about credit risk. In addition, managers may only discuss credit risk in their disclosures after it has significantly deteriorated.

Language in firm disclosures is multi-faceted and attempts to convert qualitative information into useable numerical indices inevitably results in a loss of information and context, because the numerical indices can only capture a portion of what a human interpreter would glean from the original text (Liberti and Petersen 2019). Research often relies on researcher-generated dictionaries to summarize qualitative disclosure content (e.g., Loughran and McDonald 2011; Bozanic et al. 2018). For example, Mayew et al. (2015) generate a dictionary of words and phrases relating to a firm's ability to continue as a going-concern to identify explicit discussions of bankruptcy risk. However, if a dictionary is incomplete, the dictionary-based measure may not comprehensively capture relevant economic events described in firm disclosures, leading researchers to *understate* the importance of qualitative information. Supervised machine-learning methods can improve researchers' ability to extract relevant information from firm disclosures and identify information in disclosures that relates to the construct of interest (i.e., credit risk) that may not be identified ex ante by a researcher. Supervised machine-learning methods also assign varying weights to phrases that predict the construct, which increase their power to capture the construct of interest, whereas measures based on researcher-generated dictionaries typically equally weight each word.

Using samples of disclosures from 2002 to 2016, we use three supervised machine-learning methods (support vector regressions [SVR], supervised Latent Dirichlet Allocation [sLDA], and random forest regression trees [RF]) to directly map the language in conference calls and MD&As to credit default swap (CDS) spreads. We train the models using CDS spreads, because they represent a comprehensive and timely measure of a firm's concurrent credit risk (Blanco et al. 2005; Hull et al. 2004; Norden and Weber 2004; Standard and Poor's 2007; Das et al. 2014) and therefore allow the machine learning models to identify comprehensive and timely information in disclosures related to credit risk. We apply the model parameters estimated from training samples to out-of-sample disclosures to create estimates of credit risk based on CDS spreads. We then use factor analysis to combine the output of the three machine learning models into a single text-based credit risk score (hereafter, *TCR Score*) for each disclosure. We use a split-sample and rolling-window approach to create out-of-sample estimates of credit risk. In the split-sample approach, we train the machine-learning models using all observations from 2002 to 2012 and test whether the out-of-sample estimates capture credit risk using all observations from 2013 to 2016. In the rolling-window approach, we estimate the models on a yearly basis. We train the models using all observations over the preceding four years and test whether the out-of-sample estimates capture credit risk using observations in the current year. These two approaches yield two *TCR Score* estimates for conference calls and two for the MD&A.

We validate the *TCR Score* measures by examining their associations with actual CDS spreads in out-of-sample data. We find statistically and economically significant associations between CDS spreads and all four *TCR Score* measures. To provide economic insights into the credit-risk information captured by these machine-learning methods, we also classify the most influential phrases and topics identified by each method into categories. We find that the RF model primarily identifies language that is associated with liquidity, debt, and performance. The SVR and sLDA models primarily identify phrases that are associated with performance, industry, and accounting-specific information. By connecting the language identified by the machine-learning methods to economic intuition, we draw a tighter link between the construct of credit risk and our proxy.

One benefit of our approach is that we can apply the machine learning models to any conference call or MD&A, even if the firm does not have actively traded CDS. This benefit is especially important, given that fewer than 10% of borrowers with long-term debt have CDS spreads. We demonstrate the construct validity of the *TCR Score* measures for a sample of firms *without* CDS spreads by examining their associations with *future* credit-related events. The *TCR Score* measures capture forward-looking credit-risk information, because the machine learning models are fit to CDS spreads, which implicitly capture the likelihood that the borrower will default. We first provide evidence that each *TCR Score* measure is positively associated with future bankruptcy filings, the initial interest rate spreads charged in future private debt contracts, and future credit rating downgrades. We also find that the explanatory power of each measure is incremental to alternative measures of credit risk (EDF, estimated credit ratings, O-score, and Z-score) as well as firm characteristics. Furthermore, the incremental explanatory power is economically significant.

Second, we examine whether the *TCR Score* measures explain *within-firm* variation in credit risk by including firm fixed effects, time fixed effects (year-quarter for quarterly tests and year for annual tests), and other firm-specific controls in the regression models. In all cases, we continue to find a positive and significant association between *future* credit events and the *TCR Score* estimates based on the conference call; however, we do not find consistent evidence of a significant association between future credit events and *TCR Score* estimates based on the MD&A. This evidence suggests that the *TCR Score* measures based on the conference call capture within firm variation in credit risk but the measures based on the MD&A do not. None of the existing credit risk measures (e.g., EDF) consistently explains *within-firm* variation in credit risk.

The result that the *TCR Score* measure based on the conference call is incremental to EDF, a market-based measure of credit risk, may suggest that equity prices do not fully capture credit-risk-relevant information. To further test whether our measures capture information impounded by future market participants, we examine whether *TCR Score* predicts negative equity market responses, bond yields, and market responses to credit rating downgrades. We find that *TCR Score* estimates based on the conference call but not those based on the MD&A are positively associated with future bond yields and large negative equity market price shocks, which we define as a monthly equity stock return of less −30% (Zhang 2008). We also find some evidence that *TCR Score* based on the conference call predicts the stock market response to *future* credit rating downgrade announcements.

This paper provides several contributions to the literature. First, we create a comprehensive measure of credit risk using the narrative content of disclosure that incrementally and more comprehensively predicts multiple credit-related outcomes. Studies have developed improved measures of credit risk, but the incremental improvements tend to be relatively small (e.g., Hillegeist et al. 2004). Our evidence suggests that a text-based credit risk measure generated by machine learning methods yields economically significant improvements.[3] In addition, research primarily has evaluated credit risk measures based on their ability to predict bankruptcy, an extreme outcome. For example, Mayew et al. (2015) find that explicit discussion of the firm's ability to continue as a going-concern predicts bankruptcy.[4] We provide evidence that our measure not only outperforms other credit risk measures in predicting bankruptcy but also is the *only* measure that consistently predicts another negative credit event (credit rating downgrades) and the cost of debt among creditworthy firms raising private debt capital (i.e., interest rate spreads).

Second, we compare the relative usefulness of the language in the MD&A and conference call for assessing credit risk. Our findings suggest that the narrative content of both the conference call and MD&A capture *across-firm* variation in credit risk; however, our results also suggest that *only* the narrative content of the conference call consistently explains *within-firm* variation for several future credit events (bankruptcy, interest spreads, and credit rating downgrades). We do not find evidence that the MD&A or other credit risk measures (e.g., Z-score) consistently capture *within-firm* variation in credit risk. The dynamic nature of the conference call is conducive to more spontaneous and informative responses, compared to the MD&A, which may explain

[3] Hillegeist et al. (2004) suggest that EDF is a more comprehensive measure of credit risk, due to its higher "information content" than the Altman Z-Score and Ohlson O-Score. However, Hillegeist et al. (2004) caution that the information content scores of these three measures are still quite low. Hillegeist et al. (2004) provide evidence that EDF alone yields a 12% pseudo-R^2 when predicting future bankruptcies while the O-Score, which is the next best measure in their paper, yields a 10% pseudo-R^2. In contrast, we find that one of our measures based on conference calls yields a pseudo-R^2 of 34.2% when predicting future bankruptcies while the O-score, which is the next best measure in our empirical tests, yields a pseudo-R^2 of 10.9%.

[4] Since expressing doubt that that the firm will continue as a going-concern is a dichotomous signal, the going-concern disclosure employed by Mayew et al. (2015) cannot be used to detect high or meaningful increases in credit risk for firms that are not on the verge of bankruptcy. Mayew et al. (2015) find that *explicit* discussion of the firm's ability to continue as a going-concern is very rare; only approximately 3% of all firm-years and only 39% of firms that eventually file for bankruptcy *explicitly* discuss their ability to continue as a going-concern. In contrast, our machine-learning methods can identify language that is *implicitly* (rather than only *explicitly*) related to credit risk, leading to a continuous measure of credit risk that can be applied to a broader set of firms to measure meaningful levels of or changes in credit risk. This comment is not a critique of Mayew et al. (2015), because their aim is *not* to develop a measure of credit risk but to provide empirical evidence to inform the debate on whether the FASB should require management to assess and disclose the entity's ability to continue as a going-concern in the MD&A.

these differences. Our results also suggest that disclosure tone (including the tone of the conference call) is limited in its ability to explain *within-firm* variation in credit risk.

Third, we connect the language identified by the machine-learning methods to human intuition and economic insights by manually classifying the phrases and topics identified by each machine-learning method into categories, which shows that our measures derived using machine-learning methods capture credit risk and not just spurious correlations with future credit events. Our analysis illustrates a key advantage of using machine-learning to extract content from firm disclosures, relative to other textual analysis approaches (e.g., dictionary approach). Machine-learning methods identify a wider range of phrases and topics that may not be identified ex ante by a researcher. Our results suggest a wide range of intuitive words that managers use to communicate the firm's credit risk. Other textual analysis approaches (e.g., disclosure tone) can be limited in their ability to identify relevant words, which could reduce the usefulness of these approaches.[5]

We believe that our measures derived from firm disclosures can be useful to academics and practitioners in several ways. Other measures (Z-score, O-score, or EDF) have garnered thousands of citations from researchers. This general interest in estimates of credit risk highlights the importance of assessing credit risk in academic research. For example, academics could use our measure to estimate a firm's credit risk more precisely or to study the relative importance of qualitative information when estimating credit risk. In practice, arms-length capital providers, who do not have access to private information (e.g., investors in the bond market or syndicate members who are not the lead arranger), could use our measure to supplement their models to obtain a more comprehensive and independent estimate of credit risk. We expect our measure to be particularly useful when alternative market measures of credit risk (e.g., CDS spreads and credit ratings) are not available, as the majority of borrowers with long-term debt do not have CDS spreads or credit ratings. We note that our study resembles other studies that develop and improve upon other proxies for difficult-to-measure constructs, such as disclosure quality (e.g., Chen et al. 2015) and litigation risk (e.g., Kim and Skinner 2012).

2 Background

Sophisticated debt market participants use both quantitative and qualitative information to assess borrowers' credit risk (Ganguin and Bilardello 2005; S&P 2015). One way that debt market participants gather qualitative information is from public firm disclosures (e.g., conference calls and 10-K reports). Firm disclosures likely contain credit risk relevant information, such as expectations regarding future earnings, future cash flows, current debt, and other relevant information for determining a firm's ability to make its required interest and principal payments. Disclosures also contain firm-specific, industry-specific, and management-specific information, which can be used

[5] For example, Mayew et al. (2015) creates a dictionary of language that questions the firm's ability to continue as a going-concern. We provide evidence that the machine-learning methods extract a wider range of words associated with firm credit risk (e.g., firm performance, debt, liquidity, industry), which are distinct from the explicit discussion of the firm's ability to continue as a going-concern.

by debt market participants who consider the firm's competitive position within the industry, management's ability, and management's integrity when assessing credit risk (Bonsall et al. 2016, 2017a). Therefore qualitative information in firm disclosures is likely useful in assessing a borrower's credit risk. Anecdotally, Ganguin and Bilardello (2005) suggest that credit rating agencies regularly gather and process qualitative information from quarterly earnings conference calls, among other sources.[6] Although not directly related to qualitative firm disclosures, Campbell et al. (2019) show that consumer loan officers use qualitative (or soft) information to assess consumers' credit risk. Interestingly, they find that loan officer inattentiveness impairs the usefulness of qualitative information for assessing credit risk. Costello et al. (2019) also provide evidence that discretion does not always improve upon machine-generated credit risk assessments. Collectively this evidence supports our contention that machine-learning methods may help identify credit-relevant qualitative information, because these methods are not subject to behavioral biases and because researchers may overlook important aspects of credit risk that will be captured by our approach.

Despite the potential importance of qualitative information in assessing borrowers' credit risk, academic research focuses predominantly on quantitative-based signals to develop measures of credit risk, such as the Z-score, O-score, and EDF. These measures generally include quantitative values from financial statements or historical stock price movements. Stock prices likely impound some qualitative information from firm disclosures; however, credit risk proxies do not allow us to disentangle the importance of qualitative and quantitative information. The focus on quantitative signals in research is largely due to the difficulty in *directly* capturing and quantifying relevant information from qualitative disclosures, such as conference calls. One notable exception is the work of Mayew et al. (2015), who provide evidence that the tone of the MD&A and *explicit* discussion of the firm's ability to continue as a going-concern are positively associated with the likelihood of future bankruptcy.

We use multiple supervised machine-learning methods, rather than a dictionary approach, to capture credit risk-relevant qualitative information directly from firm disclosures. Supervised machine-learning methods automate the identification of language that explains a variable of interest (Frankel et al. 2016). These methods identify a wider range of phrases and topics that may not be identified ex ante by a researcher. Other textual analysis approaches can be limited in their ability to identify relevant words, which could reduce their usefulness in capturing firm-specific characteristics (e.g., credit risk). In addition, supervised machine-learning methods provide greater flexibility by assigning varying importance to individual phrases and topics. Researchers and practitioners have used machine-learning methods in medical diagnosis, advertising, product placement, stock market trading, air flight simulators, financial portfolio management, résumé screening, speech and facial recognition, among many others. According to Barron's (2017), "[Machine learning] is now increasingly becoming accepted as a useful tool for decision-making in the corporate world." Academic researchers also use supervised machine-learning methods to quantify and directly predict firm fundamentals (e.g., contemporaneous accruals, future cash flows, future

[6] Conversations with personnel of both Moody's and S&P confirm that listening to or engaging in firms' quarterly conference calls is one of several key inputs to the rating process and is considered a mandatory component of an analyst's research function.

earnings surprises, financial statement fraud) using qualitative firm disclosures (Cecchini et al. 2010; Frankel et al. 2016, 2019).[7] One advantage to using supervised machine-learning methods to explain a firm fundamental is that these methods can be easily applied to a wide variety of dependent variables, languages, and contexts.

Conversely, qualitative disclosures may not contain a significant amount of information that is relevant for predicting credit events beyond that identified by research. For example, in a conference call context, managers may use strategies to avoid revealing credit-risk relevant information, such as scripting their responses to questions posed by analysts (Lee 2016) or refusing to respond to questions that refer to credit-risk-relevant information (Hollander et al. 2010). In addition, managers may provide little credit-risk-relevant information during conference calls, because these calls are typically designed to provide information to shareholders, whose incentives differ from those of lenders. Qualitative MD&A disclosures may also contain little credit-risk-relevant information if managers and lawyers carefully craft the MD&A to omit information that could be used to predict credit-related events. In addition, managers may only discuss credit risk when it has significantly deteriorated, thus reducing its usefulness for predicting credit events for the average firm. In addition, the methods that attempt to turn qualitative information into useable numerical indices inevitably result in a loss of information and context, because the numerical indices can only capture a portion of what a human interpreter would glean from the original text (Liberti and Petersen 2019). Therefore whether these methods capture qualitative information from firm disclosures that is relevant for assessing firms' credit risk is an empirical question.

3 Data

3.1 Sample selection

We obtain a sample of 132,060 quarterly conference call transcripts from Factiva's FD Wire between 2002 and 2016 to analyze the narrative content of conference calls. We obtain a sample of 92,353 annual 10-Ks with MD&A between 2002 and 2016 to analyze the narrative content of the MD&A. We collect financial statement data from Compustat, stock return data from CRSP, and analyst forecast data from I/B/E/S. CDS spreads are obtained from the Markit database. We collect interest spreads from private debt contracts from Dealscan.[8] S&P's firm-level credit ratings are obtained from Compustat. We obtain bankruptcy information from two sources: (1) Moody's

[7] Research also uses static dictionaries to measure disclosure characteristics, such as tone (Henry 2008; Tetlock et al. 2008; Loughran and McDonald 2011; Price et al. 2012; Henry and Leone 2016; Davis et al. 2012; Feldman et al. 2010), firm risk (Campbell et al. 2014; Kravet and Muslu 2013), or the extent of forward-looking information (Li 2010; Muslu et al. 2014). Other studies use summary measures, such as FOG, BOG, word length, and file size, to measure disclosure readability (Li 2008; Bonsall et al. 2017c; Loughran and McDonald 2014). Still other studies use more sophisticated techniques, such as Latent Dirichlet Allocation (i.e., LDA), to summarize information in disclosures (Dyer et al. 2017; Huang et al. 2018; Bao and Datta 2014). Generally, these methods are not easily adapted to alternative contexts.

[8] We thank Michael Roberts for providing the Compustat-Dealscan linking table, available on WRDS. Refer to Chava and Roberts (2008) for additional details.

Ultimate Recovery Rate Database and (2) the UCLA Bankruptcy Research Database.[9] For all tests, we exclude banks (SIC 6000–6299) and utilities (SIC 4000–4999) from our analyses to be consistent with research that suggests that these firms have significantly different credit risk characteristics than industrial firms (Bonsall et al. 2017b; Morgan 2002). The sample size for each test varies based on data availability. We discuss each subsample in greater detail below.

3.2 Measures of credit risk using machine-learning methods

The goal of our study is to produce a text-based measure of credit risk that can be applied to a wide variety of firms, such as those with private debt, with public debt, or without a market-based measure (e.g., CDS spread) of credit risk.[10] We use three supervised machine-learning methods to estimate CDS spreads using the language in the quarterly conference calls and annual MD&As.[11] We use CDS spreads to train the machine-learning methods, because they represent a comprehensive and timely measure of credit risk (Blanco et al. 2005; Hull et al. 2004; Norden and Weber 2004).[12] CDS spreads inherently reflect forward-looking information regarding the borrower's ability to make the required interest and principal payments. We measure CDS spread (*CDS*) as the average daily five-year CDS spread for firm *i* in the [−91,+1] window surrounding the conference call date or annual report date.[13] We use the average credit risk, as measured by CDS spreads, over the window to reduce the volatility that could exist in examining a shorter window. We include CDS spreads through the conference call or annual report date (day +1) to allow for the possibility that institutional investors in the CDS market learn about the firm's creditworthiness through the conference call or annual report (Kim et al. 2019).

We use the counts of all one- and two-word phrases in the conference call transcripts and MD&A disclosures as inputs to the machine-learning models (SVR, sLDA, and RF regression trees). We reduce the influence of sparsity by stemming all words using the Porter Stemmer algorithm and removing any one- or two-word phrase that is used in

[9] We thank Lynn LoPucki for providing his data, available at http://lopucki.law.ucla.edu/index.htm.

[10] Disclosures may differ for firms with public and private debt. Research suggests that firms with lower disclosure quality (e.g., Bharath et al. 2008; Dhaliwal et al. 2011) are more likely to access the private debt market. Vashishtha (2014) shows that borrowers with bank debt are less likely to issue earnings forecasts following a covenant violation. Christensen et al. (2019) find that the likelihood of issuing non-GAAP earnings decreases following a covenant violation. Despite the potential differences in disclosure for firms with public and private debt, we provide evidence that *TCR Score* explains within-firm variation in credit risk for both firms with private and public debt.

[11] Firms that hold conference calls may fundamentally differ from those that do not hold conference calls. Therefore there may be a selection bias when it comes to the conference call results. Note that 42.5% of all firm-quarters on Compustat with public equity and positive total assets host conference calls between 2002 and 2016 and 65.3% of the firms in our MD&A sample hold at least one conference call during the year.

[12] An alternative approach would be to directly train the model to predict credit events, such as bankruptcies or credit rating downgrades; however, these events occur somewhat infrequently, possibly reducing our ability to adequately train the model. Another alternative approach would be to train the model on other outcomes, such as interest spreads in loan contracts; however, these contracts are only observed immediately after contract negotiation for creditworthy firms. Therefore it is less likely that this approach will identify language that is useful if a firm's credit risk significantly deteriorates after contract inception.

[13] In an additional robustness test, we change the window over which we measure the average CDS spread to be −5 to +5 and − 45 to +45 and find qualitatively similar results. We tabulate and report these results in the Online Appendix.

fewer than 10 conference calls.[14] We also remove high-frequency words (i.e., stop words), such as “and” and “the,” and remove all words containing digits. Similar to Huang et al. (2018), we convert high-frequency phrases that refer to specific financial or technical terms into single words to better represent the meaning of common financial and technical terms. For example, we convert “earnings per share” into “earnings-per-share” and “balance sheet” into “balance-sheet.”

Each supervised machine-learning technique likely captures a different aspect of the human interpretation process and is likely to be uniquely useful in capturing the narrative content of a disclosure. SVR is more likely to identify the importance of individual words and phrases, RF is more likely to identify the importance of interactions among words and phrases, and sLDA is more likely to identify the importance of groups of words and phrases when explaining a dependent variable. We briefly describe each method below.

Similar to ordinary least squares estimation, SVR places weights on individual words and phrases to explain a dependent variable (i.e., CDS spreads). These weights are determined through an algorithm that simultaneously minimizes the coefficient vector magnitude as well as the estimation errors. Frankel et al. (2016) and Manela and Moreira (2017) discuss SVR estimation in depth. RF creates decision trees to explain a specific value or characteristic, using an iterative process called binary recursive partitioning. Disclosures are binarily partitioned based on phrases using an algorithm that minimizes the sum of squared error in the dependent variable within the resulting partitions (Breiman 2001). The disclosures are partitioned until the number of disclosures in each partition falls below a pre-specified number or when the sum-of-squared error in each partition is equal to zero. sLDA chooses latent topics that are associated with a dependent variable by grouping phrases based on the probability of the phrases co-occurring within disclosures (Blei and McAuliffe 2007).[15]

We use two different methods to estimate the machine-learning models. In our first method, similar to that of Manela and Moreira (2017), we partition our sample into three groups: 1) an early subsample of observations with CDS spreads between 2002 and 2012 to train the supervised machine-learning models to identify the language associated with credit risk, 2) a hold-out subsample of observations between 2013 and 2016 *with* CDS to test whether the out-of-sample credit risk estimates capture credit risk as measured by CDS spreads, and 3) out-of-sample observations between 2013 and 2016 *without* CDS spreads to assess whether the credit risk estimates capture credit risk

[14] We estimate additional robustness tests to ensure our results are not sensitive to the requirement of requiring phrases (NGRAMS) to be included in at least 10 conference calls. Specifically, we re-estimate *TCR Score* for the conference call and MD&A by requiring the NGRAM to appear in at least (1) five disclosures or (2) 20 disclosures. We find qualitatively similar results to those reported in the manuscript. We report these results in the Online Appendix.

[15] Supervised LDA differs from unsupervised LDA. The former identifies topics in relation to a dependent variable and is better suited for prediction. The latter identifies topics *without* respect to a dependent variable and is better for text categorization. See Blei and McAuliffe (2007) for more information on the differences.

for firms out of sample.[16] We train the machine-learning models using 14,618 firm-quarters (6290 firm-years) with conference calls (MD&As) and CDS spreads between 2002 and 2012. We label this estimation method the "split" method. Our second method resembles that of Frankel et al. (2016). We estimate the machine-learning models for each year using historical data over years *t-4* to *t-1* for observations with CDS spreads. We then apply the relevant parameters from the machine-learning models to all disclosures (MD&As or conference calls) in the current year to obtain an estimation of credit risk for firms *with* or *without* CDS spreads. We label this estimation method the "rolling" method.

Each estimation (i.e., split and rolling) method has benefits and drawbacks. The split sample method has three possible limitations. First, the relatively low number of out-of-sample observations from 2013 to 2016 could reduce the power of our tests. Second, because we use the early observations of our sample to train the models, we cannot produce out-of-sample estimates of credit risk until after 2013. Third, the split method is more likely to be affected by stationarity assumptions, which assume that the words used to explain credit risk remain constant over time. Given that the determinants of credit risk could change over time, the split method may result in stale models and reduce explanatory power in our out-of-sample tests. One benefit of the split sample method is that its larger training sample size increases its ability to extract meaningful signals of credit risk from firm disclosures. The rolling method addresses a few of these limitations while presenting limitations of its own. For example, the rolling method generates out-of-sample estimates in each year of our sample, thus increasing power in our out-of-sample tests. To achieve a larger out-of-sample set, however, the rolling method uses fewer observations in each training sample, potentially reducing the power of the models to detect credit risk. This concern is more likely to be problematic in the early years of our sample when the CDS market is relatively small. An additional benefit of the rolling method is that it updates the relevant words that explain credit risk over time and is less affected by stationarity assumptions.[17,18]

[16] We recognize that the initiation of CDS trading may change firms' disclosures, although the literature provides mixed evidence (Martin and Roychowdhury 2015; Kim et al. 2018; Kang et al. 2020). We are not aware of any evidence that firms alter the information provided to market participants in their conference calls post-CDS trade initiation. Importantly, if the disclosures of non-CDS firms differ significantly from those of CDS firms, the ability of the *TCR Score* measures to capture credit risk among firms without CDS spreads would be diminished, reducing the power of our tests performed on non-CDS firms.

[17] Machine-learning methods reduce the effects of stationarity assumptions when compared to other textual analysis methods. For example, dictionary approaches, which are often constructed using simple word counts from researcher-generated word lists, are also subject to stationarity assumptions. To reduce the effects of stationarity for dictionary methods, the researcher would need to carefully evaluate disclosures to identify words that should no longer be included in the list and words that should be added. To avoid potential researcher bias, multiple researchers would be required to undertake this process to root out idiosyncratic biases. When dealing with a construct like credit risk, this process could involve a significant amount of time and potentially reduce the timeliness of the dictionary. In contrast, ML methods are only constrained by computer processing resources. In addition, machine-learning methods are not inherently subject to researcher biases that could influence the choice of words associated with credit risk.

[18] Although unlikely, managers could identify the words and weightings that predict credit risk using our methods with the intent to avoid their use during conference calls, 10-Ks, and other disclosures. We expect this to be less possible during conferences calls than with carefully vetted 10-K reports, given that managers do not control analysts' questions during the Q&A session of the conference call. For example, managers might find it difficult to stop analysts from asking about performance and debt structure, which are two key topics that are associated with credit risk (see the Online Appendix).

After estimating each of the three machine-learning models using the split and rolling estimation methods for firm observations with CDS spreads, we generate out-of-sample text-based estimates of the CDS spread for observations *with* CDS and *without* CDS. We combine the three estimates from each machine-learning model for each disclosure (conference call or MD&A) and estimation (split or rolling) method using factor analysis including observations over the preceding 365 days. Our estimation of credit risk using the conference call (MD&A) and the split estimation method is labeled *TCR Score – Split* (*TCR Score_MD&A – Split*). Our estimation of credit risk using the conference call (MD&A) and the rolling estimation method is labeled *TCR Score – Rolling* (*TCR Score_MD&A – Rolling*).[19] (TCR stands for text-based credit risk.)

In Fig. 1, we provide descriptive evidence examining whether our measures of credit risk vary as expected across firms and over time. For the sake of parsimony, we present evidence in Figs. 1 and 2 only for *TCR Score–Rolling*, although results are qualitatively similar for *TCR Score_MD&A – Rolling*. We split all observations with CDS spreads (*CDS*) into quintiles and estimate the average *TCR Score–Rolling* by year. The results in Fig. 1 suggest two key takeaways. First, the average *TCR Score–Rolling* varies monotonically by *CDS* quintile. For example, the average *TCR Score–Rolling* for the third *CDS* quintile is always between the average *TCR Score–Rolling* for the second and third *CDS* quintile over the entire sample period. Second, for each quintile in the sample, the average *TCR Score–Rolling* spikes for each *CDS* quintile around the financial crisis, which suggests that our measure captures variation in credit risk across time. In Fig. 2, we perform a similar analysis by estimating the average *TCR Score–Rolling* for each quintile ranking of S&P credit ratings. The AAA (D) rating takes the value of 1 (22) so that higher quintiles reflect higher credit risk, consistent with CDS spreads in Fig. 1. The inferences from Fig. 2 resemble those in Fig. 1.

3.3 Influential words and phrases

A primary benefit of using machine-learning methods to extract information from disclosures is their reduced reliance on subjectivity, relative to other textual analysis methods (e.g., dictionaries). However, a drawback is their elimination of intuition when selecting the phrases useful in assessing firm credit risk. To gain some understanding of the type of information used by the machine-learning methods, we provide descriptive evidence regarding the most influential phrases identified by each machine-learning technique. For the sake of parsimony, we focus our analysis on the split estimation method using conference calls (*TCR Score - Split*). We report similar analyses for the rolling conference call approach and the two MD&A approaches in the Online Appendix. Even though machine-learning methods do not use human intuition to select relevant credit risk phrases, we can draw a link between the construct of credit risk and our proxy by connecting the language identified by the machine-learning methods to economic intuition.

[19] To decrease the costs of replication, we provide *TCR Score* through our websites. If future researchers would like to recalculate our measure or apply these methods to another variable, they can access SVR, sLDA, and random forest regression trees using the following links: http://svmlight.joachims.org/, https://pypi.org/project/slda/, http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

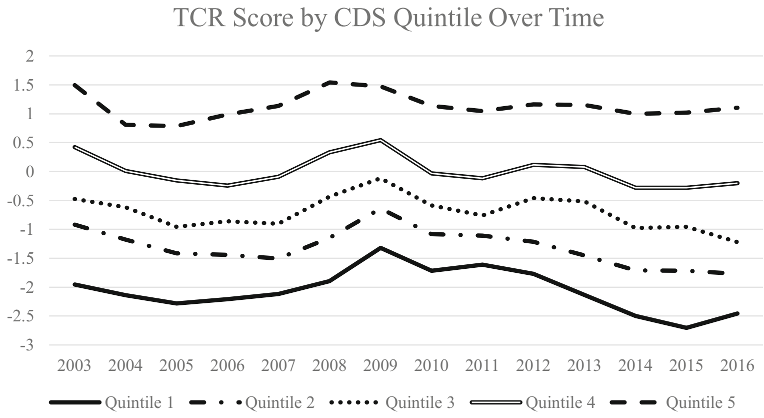

Fig. 1 *TCR Score-Rolling* by CDS Quintile Over Time

The random forest and support vector regression methods provide an importance weighting for each word and phrase used in model training. The supervised LDA method provides an importance weighting for each word and phrase within each topic. We manually categorize the top 200 most important words and phrases generated by the RF and SVR methods and the top 10 most important words and phrases for each of the 200 topics generated by sLDA into one of 13 categories that we identify during coding (performance, investing, banking, debt, liquidity, macro-economy, industry, direction of the news, risk, accounting, comparison, forward-looking, and tone). Any words or phrases that do not fit into one of these categories are left uncategorized.[20] For sLDA, we represent each topic as a weighted average of the 13 categories, where the category weights are determined by the relative importance of the assigned categories of the top 10 words and phrases for each topic. In an Online Appendix, we report the categorization of the words, phrases, and topics for each method (RF, SVR, and sLDA).

In Figs. 3, 4, and 5, we show the relative importance of the words, phrases, and topics in each of the categories when the models are estimated using RF, SVR, and sLDA, respectively. Figure 3 suggests that RF primarily identifies language associated with liquidity, firm performance, and debt. Figure 4 suggests that SVR primarily identifies words associated with the firm's performance, its industry, and accounting matters. Figure 5 presents the sum of the importance weightings for each category across all 200 sLDA topics. The results suggest that sLDA primarily identifies information related to performance, industry, and accounting. The Online Appendix provides the words lists and their categorizations for each model. Of note, we find that the two most negative sLDA topics (i.e., the topics most associated with a reduction in credit risk) discuss share repurchases and strong growth and the two most positive topics (i.e., the topics most associated with an increase in credit risk) discuss a challenging economic environment and liquidity. Overall, this evidence suggests that firm performance, liquidity, debt, and industry-specific information are among the most important categories identified by machine-learning methods to create the credit risk

[20] The words and phrases we can categorize represent 89.02% (68.61%) of the total importance of the top 200 most important words and phrases identified by the random forest (SVR) model.

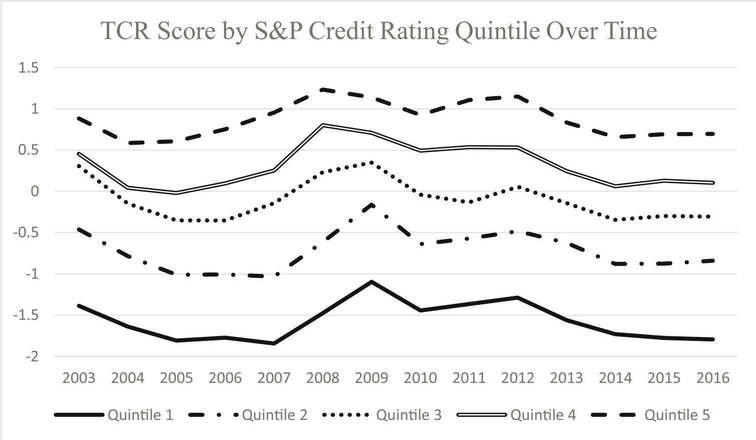

Fig. 2 *TCR Score-Rolling* by S&P Credit Rating Quintile Over Time

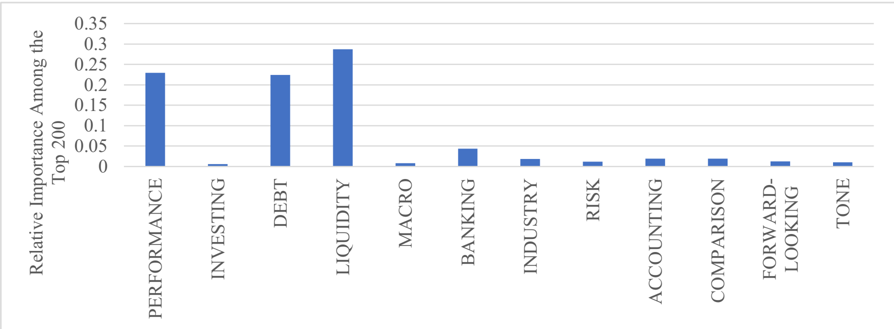

Fig. 3 Phrase Categorization for Random Forest (*TCR Score–Split*)

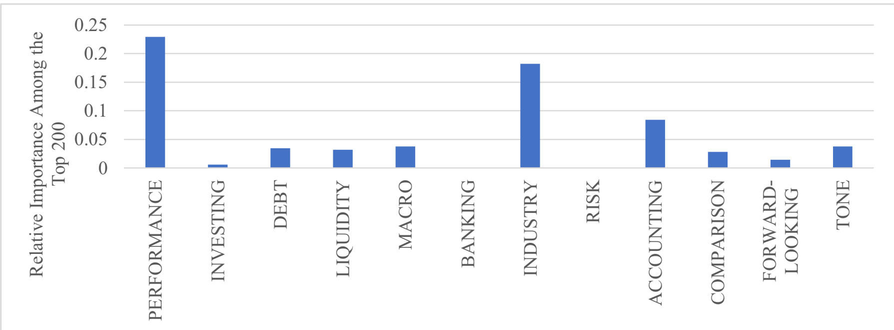

Fig. 4 Phrase Categorization for Support Vector Regression (*TCR Score–Split*)

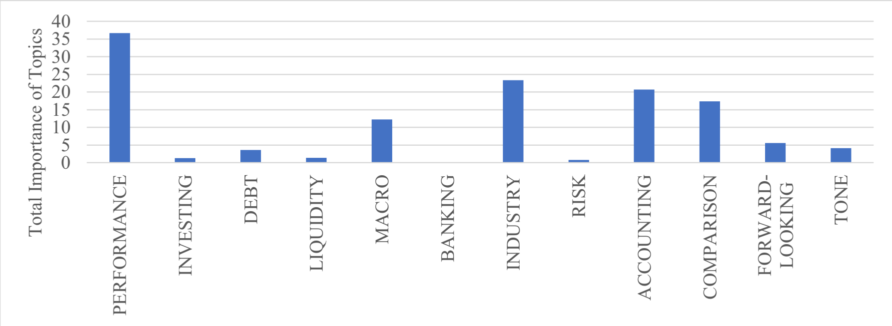

Fig. 5 Topic Categorization for Supervised Latent Dirichlet Allocation (*TCR Score–Split*)

measure. This analysis illustrates one of the key advantages of using machine-learning methods. Machine-learning methods identify a wider range of phrases and topics that may not be identified by a researcher.

We recognize that some of the phrases do not appear to directly relate to credit risk; however, they may indirectly relate. If these seemingly less intuitive words and phrases are not associated with the firm's credit risk, this would lower the effectiveness of the machine-learning methods in capturing credit risk and the estimates would have no out-of-sample explanatory power. Rather than using researcher judgment to determine the inclusion or exclusion of each word or phrase identified by each machine-learning method, we validate the machine-learning estimates of credit risk by assessing their ability to explain CDS spreads and other future credit events in out-of-sample tests. See Sections 4 and 5.

4 Validation: text-based measure of credit risk and CDS spreads

We validate our four *TCR Score* estimates of credit risk by examining their associations with CDS spreads in out-of-sample tests. We note that the sample size varies based on data availability for each method (split and rolling) and disclosure (conference call and MD&A).

Panel A of Table 1 presents descriptive statistics for the variables used in our analyses of CDS spreads. All variable definitions are included in Appendix Table 10. The average (median) CDS spread for observations in our sample is approximately 166.59 (77.57) basis points and is consistent with prior literature (Ham and Koharki 2016; Zhang et al. 2009). For this subsample of firms with CDS, the average (median) value of our text-based measure of credit risk ranges between −0.579 and − 0.217 (−0.666 and − 0.347).[21]

In an untabulated analysis, we find that firms in our CDS sample are large, with average (median) market value of equity plus book value of debt (*Size*) equal to \$22.74 billion (\$10.66 billion). The average (median) firm is profitable, with return on assets (*ROA*) of 0.014 (0.015). The average (median) firm has positive news during the fiscal

[21] We estimate *TCR Score* using factor analysis based on all conference call transcripts available over the past 365 days. In an untabulated analysis, we note that factor analysis generates a factor with a mean equal to zero and a standard deviation equal to one.

Table 1 CDS spreads and TCR score

Panel A: Descriptive Statistics						
Variable	**N**	**Mean**	**25th Pctl**	**Median**	**75th Pctl**	**Std Dev**
Dependent Variables:						
$CDS_{i,t}$	11,146	166.6	41.0	77.6	176.3	381.2
Narrative Disclosure Credit Risk Variables:						
TCR Score$_{i,t}$ – *Split*	2323	−0.492	−1.379	−0.570	0.267	1.207
TCR Score$_{i,t}$ – *Rolling*	11,146	−0.579	−1.516	−0.666	0.265	1.374
TCR Score_MD&A$_{i,t}$ – *Split*	596	−0.217	−1.200	−0.347	0.481	1.260
TCR Score_MD&A$_{i,t}$ – *Rolling*	3025	−0.314	−1.368	−0.406	0.692	1.519

Panel B: Explanatory Power of TCR Score												
	[1]			**[2]**			**[3]**			**[4]**		
TCR Score Measure	**Split Sample**			**Rolling**			**Split Sample_MD&A**			**Rolling_MD&A**		
Dependent Variable	***Log(CDS$_{i,t}$)***			***Log(CDS$_{i,t}$)***			***Log(CDS$_{i,t}$)***			***Log(CDS$_{i,t}$)***		
	Coefficient		**t-stat**	**Coefficient**		**t-stat**	**Coefficient**		**t-stat**	**Coefficient**		**t-stat**
TCR Score$_{i,t}$	0.585	***	(22.74)	0.614	***	(46.59)	0.561	***	(18.02)	0.554	***	(51.81)
Num Obs	2323			11,146			596			3025		
Adjusted R-Squared	0.628			0.655			0.620			0.614		

Table 1 examines the relation between CDS spreads and predicted credit risk based on narrative disclosures in conference calls (*TCR Score*). Panel A presents descriptive statistics for all variables used in the empirical analysis. Panel B reports the results of a regression model. The dependent variable, *Log(CDS)*, is equal to the natural log of one plus the average daily five-year CDS spread for the firm in the [−91,+1] window surrounding the conference call / MD&A filing date. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively quarter, with fiscal-quarter cumulative abnormal returns (*Return*) equal to 0.012 (0.010). Furthermore, the average (median) firm has total debt equal to 27.1% (24.8%) of total assets (*Leverage*).

To validate our measures, we estimate ordinary least squares (OLS) models with the natural log of one plus the firm's CDS spread (*Log(CDS Spread)*) as the dependent variable and each *TCR Score* measure as the sole independent variable. We report the results in Panel B of Table 1. Columns 1 and 2 (3 and 4) present the results with *TCR Score–Split* and *TCR Score–Rolling* (*TCR Score_MDA–Split* and *TCR Score_MDA–Rolling*) as the independent variables. We find that each of the coefficients on the *TCR Score* variables are positive and significant at the 1% level and the adjusted R^2s range from 61.4 to 65.5%, suggesting that both the conference call and MD&A capture credit risk as measured by CDS spreads.

5 Empirical results: text-based measure of credit risk and future credit events

In this section, we examine whether the *TCR Score* measures can be broadly applied to capture the credit risk of firms *without* CDS spreads. We do this for two reasons. First, if the *TCR Score* measures capture credit risk, they should be associated with *future* events that reflect a firm's credit risk. Second, as previously discussed, CDS spreads are not available for all borrowers (e.g., less than 10% of borrowers with long-term debt have CDS spreads). Since the *TCR Score* measures may only capture the credit risk of borrowers with CDS spreads, we examine whether the *TCR Score* measures capture credit risk for firms *without* CDS spreads.

We first examine the association between the *TCR Score* measures and bankruptcies occurring during the year after the conference call or MD&A, because debt holders and other stakeholders are interested in the probability of future insolvency (*Bankrupt*). We then examine whether the *TCR Score* measures capture the expected credit risk among firms obtaining new private debt using interest spreads in private debt contracts initiated during the period after the conference call or MD&A (*Interest Spread*). The interest rate spread represents a continuous measure of the borrower's credit risk based on information available to the bank lender during contract negotiations (Bharath et al. 2008; Plumlee et al. 2015). Therefore these tests examine how well the *TCR Score* measures capture the information used by banks when assessing the borrower's credit risk prior to debt contract inception. Lastly, we examine whether *TCR Score* predicts credit rating downgrades over the year after the conference call or MD&A (*Downgrade*). Each test is limited to a sample of firms *without* CDS spreads and with data available to measure the dependent and independent variables.

5.1 Descriptive statistics

Panel A of Table 2 presents descriptive statistics for all observations with available data for each test. For firms accessing the private debt market, the average (median) cost of debt (*Interest Spread*) for private loans issued during the quarter after the conference call is approximately 221 (175) basis points above LIBOR. Additionally, 0.3% of our sample files for bankruptcy within the next year (*Bankrupt*), consistent with prior

Table 2 Descriptive statistics

Panel A: Descriptive Statistics						
Variable	**N**	**Mean**	**25th Pctl**	**Median**	**75th Pctl**	**Std Dev**
Dependent Variables:						
$Bankrupt_{i,t+1}$	72,674	0.003	0.000	0.000	0.000	0.057
$Interest\ Spread_{i,t+1}$	6210	220.60	125.00	175.00	275.00	141.56
$Downgrade_{i,t+1}$	21,234	0.125	0.000	0.000	0.000	0.331
Narrative Disclosure Credit Risk Variables:						
$TCR\ Score_{i,t}$ – *Split*	23,292	−0.055	−0.687	−0.129	0.505	0.965
$TCR\ Score_{i,t}$ – *Rolling*	72,674	−0.024	−0.594	−0.060	0.519	0.889
$TCR\ Score\_MD\&A_{i,t}$ – *Split*	9660	0.148	−0.480	0.053	0.673	0.896
$TCR\ Score\_MD\&A_{i,t}$ – *Rolling*	30,947	0.047	−0.546	−0.036	0.567	0.883
Control Variables:						
$ZSCORE_{i,t}$	72,674	69.329	1.071	2.676	7.548	357.219
$OSCORE_{i,t}$	72,674	−1.689	−3.132	−1.925	−0.532	2.265
$Est\ Credit\ Rating_{i,t}$	72,674	13.434	12.000	13.000	15.000	2.571
$EDF_{i,t}$	72,674	0.013	0.000	0.000	0.001	0.054
$Size_{i,t}$	72,674	2425.020	205.419	647.232	1945.270	7531.690
$Leverage_{i,t}$	72,674	0.198	0.004	0.149	0.312	0.210
$ROA_{i,t}$	72,674	−0.002	−0.006	0.010	0.022	0.056
$MTB_{i,t}$	72,674	3.097	1.356	2.234	3.755	5.289
$Sales\ Growth_{i,t}$	72,674	0.138	−0.037	0.074	0.213	0.434
$Returns_{i,t}$	72,674	0.009	−0.105	0.008	0.120	0.214
$Std(Returns)_{i,t}$	72,674	0.149	0.100	0.135	0.181	0.068
$Std(Income)_{i,t}$	72,674	0.040	0.009	0.020	0.044	0.058

Table 2 (continued)

Std(Revenue)$_{i,t}$	72,674		0.059		0.024	0.041		0.071		0.061	
Analyst Following$_{i,t}$	72,674		2.972		1.000	1.000		4.000		3.495	
Tone$_{i,t}$	72,674		0.198		0.066	0.209		0.341		0.199	
Panel B: Correlation Matrix											
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[9]**	**[10]**	**[11]**
TCR Score$_{i,t}$ – Split [1]		0.830	0.581	0.560	0.539	0.100	0.249	−0.361	0.357	0.350	0.540
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
TCR Score$_{i,t}$ – Rolling [2]	0.841		0.514	0.626	0.522	0.082	0.171	−0.320	0.344	0.398	0.524
	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
TCR Score_MD&A$_{i,t}$ – Split [3]	0.610	0.561		0.798	0.519	0.099	0.126	−0.293	0.350	0.216	0.407
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
TCR Score_MD&A$_{i,t}$ – Rolling [4]	0.583	0.679	0.825		0.521	0.073	0.088	−0.275	0.342	0.268	0.537
	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Log (Spread$_{i,t+1}$) [5]	0.509	0.500	0.469	0.487		0.106	0.157	−0.414	0.373	0.465	0.530
	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Bankrupt$_{i,t+1}$ [6]	0.156	0.106	0.126	0.085	0.138		0.213	−0.071	0.074	0.042	0.081
	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001
Downgrade$_{i,t+1}$ [7]	0.280	0.177	0.126	0.084	0.157	0.213		−0.214	0.179	0.141	0.188
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
ZSCORE$_{i,t}$ [8]	−0.067	−0.033	−0.088	−0.053	−0.075	−0.011	−0.028		−0.343	−0.274	−0.328
	<.0001	<.0001	<.0001	<.0001	<.0001	0.00	<.0001		<.0001	<.0001	<.0001
OSCORE$_{i,t}$ [9]	0.330	0.310	0.321	0.307	0.339	0.080	0.179	−0.086		0.402	0.455
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
Est Credit Rating$_{i,t}$ [10]	0.357	0.415	0.227	0.288	0.461	0.040	0.149	0.048	0.428		0.446

Table 2 (continued)

	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001
$EDF_{i,t}$ [11]	0.254	0.254	0.183	0.283	0.264	0.127	0.113	−0.034	0.261	0.205	
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	

Table 2 presents descriptive statistics for all variables used in the empirical analysis. Panel A presents the distribution of dependent and independent variables. Panel B reports Pearson and Spearman correlations between key variables used in the analysis. Pearson correlations are presented below the diagonal, and Spearman correlations are presented above the diagonal. Correlation coefficients and p-values are presented for all correlations. All variables are defined in Appendix Table 10 literature suggesting that bankruptcy is rare. Approximately 12.5% of firm-quarters are downgraded by S&P (*Downgrade*) within the next year.

Descriptive statistics for the *TCR Score* variables, other credit risk proxies, and control variables are based on the sample of firms used in the bankruptcy sample, which is the largest subsample used in this section. The average (median) value of the *TCR Score* measures range between −0.055 and 0.148 (−0.129 and 0.053) and the standard deviation ranges between 0.883 and 0.965. Additionally, firms in this subsample have average (median) market value of equity plus book value of debt (*Size*) equal to \$2.43 billion (\$0.65 billion). The average (median) firm nearly breaks-even, with return on assets (*ROA*) of −0.002 (0.010). Furthermore, the average (median) firm has total debt equal to 19.8% (14.5%) of total assets (*Leverage*). Finally, the average and median firm is assigned a speculative-grade credit rating by S&P of BB- (*Est Credit Rating*).

We report univariate correlations in Panel B of Table 2. Pearson (Spearman) correlations are below (above) the diagonal. As expected, the *TCR Score* variables are positively associated with *Interest Spread*, *Bankrupt*, and *Downgrade.* These correlations provide preliminary evidence that the *TCR Score* variables capture both expected credit risk as well as credit risk outcomes. As expected, the *TCR Score* variables are positively (negatively) correlated with *OSCORE*, *Est Credit Rating*, and *EDF* (*ZSCORE*).[22]

5.2 Explanatory power of text-based measure of credit risk and other measures of credit risk

We next estimate a set of tests to examine the overall usefulness of each *TCR Score* variable as a summary measure of credit risk. We first estimate nonparametric tests to determine whether the *TCR Score* variables and other credit risk proxies help predict bankruptcy over the subsequent year (*Bankrupt*). In Table 3, we percentile rank (100 groups) all credit risk proxies (*TCR Score*, *ZSCORE*, *OSCORE*, *Est Credit Rating*, and *EDF*) and count the number of firms filing bankruptcy over the subsequent year within each percentile.[23] For a particular credit risk proxy, a perfect prediction would have all bankruptcies contained within the 99th percentile of the distribution. The results reported in Table 3 suggest that each of the credit risk measures reasonably predict bankruptcy, as indicated by the increasing number of firms entering default based on the percentile rank of each proxy. However, we note that *TCR Score–Split*, *TCR Score–Rolling*, and *TCR Score_MD&A–Split* better predict bankruptcy than the other measures. We note that 40.21%, 18.26%, and 17.07% of the total bankruptcies are found in the 99th percentile of the *TCR Score–Split*, *TCR Score–Rolling*, and *TCR*

[22] In an additional untabulated analysis for this sample, we find that the *TCR Score* variables are negatively correlated with firm performance (*ROA*), growth (*MTB* and *Sales Growth*), size (*Size*), stock returns (*Returns*), and analyst following (*Analyst Following*). We also note that *TCR Score* is positively correlated with firm leverage (*Leverage*) and performance volatility (*Std[Returns]*, *Std[Income]*, *Std[Revenue]*).

[23] For brevity, we combine percentile rankings by summing the number of observations and total future bankruptcies that fall in each group below the 96th percentile (e.g., the 80th through 89th percentiles are grouped together). Additionally, to facilitate comparison, we multiply *ZSCORE* by negative one for this analysis only so that greater values of *ZSCORE* indicate greater credit risk. In addition, *Est Credit Rating* consists of discrete categories for each S&P credit rating and therefore does not have a value for each percentile. For this reason, there are no observations that fall in the 50th–59th and 96th–97th percentiles.

Table 3 Nonparametric analysis of future default

[1]				[2]				[3]				[4]			
TCR Score – Split				*TCR Score – Rolling*				*TCR Score_MD&A – Split*				*TCR Score_MD&A – Rolling*			
Pct	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*
0–9	2329	0	0.00%	0–9	7267	0	0.00%	0–9	966	0	0.00%	0–9	3094	1	0.88%
10–19	2329	0	0.00%	10–19	7267	1	0.41%	10–19	966	0	0.00%	10–19	3095	0	0.00%
20–29	2329	0	0.00%	20–29	7268	0	0.00%	20–29	966	0	0.00%	20–29	3095	0	0.00%
30–39	2330	0	0.00%	30–39	7267	4	1.66%	30–39	966	0	0.00%	30–39	3095	3	2.63%
40–49	2329	1	1.03%	40–49	7268	1	0.41%	40–49	966	1	2.44%	40–49	3094	1	0.88%
50–59	2329	1	1.03%	50–59	7267	1	0.41%	50–59	966	0	0.00%	50–59	3095	5	4.39%
60–69	2330	1	1.03%	60–69	7268	7	2.90%	60–69	966	0	0.00%	60–69	3095	6	5.26%
70–79	2329	2	2.06%	70–79	7267	17	7.05%	70–79	966	4	9.76%	70–79	3095	12	10.53%
80–89	2329	10	10.31%	80–89	7268	37	15.35%	80–89	966	6	14.63%	80–89	3095	20	17.54%
90–95	1398	18	18.56%	90–95	4360	52	21.58%	90–95	580	15	36.59%	90–95	1857	27	23.68%
96	233	4	4.12%	96	727	18	7.47%	96	97	2	4.88%	96	309	7	6.14%
97	233	8	8.25%	97	727	26	10.79%	97	96	5	12.20%	97	310	12	10.53%
98	233	13	13.40%	98	727	33	13.69%	98	97	1	2.44%	98	309	7	6.14%
99	232	39	40.21%	99	726	44	18.26%	99	96	7	17.07%	99	309	13	11.40%
[1]				[2]				[3]				[4]			
$ZSCORE_{i,t}$				*$OSCORE_{i,t}$*				*Est Credit $Rating_{i,t}$*				*$EDF_{i,t}$*			
Pct	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*	**Pct**	**N**	***Bankrupt***	*%*
0–9	7267	0	0.00%	0–9	7267	4	1.66%	0–9	9284	2	0.83%	0–9	11,710	7	2.90%
10–19	7267	0	0.00%	10–19	7267	1	0.41%	10–19	8120	11	4.56%	10–19	2819	0	0.00%
20–29	7268	2	0.83%	20–29	7268	1	0.41%	30–39	10,066	14	5.81%	20–29	7273	0	0.00%
30–39	7267	3	1.24%	30–39	7267	1	0.41%	40–49	10,662	19	7.88%	30–39	7267	0	0.00%
40–49	7268	3	1.24%	40–49	7268	3	1.24%	60–69	11,114	36	14.94%	40–49	7268	0	0.00%

Table 3 (continued)

50–59	7267	4	1.66%	50–59	7267	5	2.07%	70–79	7273	45	18.67%	50–59	7267	1	0.41%
60–69	7268	10	4.15%	60–69	7268	4	1.66%	80–89	6436	49	20.33%	60–69	7268	5	2.07%
70–79	7267	19	7.88%	70–79	7267	21	8.71%	90–95	7751	54	22.41%	70–79	7267	13	5.39%
80–89	7268	102	42.32%	80–89	7268	53	21.99%	98	1470	9	3.73%	80–89	7268	29	12.03%
90–95	4360	82	34.02%	90–95	4360	96	39.83%	99	498	2	0.83%	90–95	4360	76	31.54%
96	727	6	2.49%	96	727	20	8.30%					96	727	36	14.94%
97	727	7	2.90%	97	727	12	4.98%					97	727	23	9.54%
98	711	2	0.83%	98	706	9	3.73%					98	727	21	8.71%
99	742	1	0.41%	99	747	11	4.56%					99	726	30	12.45%

Table 3 examines provides nonparametric analysis testing the ability of individual measures of credit risk to predict future default. We rank each measure of credit risk into 100 groups and provide the number of firms entering bankruptcy over the subsequent one-year period in each percentile. All variables are defined in Appendix Table 10

Score_MD&A–Split measures, respectively. *EDF* provides next best prediction by placing 12.45% of the bankruptcies in the 99th percentile. The *TCR Score_MD&A–Rolling* measure is just behind *EDF* by including 11.40% of the bankruptcies in the 99th percentile. This nonparametric analysis suggests that the *TCR Score* variables perform well in predicting future bankruptcy.

Next, we use regression analysis to compare the *individual* explanatory power of each *TCR Score* measure for predicting bankruptcy, interest rate spreads, and credit rating downgrades to that of each other credit risk measure used in the literature (*ZSCORE*, *OSCORE*, *Est Credit Rating*, and *EDF*) to assess the relative usefulness of each proxy. This evidence is important to future researchers evaluating the appropriateness of credit risk proxies in a particular context. We estimate the following model to estimate the usefulness of each credit risk proxy.

$$DEP\ VAR = \alpha + \beta 1\ Credit\ Risk\ Proxy + \varepsilon. \tag{1}$$

The *DEP VAR* is equal to each of the *future* events that reflect a firm's credit risk (*Bankrupt*, *Spread*, and *Downgrade*). A separate OLS or probit regression is estimated for each dependent variable and each credit risk proxy, which include each *TCR Score* variable (*TCR Score – Split*, *TCR Score – Rolling*, *TCR Score_MD&A – Split*, *TCR Score_MD&A – Rolling*) and other firm-specific measures of credit risk used in the literature (Altman Z-score [*ZSCORE*], O-score [*OSCORE*], expected default frequency [*EDF*], and the estimated credit rating [*Est Credit Rating*]).[24] The adjusted-R^2 or pseudo-R^2 in each regression assesses each credit risk proxy's ability to capture firm credit risk. The results are presented in Panel A of Table 4. In Column 1, we find evidence that all of the *TCR Score* variables yield a higher pseudo-R^2, compared to the other measures of credit risk, when *Bankrupt* is the dependent variable. *TCR Score – Split* yields the largest pseudo-R^2 equal to 34.2%. Among the other measures, the *OSCORE* yields the highest pseudo-R^2 equal to 10.9%. In untabulated results, we note similar inferences when evaluating each measure using the area under the ROC curve. Similarly, in Column 2, all the *TCR Score* variables yield a higher adjusted-R^2 than all other credit risk variables when explaining future interest spreads (*Spread*), with the highest adjusted-R^2, that is, 22.6% for *TCR Score – Split*. Of the other measures, the *Est Credit Rating* yields the highest adjusted-R^2, that is, 17.8%. In Column 3, when predicting future downgrades, *TCR Score – Split* and *TCR Score – Rolling* yield the highest pseudo-R^2s (9.5% and 4.0%, respectively), and *OSCORE* yields the highest pseudo-R^2 (3.9%) among the other measures.

[24] The parameters from Altman (1968) and Ohlson (1980) could be updated to provide a better credit risk measure. We re-estimated the parameters for the *ZSCORE* and *OSCORE* in case the parameter estimates have changed since the publication of Altman (1968) and Ohlson (1980). Specifically, we estimated rolling regressions, where the dependent variable is an indicator variable equal to one if the firm enters bankruptcy over the subsequent two years, and we included all of the financial ratios from Altman (1968) and Ohlson (1980) as independent variables. For each quarter in the sample, we estimated this regression using all available data (beginning in 1985) up to that point in time. In addition to re-estimating the coefficients from the models of Altman (1968) and Ohlson (1980), we also used rolling regressions to update the coefficient estimates for estimated credit rating (*Est Credit Rating*) from Barth et al. (2008) and Beatty et al. (2008). After replacing *ZSCORE*, *OSCORE*, and *Est Credit Rating* with the updated variables, all regression results reported in Tables 4, 5, 6, 7, 8, and 9 are qualitatively similar. These results can be found in the online appendix.

Collectively, the evidence in Panel A of Table 4 suggests that the *TCR Score* variables represent a more consistent *summary* measure of credit risk than other measures identified in the literature. By comparison, *OSCORE* is the next most consistent measure, as it ranks first, second, and first among the other measures across the three columns. For the sake of comparison, we compare our results to those of Hillegeist et al. (2004). Hillegeist et al. (2004) show that EDF is a more comprehensive measure than the Z-Score and O-Score. The authors provide evidence that EDF alone yields a 12% pseudo-R^2 while the O-Score, which is the next best measure in their paper, yields a 10% pseudo-R^2 when predicting bankruptcies. We find that *TCR* $Score_{i,t}$*–Split* yields a pseudo-R^2 of 34.2% while the O-score, which is the next best measure in our empirical tests, yields a pseudo-R^2 of 10.9% when predicting bankruptcies. Based on the above evidence, we believe that our findings are economically significant.

We next examine the incremental usefulness of each *TCR Score* variable in capturing credit risk, relative to that of the other credit risk proxies. To do so, we compare the explanatory power of the following two regression models for each dependent variable and each measure of credit risk previously discussed.

$$\begin{aligned} DEP\ VAR = {} & \alpha + \beta 1 Credit\ Risk\ Proxy + \beta 2 Other\ Credit\ Risk\ Proxies \\ & + \beta 3 Controls + \varepsilon. \end{aligned} \tag{2}$$

$$DEP\ VAR = \alpha + \gamma 1\ Other\ Credit\ Risk\ Proxies + \gamma 2 Controls + \varepsilon. \tag{3}$$

Control variables in Eqs. 2 and 3 include the firm's market value of equity plus book value of debt (*Size*), leverage (*Leverage*), profitability (*ROA*), growth (*MTB*, *Sales Growth*), information environment (*Analyst Following*), and economic performance during the fiscal quarter as reflected by abnormal stock returns (*Returns*). We also include control variables for the inherent volatility of the firm's operating environment using the historical standard deviation of the firm's stock returns [*Std(Returns)*], net income [*Std(Income)*], and sales [*Std(Revenue)*]. We note that these variables control for overall volatility, as examined by Bao and Datta (2014), Campbell et al. (2014), and Kravet and Muslu (2013).[25] Similar to *EDF*, we note that *Size*, *Return*, and *Std(Returns)* also represent market-based measures of credit risk (Beaver et al. 2006). Lastly, we control for the tone of the conference call (*Tone*) to reduce the likelihood that we are not simply measuring the tone of the disclosure with our measure of credit risk. All variable definitions can be found in Appendix Table 10. We winsorize all continuous variables at the first and 99th percentile values by year. In all reported regression models throughout the paper, we cluster standard errors by three-digit SIC.

Comparing the explanatory power of the full model using Eq. 2 to the model omitting a proxy of credit risk using Eq. 3 allows us to isolate the incremental

[25] Kravet and Muslu (2013) also examine analyst forecast dispersion. Although we do not include a specific control for analyst forecast dispersion in our primary tests due to data constraints, we explicitly control for the standard deviation of returns [*Std(Returns)*], earnings [*Std(Income)*], and sales [*Std(Revenue)*]. In additional robustness tests, we explicitly control for analyst forecast dispersion and find qualitatively similar results.

Table 4 Explanatory power of individual measures of credit risk

Panel A: Explanatory Power of Each Individual Measure									
	[1]			**[2]**			**[3]**		
	$Bankrupt_{i,t+1}$			***$Log(Spread_{i,t+1})$***			***$Downgrade_{i,t+1}$***		
	Coefficient	**t-stat**	**Pseudo R-Sq**	**Coeff-icient**	**t-stat**	**Adj. R-Sq**	**Coeff-icient**	**t-stat**	**Pseudo R-Sq**
$TCR\ Score_{i,t}$ – Split	0.734	(12.33)	0.342	0.246	(17.95)	0.226	0.380	(5.82)	0.095
$TCR\ Score_{i,t}$ – Rolling	0.646	(19.38)	0.220	0.285	(19.18)	0.208	0.261	(7.79)	0.040
$TCR\ Score\_MD\&A_{i,t}$ – Split	0.606	(13.14)	0.217	0.224	(12.28)	0.179	0.186	(4.86)	0.020
$TCR\ Score\_MD\&A_{i,t}$ – Rolling	0.486	(14.82)	0.146	0.266	(20.77)	0.199	0.133	(5.93)	0.011
$ZSCORE_{i,t}$	−0.0599	(−5.65)	0.067	−0.0002	(−5.31)	0.008	−0.001	(−1.49)	0.002
$OSCORE_{i,t}$	0.144	(10.03)	0.109	0.101	(14.19)	0.107	0.134	(13.07)	0.039
$Est\ Credit\ Rating_{i,t}$	0.093	(5.70)	0.037	0.124	(20.76)	0.178	0.106	(8.03)	0.028
$EDF_{i,t}$	3.162	(16.38)	0.102	3.071	(9.45)	0.060	2.386	(7.45)	0.013
Panel B: Incremental Explanatory Power of Each Individual Measure									
	[1]			**[2]**			**[3]**		
	$Bankrupt_{i,t+1}$			***$Log(Spread_{i,t+1})$***			***$Downgrade_{i,t+1}$***		
	Incremental Pseudo R-Squared			**Incremental Adj. R-Squared**			**Incremental Pseudo R-Squared**		
$TCR\ Score_{i,t}$ – Split	26.158%			2.759%			10.191%		
$TCR\ Score_{i,t}$ – Rolling	22.581%			4.175%			2.586%		
$TCR\ Score\_MD\&A_{i,t}$ – Split	27.986%			1.579%			0.000%		
$TCR\ Score\_MD\&A_{i,t}$ – Rolling	24.762%			3.376%			0.000%		
$ZSCORE_{i,t}$	0.588%			0.605%			0.000%		
$OSCORE_{i,t}$	0.885%			0.201%			4.386%		

Table 4 (continued)

Est Credit Rating$_{i,t}$	0.293%	1.012%	0.000%
EDF$_{i,t}$	3.636%	0.808%	0.847%

Table 4 examines the explanatory power of individual measures of credit risk in predicting future credit events. Panel A presents the individual explanatory power of each measure using only each individual measure of credit risk on the right-hand side of the regression model. Panel B presents the incremental explanatory power of each measure of credit risk, calculated by measuring the improvement in the explanatory power of (i) the regression model with all credit risk proxies and control variables, relative to (ii) the regression model omitting that particular measure of credit risk while controlling for all other measures of credit risk and control variables. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively usefulness of each individual credit risk proxy and compare it to other proxies. We report the incremental explanatory power of each individual credit risk proxy in Panel B of Table 4. We measure the incremental explanatory power by subtracting the adjusted-R^2 using Eq. 3 from the adjusted-R^2 using Eq. 2. We then divide the difference by the adjusted-R^2 from Eq. 3.[26]

In Column 1 of Panel B in Table 4, we find that all the *TCR Score* variables have the highest incremental explanatory power when *Bankrupt* is the dependent variable. In Columns 2, when *Log(Spread)* is the dependent variable, the incremental explanatory power of all *TCR Score* variables exceeds that of the other credit risk proxies. In Column 3, when *Downgrade* is the dependent variable, the incremental explanatory of *TCR Score–Split* yields the highest increase in explanatory power, with *OSCORE* and *TCR Score-Rolling* yielding the next highest.

Overall, the evidence in Table 4 suggest that the *TCR Score* measures are useful and consistent summary measures of the firm's credit risk. While the evidence presented in Panel A of Table 4 suggests that the *TCR Score* measures based on both the conference call and the MD&A explain across-firm variation in credit risk, the evidence in Panel B suggests that the *TCR Score* measures yield the highest incremental explanatory power in the regressions, relative to other credit risk proxies in the literature. We emphasize that the *TCR Score* variables are positive and significant in each regression, which suggests that they identify information that is not captured by other measures of credit risk, including *EDF*, which is a market-based measure of credit risk. This result suggests that the *TCR Score* measures capture variation in credit risk not already captured by the equity market. We further test this conjecture in Section 6 by examining whether our measures capture information impounded by future market participants.

5.3 Bankruptcy and text-based credit risk measures

The tests in the previous subsection suggest that the *TCR Score* measures capture *across-firm* variation in credit risk. We now examine whether the *TCR Score* measures capture *within-firm* variation in credit risk. To do so, we augment the regression model in Eq. 2 with firm fixed effects. We also include time fixed effects to control for macroeconomic factors.[27] In Table 5, we start by including *Bankrupt* as the dependent variable using a linear probability model to avoid an incidental parameters problem (Greene 2004; Allison 2009).[28] We find that both credit risk measures based on the conference call (*TCR Score–Split* and *TCR Score–Rolling*) are positively and significantly (5% level) associated with the likelihood of bankruptcy, which suggests that these measures identify information about credit risk that is not captured by other measures or other firm characteristics. Because we include firm fixed effects in the models, these results suggest that the conference call credit risk measures capture *within-firm* variation in credit risk. In contrast, the coefficient on *TCR*

[26] We follow the same calculation using the Pseudo-R^2 for nonlinear probit models.

[27] For quarterly (annual) tests, we include year-quarter (year) fixed effects.

[28] In an additional untabulated robustness test, we use a hazard model to examine future bankruptcy and find that the *TCR Score* measures based on the conference call significantly predict bankruptcy over the subsequent one and three years. We report these results in the Online Appendix.

Table 5 Bankruptcy and TCR score

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	$Bankrupt_{i,t+1}$			$Bankrupt_{i,t+1}$			$Bankrupt_{i,t+1}$			$Bankrupt_{i,t+1}$		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
$TCR\ Score_{i,t}$	0.006	**	(2.21)	0.004	***	(3.34)	0.000		(0.01)	0.002	*	(1.75)
$ZSCORE_{i,t}$	0.000	*	(1.98)	0.000	***	(2.81)	0.000	*	(1.89)	0.000	*	(1.98)
$OSCORE_{i,t}$	0.000		(−0.64)	0.000		(−0.67)	0.000		(−0.24)	0.000		(−1.15)
$Est\ Credit\ Rating_{i,t}$	−0.001		(−1.01)	0.000		(0.44)	0.002		(1.54)	0.002	***	(3.57)
$EDF_{i,t}$	0.216	**	(2.26)	0.121	***	(4.24)	0.143	*	(1.82)	0.038	**	(2.36)
$Size_{i,t}$	0.000		(−1.38)	0.000		(−1.62)	0.000		(−1.33)	0.000		(−1.53)
$Leverage_{i,t}$	0.046		(1.57)	0.031	**	(2.31)	0.050		(1.23)	0.022	*	(1.75)
$ROA_{i,t}$	−0.122		(−1.61)	−0.044	**	(−2.07)	0.000		(−0.62)	0.002		(1.10)
$MTB_{i,t}$	0.000		(1.42)	0.000		(0.41)	0.000		(−1.03)	0.000		(−0.12)
$Sales\ Growth_{i,t}$	−0.005		(−0.99)	−0.002	*	(−1.78)	−0.001		(−0.71)	0.000		(−1.02)
$Returns_{i,t}$	−0.012	**	(−2.19)	−0.009	***	(−3.86)	0.003		(0.53)	−0.007	***	(−2.88)
$Std(Returns)_{i,t}$	0.091	*	(1.73)	0.029	*	(1.75)	−0.005		(−0.09)	0.013		(1.19)
$Std(Income)_{i,t}$	0.045	**	(2.08)	−0.016		(−1.13)	0.002		(1.02)	0.000		(−0.61)
$Std(Revenue)_{i,t}$	−0.001		(−0.11)	−0.002		(−0.20)	−0.003		(−0.36)	−0.004		(−0.97)
$Analyst\ Following_{i,t}$	0.000		(−1.27)	0.000		(−0.95)	0.000		(−0.43)	0.000		(−0.76)
$Tone_{i,t}$	0.002		(0.48)	0.000		(−0.08)	−0.002		(−0.29)	0.001		(0.18)
Firm Fixed Effects	Yes			Yes			Yes			Yes		
Time Fixed Effects	Yes			Yes			Yes			Yes		
Num Obs	23,292			72,674			9660			30,947		
Adjusted R-Squared	0.362			0.245			0.152			0.246		

Table 5 examines the association between bankruptcy filings and predicted credit risk based on narrative disclosures (*TCR Score*). The dependent variable, *Bankrupt*, is an indicator variable equal to one if the firm files for bankruptcy over the one-year period and zero otherwise. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively

Score_MD&A–Split is insignificant, and the coefficient on *TCR Score_MD&A–Rolling* is positive and marginally significant at the 10% level. These results provide limited evidence that the MD&A captures *within-firm* variation in firm credit risk. Among the alternative measures, only *EDF* appears to consistently capture within-firm variation when *Bankrupt* is the dependent variable. Interestingly, the coefficient on *ZSCORE* is consistently positive, which is opposite to that expected.[29] Finally, we find that the tone of the conference call or MD&A (*Tone*) is not significantly associated with the likelihood of future bankruptcies, which provides no evidence that the overall negativity of the firm's qualitative disclosure is associated with future bankruptcies.

5.4 Cost of debt and text-based credit risk measures

In Table 6, we report the results of Eq. 2 with *Log(Spread)* as the dependent variable including firm and time fixed effects.[30] We find that the coefficient on *TCR Score – Split* (*TCR Score – Rolling*) is positive and significant at the 10% (1%) level, which suggests that these measures identify *within-firm* variation in credit risk that is not captured by the other credit risk measures and firm characteristics. A one standard deviation increase in *TCR Score – Split* (*TCR Score – Rolling*) increases the cost of debt in future private debt contracts by approximately 5.6% (9.2%), which represents approximately 9.8 (16.1) basis points based on the median interest spread in our sample. These results suggest that *TCR Score – Split* and *TCR Score – Rolling* explain an economically significant portion of the variation in *Log(Spread)*. Overall, these results are particularly important, as they show that our measure of credit risk (*TCR Score*) captures credit risk among creditworthy firms obtaining new debt from private lenders.

Interestingly, we find a negative and marginally significant coefficient on *TCR Score_MD&A–Split* and an insignificant coefficient on *TCR Score_MD&A–Rolling*. Consistent with the results in Table 5, these results do not provide evidence that the MD&A captures *within-firm* variation in credit risk as measured by interest rate spreads in private debt contracts. Additionally, none of the other proxies (*EDF*, *ZSCORE*, *OSCORE*, or *Est Credit Rating*) are consistently significant in the predicted direction.[31] These results provide no evidence that the other credit risk proxies consistently capture *within-firm* variation.

5.5 Credit rating downgrades and text-based credit risk measures

In Table 7, we report the results of Eq. 2 with *Downgrade* as the dependent variable including firm and time fixed effects.[32] Credit rating data availability limits the sample size for this test. As Table 2 indicates, approximately 12.5% of firms are downgraded by S&P (*Downgrade*) within the year following the conference call. Similar to Tables 5 and 6, we continue to find a positive and significant (1% level) coefficient on *TCR Score – Split* and *TCR Score – Rolling*, which suggests that the conference call captures *within-firm* variation in credit risk as measured by credit rating downgrades. The coefficient of 0.061 (0.032) on *TCR Score – Split* (*TCR Score – Rolling*) suggests that

[31] In additional robustness tests, we replace *Est Credit Rating* with the estimated credit rating measures developed by Alp (2013) and Baghai et al. (2014). We find qualitatively similar results to those reported in Tables 5, 6, 7, 8, 9. The correlation between *Est Credit Rating* and the measure produced by Alp (2013) is equal to 0.781, and the correlation between *Est Credit Rating* and the measure produced by Baghai et al. (2014) is equal to 0.827.

[32] In an untabulated analysis, we perform a nonparametric test examining the association between each credit risk proxy and future credit rating downgrades. Similar to results reported in Table 3, we find nonparametric evidence that *TCR Score* outperforms other credit risk proxies in predicting credit rating downgrades.

a one standard deviation increase is associated with an increase in the probability of a credit rating downgrade of approximately 48.8% (25.6%) relative to sample mean of 12.5% (Table 2).

Neither the coefficient on *TCR Score_MD&A – Split* nor the coefficient on *TCR Score_MD&A – Rolling* is significant in the predicted direction, which does not provide evidence that the MD&A predicts *within-firm* variation in credit risk. In addition, none of the other credit risk variables (*EDF*, *ZSCORE*, *OSCORE*, or *Est Credit Rating*) has coefficients that are consistently significant in the predicted direction. Further, we find that *Tone* is negatively associated with future credit rating downgrades, which suggests that the overall tone of the conference call or MD&A provides useful information in assessing the likelihood of future credit rating downgrades.

5.6 Summary of results

We highlight several inferences that are noteworthy based on the results reported in Tables 4, 5, 6, 7. First, the qualitative credit risk measures based on conference calls (*TCR Score – Split* and *TCR Score – Rolling*) consistently capture both *across-firm* (Tables 3 and 4) and *within-firm* (Tables 5, 6, and 7) variation in credit risk that is incremental to other measures. This suggests that the conference call is particularly useful for assessing credit risk. Second, the qualitative credit risk measures based on the MD&A (*TCR Score_MD&A – Split* and *TCR Score_MD&A – Rolling*) do not consistently capture *within-firm* variation in credit risk but do capture *across-firm* variation. These results suggest that the MD&A is limited to providing information useful in identifying cross-sectional variation in credit risk across firms. Third, the other measures of credit risk (*EDF*, *ZSCORE*, *OSCORE*, or *Est Credit Rating*) also do not consistently explain *within-firm* credit risk but do explain *across-firm* variation in credit risk. Fourth, *TCR Score* measures based on the conference call capture information that is not already captured by equity-based measures of credit risk (e.g., *EDF*), which suggests that *TCR Score* measures capture information impounded by future market participants. We more directly test whether the *TCR Score* measures capture information impounded by future market participants in the next section.

6 Future market reactions to credit risk measures

We estimate three additional tests to examine how market participants respond to future credit events. We specifically examine the relation between the *TCR Score* measures and future bond yields, future negative equity market shocks, and the stock market reaction to future credit rating downgrades. These tests provide evidence on whether market participants incorporate information about credit risk from corporate disclosures immediately or with delay.

6.1 Future bond yields

In Table 8, we examine the association between our measures and one-year-ahead bond yields, which are measured based on the daily trade summary file provided by TRACE. While all prior analyses were performed at the firm level, this analysis is performed at

Table 6 Cost of debt and TCR score

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
$TCR\ Score_{i,t}$	0.056	*	(1.93)	0.092	***	(4.16)	−0.081	*	(−1.89)	0.020		(1.27)
$ZSCORE_{i,t}$	0.000		(0.21)	0.000		(−0.52)	0.000		(−0.60)	0.000		(−0.63)
$OSCORE_{i,t}$	0.001		(0.13)	0.005		(1.08)	0.000		(0.12)	−0.001		(−0.44)
$Est\ Credit\ Rating_{i,t}$	−0.006		(−0.34)	−0.003		(−0.39)	0.018		(0.88)	0.005		(0.46)
$EDF_{i,t}$	1.088		(1.16)	0.023		(0.10)	2.428	***	(3.58)	−0.021		(−0.11)
$Size_{i,t}$	0.000	*	(−1.65)	−0.003	*	(−1.67)	−0.007	*	(−1.94)	−0.001		(−0.84)
$Leverage_{i,t}$	0.163		(0.83)	0.240	**	(2.14)	−0.029		(−0.15)	0.134		(1.35)
$ROA_{i,t}$	−0.495		(−0.61)	−0.583		(−1.41)	−0.096		(−0.35)	−0.286	**	(−2.01)
$MTB_{i,t}$	−0.003		(−1.11)	0.000		(0.12)	0.000		(0.48)	0.001		(0.62)
$Sales\ Growth_{i,t}$	−0.020		(−0.61)	0.012		(0.51)	0.000		(−0.01)	0.019		(0.87)
$Returns_{i,t}$	0.014		(0.17)	−0.055		(−1.45)	−0.021		(−0.17)	0.004		(0.08)
$Std(Returns)_{i,t}$	0.570		(1.07)	0.390		(1.54)	0.980	*	(1.93)	0.442		(1.59)
$Std(Income)_{i,t}$	0.212		(0.35)	0.016		(0.07)	−1.502		(−1.46)	−0.090		(−0.37)
$Std(Revenue)_{i,t}$	−0.818		(−1.08)	0.441		(1.43)	0.275		(0.40)	0.367		(1.27)
$Analyst\ Following_{i,t}$	−0.005		(−1.37)	0.002		(0.52)	0.021		(1.39)	−0.014		(−1.11)
$Tone_{i,t}$	−0.033		(−0.29)	−0.040		(−0.76)	−0.110		(−0.66)	−0.268	***	(−3.57)
$BS\ Covenant_{i,t+1}$	−0.308	*	(−1.92)	−0.070		(−1.58)	−0.330	**	(−2.17)	−0.052		(−1.24)

Table 6 (continued)

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$			$Log(Spread_{i,t+1})$		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
$IS\ Covenant_{i,t+1}$	−0.178	*	(−1.67)	−0.069		(−1.36)	−0.120		(−1.03)	−0.061		(−1.18)
$PP_{i,t+1}$	−0.088	***	(−2.80)	−0.105	***	(−5.57)	−0.106	***	(−3.11)	−0.112	***	(−6.15)
$Revolver_{i,t+1}$	−0.120	***	(−4.01)	−0.180	***	(−7.59)	−0.110	***	(−3.87)	−0.165	***	(−6.16)
$Collateral_{i,t+1}$	0.120	*	(1.77)	0.157	***	(4.69)	0.144	**	(2.13)	0.189	***	(5.79)
$Inst\ Tranche_{i,t+1}$	0.000		(−0.15)	−0.023		(−0.87)	−0.001		(−0.03)	−0.004		(−0.18)
$Capex\ Restrict_{i,t+1}$	0.018		(0.20)	0.116	***	(3.83)	−0.025		(−0.23)	0.119	***	(3.80)
$Sweep\ Cov_{i,t+1}$	0.251	***	(4.00)	0.208	***	(7.79)	0.259	***	(3.67)	0.228	***	(8.49)
$Dividend\ Restrict_{i,t+1}$	−0.134	**	(−2.20)	−0.098	***	(−3.36)	−0.153	**	(−2.24)	−0.109	***	(−4.13)
$Log(Debt\ Size_{i,t+1})$	−0.006		(−0.47)	−0.025	**	(−2.51)	0.001		(0.08)	−0.035	***	(−3.73)
$Log(Maturity_{i,t+1})$	0.052		(1.37)	0.012		(0.49)	0.048		(1.19)	0.021		(0.89)
Firm Fixed Effects	Yes			Yes			Yes			Yes		
Time Fixed Effects	Yes			Yes			Yes			Yes		
Num Obs	2169			6210			2148			6997		
Adjusted R-Squared	0.816			0.754			0.817			0.732		

Table 6 examines the association between the cost of debt and predicted credit risk based on narrative disclosures (*TCR Score*). The dependent variable, *Log(Spread)*, is equal to the natural log of one plus the interest spread of the debt contract. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively

Table 7 Credit rating downgrades and TCR score

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	$Downgrade_{i,t+1}$			$Downgrade_{i,t+1}$			$Downgrade_{i,t+1}$			$Downgrade_{i,t+1}$		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
$TCR\ Score_{i,t}$	0.061	***	(3.49)	0.032	***	(3.24)	0.024		(0.68)	−0.008		(−1.00)
$ZSCORE_{i,t}$	0.000		(1.02)	0.000		(1.38)	0.000	**	(−2.26)	0.000	*	(−1.85)
$OSCORE_{i,t}$	0.003		(1.49)	0.008	***	(3.05)	0.000		(0.10)	0.000		(1.08)
$Est\ Credit\ Rating_{i,t}$	0.006		(0.72)	0.002		(0.36)	0.014		(1.24)	0.008		(1.50)
$EDF_{i,t}$	0.000		(−0.01)	0.174	*	(1.98)	0.251		(0.46)	0.077		(0.73)
$Size_{i,t}$	0.000	*	(−1.64)	−0.001	*	(−1.66)	−0.003	*	(−1.78)	−0.001		(−1.01)
$Leverage_{i,t}$	−0.180	**	(−2.10)	0.094	*	(1.74)	−0.138		(−1.31)	0.080		(1.41)
$ROA_{i,t}$	−0.719	***	(−3.34)	−0.620	***	(−4.80)	−0.188	***	(−2.57)	−0.111	*	(−1.76)
$MTB_{i,t}$	−0.001		(−0.73)	−0.001		(−0.76)	−0.001		(−1.56)	0.000		(−1.24)
$Sales\ Growth_{i,t}$	−0.100	***	(−5.26)	−0.072	***	(−6.57)	−0.009		(−0.38)	−0.043	***	(−3.63)
$Returns_{i,t}$	−0.119	***	(−5.06)	−0.133		(−10.20)	−0.197	***	(−2.98)	−0.206	***	(−7.76)
$Std(Returns)_{i,t}$	−0.259		(−0.69)	−0.880	***	(−5.24)	−0.177		(−0.36)	−0.676	***	(−3.18)
$Std(Income)_{i,t}$	−0.440		(−0.98)	−0.719	***	(−3.39)	−0.509		(−1.33)	−0.604	**	(−2.49)
$Std(Revenue)_{i,t}$	0.257		(0.91)	−0.218	*	(−1.73)	0.074		(0.34)	−0.101		(−0.91)
$Analyst\ Following_{i,t}$	0.001		(1.61)	0.002	**	(2.04)	−0.005		(−0.55)	−0.004		(−0.78)
$Tone_{i,t}$	−0.089	**	(−2.33)	−0.178	***	(−7.18)	−0.001		(−0.01)	−0.146	***	(−3.26)
Firm Fixed Effects	Yes			Yes			Yes			Yes		

Table 7 (continued)

	[1]		[2]		[3]		[4]	
TCR Score Measure	Split Sample		Rolling		Split Sample_MD&A		Rolling_MD&A	
Dependent Variable	$Downgrade_{i,t+1}$		$Downgrade_{i,t+1}$		$Downgrade_{i,t+1}$		$Downgrade_{i,t+1}$	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Time Fixed Effects	Yes		Yes		Yes		Yes	
Num Obs	8641		21,234		2675		6739	
Adjusted R-Squared	0.422		0.293		0.219		0.165	

Table 7 examines the association between credit rating downgrades and predicted credit risk based on narrative disclosures (*TCR Score*). The dependent variable, *Downgrade*, is an indicator variable equal to one if the firm receives a credit rating downgrade from S&P over the subsequent one-year period and zero otherwise. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively

Table 8 Bond yields and TCR score

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	*Log(Bond Yield$_{i,t+1}$)*			*Log(Bond Yield$_{i,t+1}$)*			*Log(Bond Yield$_{i,t+1}$)*			*Log(Bond Yield$_{i,t+1}$)*		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
TCR Score$_{i,t}$	0.507	***	(3.79)	0.247	**	(2.36)	0.748	*	(1.83)	−0.298		(−1.16)
ZSCORE$_{i,t}$	−0.001		(−0.61)	0.000		(−0.07)	0.042		(1.09)	0.008	**	(2.18)
OSCORE$_{i,t}$	0.025		(0.67)	0.039		(1.02)	−0.017		(−1.15)	−0.014		(−1.10)
Est Credit Rating$_{i,t}$	−0.156		(−1.41)	−0.109		(−1.12)	−0.435	*	(−1.80)	−0.322	*	(−1.79)
EDF$_{i,t}$	−7.380		(−1.13)	−6.010		(−0.93)	−15.600		(−1.50)	−11.240		(−1.19)
Size$_{i,t}$	−0.008	***	(−2.86)	−0.012	***	(−3.18)	−0.030	**	(−2.44)	−0.039	***	(−3.11)
Leverage$_{i,t}$	1.264		(1.23)	1.725		(1.40)	−0.297		(−0.18)	−0.042		(−0.03)
ROA$_{i,t}$	−2.635		(−0.79)	−3.243		(−1.01)	−8.758	*	(−1.68)	−8.527	*	(−1.70)
MTB$_{i,t}$	0.000		(−0.01)	−0.001		(−0.12)	0.006		(0.63)	0.003		(0.37)
Sales Growth$_{i,t}$	−0.269		(−0.63)	−0.360		(−1.13)	−0.756		(−1.37)	−0.209		(−0.42)
Returns$_{i,t}$	−2.070	**	(−2.41)	−1.973	**	(−2.42)	−6.248	***	(−2.64)	−5.436	***	(−2.66)
Std(Returns)$_{i,t}$	−1.958		(−0.42)	−0.824		(−0.19)	−4.126		(−0.36)	−1.335		(−0.13)
Std(Income)$_{i,t}$	−20.83	*	(−1.91)	−21.47	*	(−1.81)	−39.54	**	(−2.20)	−38.12	**	(−2.17)
Std(Revenue)$_{i,t}$	−2.766		(−0.71)	−2.853		(−0.80)	−3.592		(−0.80)	−3.103		(−0.78)
Analyst Following$_{i,t}$	−0.017	***	(−2.75)	−0.016	**	(−2.21)	0.042		(0.39)	0.037		(0.35)
Tone$_{i,t}$	0.153		(0.41)	−0.001		(−0.00)	−0.014		(−0.01)	−0.765		(−0.50)
Firm Fixed Effects	Yes			Yes			Yes			Yes		

Table 8 (continued)

	[1]		[2]		[3]		[4]	
TCR Score Measure	Split Sample		Rolling		Split Sample_MD&A		Rolling_MD&A	
Dependent Variable	*Log(Bond Yield$_{i,t+1}$)*		*Log(Bond Yield$_{i,t+1}$)*		*Log(Bond Yield$_{i,t+1}$)*		*Log(Bond Yield$_{i,t+1}$)*	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Time Fixed Effects	Yes		Yes		Yes		Yes	
Num Obs	35,158		41,038		10,383		11,972	
Adjusted R-Squared	0.752		0.734		0.669		0.646	

Table 8 examines the association between bond yields and predicted credit risk based on narrative disclosures (*TCR Score*). The dependent variable, *Log(Bond Yield)*, is the natural log of one plus the firm's future bond yield one-year-ahead from TRACE. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively the individual bond level. Rather than including firm fixed effects, we include bond-level fixed effects to examine within-bond variation in bond yields and control for all bond-specific attributes.

Similar to the results in Tables 5, 6, and 7, we find a positive and significant (5% level or better) coefficient on *TCR Score – Split* and *TCR Score – Rolling*, which suggests that the qualitative information in the conference call helps predict changes in credit risk as measured by the natural log of one-year-ahead bond yields. These results suggest that a one standard deviation increase in *TCR Score – Split* (*TCR Score – Rolling*) is associated with a 12.3% (6.0%) increase in bond yields relative to the sample median. The coefficient on *TCR Score_MD&A – Split* is marginally significant at the 10% level, but that the coefficient on *TCR Score_MD&A – Rolling* is insignificant. These results suggest that the MD&A does not consistently provide qualitative information that helps predict bond yields. The other credit risk proxies (*EDF*, *ZSCORE*, *OSCORE*, or *Est Credit Rating*) do not consistently predict bond yields. These results provide additional evidence that our conference call-based credit risk measures capture information that is reflected in one-year-ahead bond yields.

6.2 Negative price shocks and conference call predictions

We next examine whether the *TCR Score* measures capture information that has not yet been identified/impounded in equity markets. We create an indicator variable, *Price Shock*, equal to one if the firm experiences a monthly return of less than −30% during any month over the subsequent year and zero otherwise (Zhang 2008). This measure reflects an extremely negative outcome for firms but one that is less severe and more commonly experienced than bankruptcy. Consistent with this argument, approximately 15.2% of firms in the sample experience a future negative shock, while only 0.3% of firms file for bankruptcy. We expect that the *TCR Score* variables positively predict a negative shock during the subsequent year if the equity markets do not completely impound the information captured by the *TCR Score* measures. We continue to include firm fixed effects.

In Table 9, we find that both of the conference call measures (*TCR Score – Split* and *TCR Score – Rolling*) are positive and significant (1% level) in Column 1 and 2. This evidence suggests that the narrative content from the conference call can be used to predict significant negative stock returns. Additionally, we find that neither of the MD&A measures are significant at conventional levels, which suggests that the MD&A is not useful in predicting negative stock returns. Similar to the preceding tests, we find that the other credit risk proxies do not consistently predict significant negative stock returns. These results provide additional evidence that the *TCR Score* measures based on conference calls capture information that is not fully impounded by the equity markets at the time the conference call is released.

6.3 Stock market reaction to future credit rating downgrades

Finally, building on the evidence that *TCR Score* is positively associated with future credit rating downgrades, we examine the association between the *TCR Score* measures and the stock market reaction to future credit rating downgrades. Using the date of the credit rating downgrade available from Capital IQ, in an untabulated analysis, we find that *TCR Score -Split* (*TCR Score -Rolling*) predicts credit rating downgrade announcement stock returns, which are two-day cumulative abnormal returns, at the 10% (1%) level. When we include the other credit risk measures with each conference call based *TCR Score* measure, the coefficient on *TCR Score-Rolling* continues to be negative and significant at the 1% level (untabulated) and the coefficient on *TCR Score-Split* does not. None of the other measures of credit risk predict credit rating downgrade announcement stock returns in the predicted direction. In additional untabulated analyses, we do not find evidence that the *TCR Score* measures based on the MD&A are significantly associated with credit rating downgrade returns.

We hesitate to draw too many conclusions from these tests, due to the small sample sizes. Our sample yields only 533 and 1325 firm-quarter observations with credit rating downgrades for the split and rolling conference call samples, respectively. Thus the split sample test may lack statistical power. Furthermore, the tests examining equity market price shocks in Table 9 suggest that our conference call-based credit risk measures help predict stock market responses to decreases in creditworthiness.

7 Additional analyses

7.1 Machine-learning versus dictionary

In additional robustness tests, we examine whether the *TCR Score* measures incrementally explain credit risk, relative to the measures based on dictionaries (e.g., Mayew et al. 2015; Sethuraman 2019). While we believe we significantly contribute to the literature without controlling for these dictionary-based measures, it is also important to test whether our empirical results are distinct from those of Mayew et al. (2015) and Sethuraman (2019).

First, we control for explicit disclosure of the entity's ability to continue as a going-concern (*GC*) using the dictionary provided by Mayew et al. (2015). Consistent with Mayew et al. (2015), we find that *GC* has incremental ability to predict bankruptcy; however, *GC* is not consistently positively associated with other important aspects of a firm's credit risk, as captured by the cost of debt, credit rating downgrades, bond yields, or future equity price shocks. Therefore it does not appear that *explicit* disclosure of the firm's ability to continue as a going-concern comprehensively explains credit risk. In contrast, our *TCR Score* measures computed using the conference call remain positively associated with these future credit events after controlling for *GC* disclsoure. We report these results in the Online Appendix.

Second, we control for the proportion of words in the disclosure that are credit-relevant words (*PROPCRD%*), as defined by Sethuraman (2019). Sethuraman (2019) does not examine the usefulness of this dictionary in predicting credit events but rather studies how firms change their disclosures in response to changes in the reputation of the credit rating agency. Nevertheless, we test whether our measures continue to explain an economically significant amount of the variation in credit risk after controlling for PROPCRD%. After including PROPCRD%, the results in Tables 5, 6, 7, 8, and

Table 9 Negative equity market price shocks and TCR score

	[1]			[2]			[3]			[4]		
TCR Score Measure	Split Sample			Rolling			Split Sample_MD&A			Rolling_MD&A		
Dependent Variable	*Price Shock*$_{i,t+1}$			*Price Shock*$_{i,t+1}$			*Price Shock*$_{i,t+1}$			*Price Shock*$_{i,t+1}$		
	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat	Coefficient		t-stat
TCR Score$_{i,t}$	0.027	***	(2.88)	0.016	***	(2.90)	0.024		(1.47)	0.003		(0.40)
ZSCORE$_{i,t}$	0.000		(−0.41)	0.000		(1.51)	0.000		(0.54)	0.000	*	(1.98)
OSCORE$_{i,t}$	0.002		(1.26)	0.004	***	(4.58)	0.000		(0.10)	0.000		(−0.77)
Est Credit Rating$_{i,t}$	−0.005		(−1.09)	−0.004	*	(−1.80)	−0.002		(−0.46)	0.007	**	(2.54)
EDF$_{i,t}$	−0.430	***	(−2.93)	−0.067		(−0.96)	−0.340	*	(−1.83)	−0.202	***	(−3.75)
Size$_{i,t}$	0.000		(−0.42)	0.001		(1.33)	0.000		(−0.33)	0.001	*	(1.84)
Leverage$_{i,t}$	0.114	***	(2.69)	0.145	***	(6.07)	0.071		(0.96)	0.107	***	(3.17)
ROA$_{i,t}$	−0.129		(−1.19)	−0.301	***	(−7.40)	0.015		(1.17)	0.021		(1.43)
MTB$_{i,t}$	0.000		(0.06)	0.001		(1.54)	0.000		(0.31)	0.000		(0.97)
Sales Growth$_{i,t}$	0.012		(1.21)	0.018	***	(2.81)	0.016	**	(2.25)	0.007	**	(2.10)
Returns$_{i,t}$	−0.003		(−0.17)	−0.002		(−0.35)	−0.014		(−0.29)	−0.010		(−0.56)
Std(Returns)$_{i,t}$	−0.747	***	(−3.46)	−0.307	**	(−2.54)	−0.877	***	(−4.00)	−0.168		(−1.32)
Std(Income)$_{i,t}$	−0.097		(−0.73)	0.069		(1.11)	−0.004		(−0.59)	−0.012		(−1.36)
Std(Revenue)$_{i,t}$	0.066		(0.42)	0.155	**	(2.07)	−0.032		(−0.22)	0.060		(1.03)
Analyst Following$_{i,t}$	−0.001	**	(−2.23)	0.001	***	(3.30)	−0.002		(−0.58)	0.005		(1.36)
Tone$_{i,t}$	−0.026	*	(−1.74)	−0.064	***	(−6.14)	0.028		(0.68)	0.023		(1.04)
Firm Fixed Effects	Yes			Yes			Yes			Yes		

Table 9 (continued)

	[1]		[2]		[3]		[4]	
TCR Score Measure	Split Sample		Rolling		Split Sample_MD&A		Rolling_MD&A	
Dependent Variable	*Price Shock$_{i,t+1}$*		*Price Shock$_{i,t+1}$*		*Price Shock$_{i,t+1}$*		*Price Shock$_{i,t+1}$*	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Time Fixed Effects	Yes		Yes		Yes		Yes	
Num Obs	23,292		72,674		9660		30,947	
Adjusted R-Squared	0.486		0.374		0.325		0.251	

Table 9 examines the association between bond yields and predicted credit risk based on narrative disclosures (*TCR Score*). The dependent variable, *Price Shock*, is an indicator variable equal to one if the firm experiences monthly returns <−30% (Zhang 2008) over the subsequent one-year period and zero otherwise. Standard errors are clustered by industry. All variables are defined in Appendix Table 10. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively

9 are qualitatively similar for the TCR Score measures. We do not find consistent evidence that PROPCRD% explains within-firm variation in credit risk as measured using the dependent variables in Tables 5, 6, 7, 8, and 9. These results are also reported in the Online Appendix.

7.2 Individual machine-learning methods

We separately examine the predictive ability of each of the three machine-learning methods (i.e., SVR, RF, and sLDA) in explaining credit risk. We have no theoretical prediction regarding which machine-learning method yields the highest explanatory power. Each method likely captures a different aspect of the human interpretation process. SVR is more likely to identify the importance of individual words and phrases. RF is more likely to identify the importance of interactions among words and phrases. And sLDA is more likely to identify the importance of groups of words and phrases when explaining a dependent variable. Nevertheless, we examine which one produces the measures with the most predictive ability. We focus our analysis on the individual machine-learning methods using the conference call since these measures consistently capture *within-firm* variation in credit risk across several measures.

We find that the individual machine-learning predictions do not predict credit risk as well as the combined measures (i.e., *TCR Score*). However, both RF and sLDA perform slightly better than SVR. We present these results in the Online Appendix. One possibility for why SVR performs worst is that effective communication requires an understanding of the context in which words and phrases are used (Liberti and Petersen 2019). RF and sLDA do this by taking into consideration the interaction of words and phrases (RF) and by identifying topics discussed in the disclosure (sLDA). Nevertheless, these results suggest that the combined credit risk measure (i.e., *TCR Score*) better predicts credit risk, compared to any measures based on the individual machine-learning models. Therefore we recommend that future researchers combine the estimates of several machine-learning methods, as each individual method is likely to only capture a portion of the disclosures' narrative content.

7.3 Unintuitive words and phrases

We re-estimate each of the *TCR Score* measures using the machine-learning models excluding the "uncategorized" phrases that we identify in our category classification reported in the Online Appendix. Generally, the results are qualitatively similar and are tabulated in the Online Appendix. The similar results are potentially driven by the fact that the machine-learning models continue to identify important phrases and simply replace the eliminated uncategorized phrases with new phrases that are similarly unintuitive. Our analysis suggests that an iterative process that continuously updates the models to exclude newly uncategorized phrases is unlikely to yield additional explanatory power. The inclusion of unintuitive phrases in model estimation is a limitation of machine-learning models and likely increases noise in the out-of-sample prediction. Their inclusion, however, does not bias the out-of-sample estimates of credit risk.

7.4 Footnotes in the 10-K

In the preceding analysis, we estimate *TCR Score_MD&A* based on the narrative content of the firm's MD&A in the Form 10-K. However, the language in the footnotes may provide information that is useful in estimating credit risk. Therefore we re-estimate the *TCR Score* measures using the language from the footnotes in the 10-K. We find limited and inconsistent evidence that the measures based on the footnotes in the 10-K (*TCR Score_FN*) can explain within-firm variation in credit risk. When combining the estimates from the MD&A with the estimates from the footnotes (*TCR Score_10K*), we continue to find limited and inconsistent evidence that *TCR Score_10K* explains within-firm variation in credit risk. We tabulate these results and discuss them in greater detail in the Online Appendix.

8 Conclusion

We test whether supervised machine-learning methods extract qualitative information from firm disclosures that explains borrower credit risk. We train support vector regressions, supervised LDA, and random forest regression trees to identify language that explains credit risk. We create a summary measure for the borrower's credit risk by combining the estimates generated by each of the three methods using factor analysis for conference calls and MD&As, separately. Using holdout samples, we verify that our credit risk measures explain a substantial portion of the borrower's credit risk as measured by CDS spreads.

In out-of-sample tests, we find that our text-based measures based on conference calls and the MD&A predict *across-firm* variation in interest rate spreads, credit rating downgrades, and bankruptcy filings. However, we *only* find that the credit risk measures based on conference calls predict *within-firm* variation in interest rate spreads, credit rating downgrades, and bankruptcy filings. We also find that the credit risk measures based on conference calls predict *within-firm* variation in future market outcomes, which we measure as future bond yields and significant negative future equity returns. However, we find no evidence that the credit risk measures based on the MD&A or other credit risk proxies in the literature consistently capture *within-firm* variation in credit risk or future market reactions.

One potential drawback to using machine learning methods is that these methods could identify phrases that relate statistically but not economically to credit risk. If so, noise could be introduced into our measure. To address this concern, we categorize the phrases and topics into credit risk related topics and use out-of-sample tests (e.g., bankruptcy) to reduce the likelihood that we identify a purely statistical relation between the phrases identified by the machine learnings methods and credit risk.

Our study adds to the growing body of research using machine-learning methods to gather information from conference calls and 10-Ks to explain key outcomes (e.g., accruals, future cash flows, and fraud). In addition, our study adds to a large body of research that examines other useful signals extracted from conference calls (e.g., vocal and video cues, tone, among others). While sophisticated intermediaries use both quantitative and qualitative information to assess firms' credit risk, academic research has tended to focus on using quantitative signals (e.g., return volatility) to measure credit risk. Our study helps fill this void by developing a comprehensive and continuous measure of credit risk from firm disclosures. We expect that practitioners and academics could use our measure to supplement other models to obtain a more comprehensive and independent estimate of credit risk. In addition, our measure can be applied to borrowers when other market-based measures of credit risk (e.g., credit ratings and CDS spreads) are not available. Our measure provides an economically significant improvement in predicting credit events. We also show that conference calls are particularly useful at predicting *across-firm* and *within-firm* variation in credit risk while the MD&A is only useful at predicting *across-firm* variation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11142-020-09575-4.

Acknowledgments The authors thank Sam Bonsall, Mark Bradshaw, Bill McDonald, Scott Richardson, Terry Shevlin, Lakshmanan Shivakumar, Irem Tuna, and workshop participants at Brigham Young University, the CUHK Accounting Research Conference, the Duke/UNC Fall Camp, the Frankfurt School of Finance and Management, London Business School, the Swiss Accounting Research Alpine Camp, the University of California - Irvine, the University of Notre Dame, and the University of Iowa for helpful comments. The authors thank the Mendoza College of Business, Olin Business School, Krannert School of Management, and Terry College of Business for financial support. Data is available upon request from the authors.

Appendix

Table 10 Variable definitions

Variable	Definition (quarterly tests are measured at the quarterly level and annual tests are at the annual level)
$Analyst\ Following_{i,t}$	Number analysts following firm *i* during the period.
$Bankrupt_{i,t+1}$	Indicator variable equal to one if firm *i* enters bankruptcy during the one-year period following the conference call or 10-K filing date.
$Bond\ Yield_{i,t+1}$	The bond yield for firm *i* measured one year after the conference call or 10-K filing date.
$BS\ Covenant_{i,t+1}$	Indicator variable equal to one if firm *i*'s debt contract includes a leverage ratio, debt-to-equity ratio, net worth, current ratio, or quick ratio covenant and zero otherwise.
$Capex\ Restrict_{i,t+1}$	Indicator variable equal to one if firm *i*'s debt contract includes a covenant restricting the level of capital expenditures and zero otherwise.
$CDS_{i,t}$	Average daily CDS spread for firm *i* in the [−91,+1] window, relative to the conference call or 10-K filing date in quarter *q*.
$Collateral_{i,t+1}$	Indicator variable equal to one if firm *i*'s debt contract is secured.
$Debt\ Size_{i,t+1}$	Face value (in millions) of the debt contract for firm *i*.
$Dividend\ Restrict_{i,t+1}$	Indicator variable equal to one if the debt contract includes a dividend restriction and zero otherwise.
$Downgrade_{i,t+1}$	Indicator variable equal to one if the firm receives a credit rating downgrade from S&P over the subsequent one-year period following the conference call or 10-K filing date.

Table 10 (continued)

$EDF_{i,t}$	Expected default frequency at quarter q, derived from Merton (1974). We calculate a firm's EDF derived from the Black and Scholes (1973) and Merton (1974) models and based on the empirical approach of Hillegeist et al. (2004). Specifically, we use Stata code developed from the SAS script provided in the appendix of Hillegeist et al. (2004) to compute the EDF measure.
$Est\ Credit\ Rating_{i,t}$	Estimated S&P credit rating at quarter q following Barth et al. (2008) and Beatty et al. (2008). We estimate a firm's credit rating by regressing S&P credit ratings on profitability (*ROA*), leverage (*Leverage*), the natural log of total assets (Compustat *atq*), and indicator variables for loss firms (Compustat $ibq < 0$), firms with subordinated debt (Compustat $dltsub > 0$), and dividend paying firms (Compustat $dvt > 0$) as well as year and industry (one-digit SIC) fixed effects in a pooled regression. We then multiply the coefficients from the regression by the variables to obtained fitted values for each firm, which represents estimated credit ratings.
$Inst\ Tranche_{i,t+1}$	Indicator variable equal to one if the debt contract has a Term Loan B or higher and zero otherwise.
$Interest\ Spread_{i,t+1}$	The interest spread on private debt initiated after the conference call or 10-K filing date.
$IS\ Covenant_{i,t+1}$	Indicator variable equal to one if the debt contract includes an interest coverage ratio, fixed charge, debt service, minimum EBITDA, or debt-to-earnings covenant and zero otherwise.
$Leverage_{i,t}$	Total debt scaled by total assets ($(dlcq + dlttq)/atq$).
$Log(Spread_{i,t+1})$	Natural log of one plus the interest spread on private debt initiated after the conference call or 10-K filing date.
$Maturity_{\boldsymbol{i,t}}$	Maturity (in months) of the debt contract (*Maturity*).
$MTB_{i,t}$	Market value of equity scaled by book value of equity ($(prccq * cshoq)/ceqq$).
$OSCORE_{i,t}$	O-Score measure of credit risk following Ohlson (1980).
$PP_{i,t+1}$	Indicator variable equal to one if the debt contract includes a performance pricing provision and zero otherwise.
$Price\ Shock_{i,t+1}$	Indicator variable equal to one if the firm experiences a negative shock to the firm's stock price over the subsequent year, defined as monthly stock returns $< -30\%$ following Zhang (2008) and zero otherwise.
$Returns_{i,t}$	Cumulative abnormal daily return for the period. Abnormal returns are calculated using the value-weighted market return.
$Revolver_{i,t+1}$	Indicator variable equal to one if the debt contract is a revolving credit facility and zero otherwise.
$ROA_{i,t}$	Income before extraordinary items scaled by total assets (*ibq*/*atq* for quarterly tests and *ib*/*at* for annual tests).
$Sales\ Growth_{i,t}$	Percentage change in revenue. For quarterly tests, we measure the percentage changes in revenue (*revtq*) from the same quarter, prior year. For annual tests, we measure the percentage changes in revenue (*revt*) from the prior year.
$Size_{i,t}$	Market value of equity plus the book value of debt.
$Std(Income)_{i,t}$	Standard deviation of quarterly income before extraordinary items measured over the previous five years.
$Std(Returns)_{i,t}$	Standard deviation of CRSP monthly stock returns measured over the previous five years.
$Std(Revenue)_{i,t}$	Standard deviation of total revenues measured over the previous five years.
$Sweep\ Cov_{i,t+1}$	Indicator variable equal to one if the debt contract includes an excess cash flow sweep, asset sales sweep, debt issuance sweep, equity issuance sweep, or insurance proceeds sweep and zero otherwise.

Table 10 (continued)

TCR Score$_{i,t}$ – *Rolling*	Text-based estimate of the average CDS spread during quarter *q* based on conference call narrative disclosures using the rolling estimation method.
TCR Score$_{i,t}$ – *Split*	Text-based estimate of the average CDS spread during quarter *q* based on conference call narrative disclosures using the split estimation method.
TCR Score_MD&A$_{i,t}$ – *Rolling*	Text-based estimate of the average CDS spread during year *t* based on the MD&A narrative disclosures using the rolling estimation method.
TCR Score_MD&A$_{i,t}$ – *Split*	Text-based estimate of the average CDS spread during year *t* based on the MD&A narrative disclosures using the split estimation method.
Tone$_{i,t}$	The difference between positive and negative word counts divided by the sum of positive and negative word counts using the Loughran and McDonald (2011) dictionaries.
Total Assets$_{i,t}$	Total assets (*atq*).
ZSCORE$_{i,t}$	Z-Score measure of bankruptcy risk following Altman (1968).

References

Agarwal, S., and R. Hauswald. 2010. Distance and private information in lending. *The Review of Financial Studies* 23 (7): 2757–2788.

Allison, P. 2009. *Fixed effect regression models*. Sage Publications.

Alp, A. 2013. Structural shifts in credit rating standards. *Journal of Finance* 68 (6): 2435–2470.

Altman, E. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23: 589–609.

Baghai, R., B. Becker, and A. Tamayo. 2014. Have rating agencies become more conservative? Implications for capital structure and debt pricing. *Journal of Finance* 69 (5): 1961–2005.

Bao, Y., and A. Datta. 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Management Science* 60 (6): 1371–1391.

Barron's. (2017). Artificial Intelligence, Explained. Available at:https://www.barrons.com/articles/sponsored/artificial-intelligence-explained-1508530169

Barth, M., L. Hodder, and S. Stubben. 2008. Fair value accounting for liabilities and own credit risk. *The Accounting Review* 83 (3): 629–664.

Beatty, A., J. Weber, and J. Yu. 2008. Conservatism and debt. *Journal of Accounting and Economics* 45 (2–3): 154–174.

Beaver, W., C. Shakespeare, and M. Soliman. 2006. Differential properties in the ratings of certified versus non-certified bond-rating agencies. *Journal of Accounting and Economics* 42 (3): 303–334.

Bharath, S., J. Sunder, and S. Sunder. 2008. Accounting quality and debt contracting. *The Accounting Review* 83 (1): 1–28.

Black, F., and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (3): 637–654.

Blanco, R., S. Brennan, and I. Marsh. 2005. An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *The Journal of Finance* 60 (5): 2255–2281.

Blei, D., and J. McAuliffe. 2007. Supervised topic models. *Advances in neural infomration processing systems* 20: 1–128.

Bonsall, S., E. Holzman, and B. Miller. 2016. Managerial ability and credit risk assessment. *Management Science* 63 (5): 1425–1449.

Bonsall, S., K. Koharki, and M. Neamtiu. 2017a. When do differences in credit rating methodologies matter? Evidence form high information uncertainty borrowers. *The Accounting Review* 92 (4): 53–79.

Bonsall, S., K. Koharki, and L. Watson. 2017b. Deciphering tax avoidance: Evidence from credit rating disagreements. *Contemporary Accounting Research* 43 (2): 818–848.

Bonsall, S., A. Leone, B. Miller, and K. Rennekamp. 2017c. A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63: 329–357.

Bozanic, Z., and P. Kraft. 2015. Qualitative corporate disclosures and credit analysts' soft rating adjustments. Working paper, Florida State University.
Bozanic, Z., D.T. Roulstone, and A. Van Buskirk. 2018. Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics* 65 (1): 1–20.
Breiman, L. 2001. Random forests. *Machine Learning* 45: 5–32.
Campbell, J., H. Chen, D. Dhaliwal, H. Lu, and L. Steele. 2014. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies* 19: 396–455.
Campbell, D., M. Loumioti, and R. Wittenberg-Moerman. 2019. Making sense of soft information: Interpretation bias and loan quality. *Journal of Accounting and Economics* 68 (2–3): 101240.
Cecchini, M., H. Aytug, G. Koehler, and P. Pathak. 2010. Detecting management fraud in public companies. *Management Science* 56 (7): 1146–1160.
Chava, S., and M. Roberts. 2008. How does financing impact investment? The role of debt covenants. *The Journal of Finance* 63 (5): 2085–2121.
Chen, S., B. Miao, and T. Shevlin. 2015. A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research* 53 (5): 1017–1054.
Christensen, T.E., Pei, H., Pierce, S.R. and Tan, L. 2019. Non-GAAP reporting following debt covenant violations. *Review of Accounting Studies* 24 (2): 629–664.
Costello, A., Down, A. & Mehta, M. (2019). Machine + man: A field experiment on the role of discretion in augmenting AI-based lending models. Working paper.
Das, S., M. Kalimipalli, and S. Nayak. 2014. Did CDS trading improve the market for corporate bonds? *Journal of Financial Economics* 111 (2): 495–525.
Davis, A., J. Piger, and L. Sedor. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research* 29 (3): 845–868.
Dhaliwal, D.S., Khurana, I.K., and Pereira, R. 2011. Firm disclosure policy and the choice between private and public debt. *Contemporary Accounting Research* 28 (1): 293–330.
Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics* 64 (2–3): 221–245.
Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2010. Management's tone change, post earnings announcement drift and accruals. *Review Accounting Studies* 15: 915–953.
Frankel, R., J. Jennings, and J. Lee. 2016. Using unstructured and qualitative disclosures to explain accruals. *Journal of Accounting and Economics* 45 (2): 209–227.
Frankel, R., Jennings, J., & Lee, J. (2019). Assessing the relative explanatory power of narrative content measures using conference calls, earnings predictions and analyst revisions. Working paper.
Ganguin, B., and J. Bilardello. 2005. *Fundamentals of corporate credit analysis*. New York: McGraw-Hill.
Greene, W. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal* 7: 98–119.
Ham, C., and K. Koharki. 2016. The association between corporate general counsel and firm credit risk. *Journal of Accounting and Economics* 61: 274–293.
Henry, E. 2008. Are investors influenced by how earnings press releases are written? *International Journal of Business Communication* 45 (4): 363–407.
Henry, E., and A. Leone. 2016. Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review* 91 (1): 153–178.
Hillegeist, S., E. Keating, D. Cram, and K. Lundstedt. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9 (1): 5–34.
Hollander, S., M. Pronk, and E. Roelofsen. 2010. Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research* 48 (3): 531–563.
Huang, A., R. Lehavy, A. Zang, and R. Zheng. 2018. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science* 64 (6): 2833–2855.
Hull, J., M. Predescu, and A. White. 2004. The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking and Finance* 28: 2789–2811.
Kang, J.K., Williams, C., and Wittenberg-Moerman, R. (2020). CDS trading and non-relationship lending dynamics. *Review of Accounting Studies*, forthcoming.
Kim, I., and D. Skinner. 2012. Measuring securities litigation risk. *Journal of Accounting and Economics* 53 (1–2): 290–310.
Kim, J., P. Shroff, D. Vyas, and R. Wittenberg-Moerman. 2018. Credit default swaps and managers' voluntary disclosure. *Journal of Accounting Research* 56 (3): 953–988.
Kim, E., Sethuraman, M., & Steffen, T. (2019). The informational role of investor relations officers: Evidence from the debt market. Working paper.

Kravet, T., and V. Muslu. 2013. Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies* 18: 1088–1122.
Lee, J. 2016. Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review* 91 (1): 229–250.
Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45: 221–447.
Li, F. 2010. The information content of forward-looking statements in corporate filings–a naïve Bayesian machine learning approach. *Journal of Accounting Research* 48 (5): 1049–1102.
Liberti, J., and M. Petersen. 2019. Information: Hard and soft. *The Review of Corporate Finance Studies* 8 (1): 1–41.
Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65.
Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69: 1643–1671.
Manela, A., and A. Moreira. 2017. News implied volatility and disaster concerns. *Journal of Financial Economics* 123: 137–162.
Martin, X., and S. Roychowdhury. 2015. Do financial market developments influence accounting practices? Credit default swaps and borrowers' reporting conservatism. *Journal of Accounting and Economics* 59 (1): 80–104.
Mayew, W.J. 2008. Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research* 46 (3): 627–659.
Mayew, W.J., M. Sethuraman, and M. Venkatachalam. 2015. MD&A Disclosure and the Firm's ability to continue as a going concern. *The Accounting Review* 90 (4): 1621–1651.
Merton, R. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance* 29 (2): 449–470.
Morgan, D. 2002. Risk and uncertainty in an opaque industry. *American Economic Review* 92 (4): 874–888.
Muslu, V., S. Radhakrishnan, K.R. Subramanyam, and D. Lim. 2014. Forward-looking MD&a disclosures and the information environment. *Management Science* 61 (5): 931–948.
Norden, L., and M. Weber. 2004. Informational efficiency of credit default swap and stock markets: The impact of credit rating announcements. *Journal of Banking and Finance* 28: 2813–2843.
Ohlson, J. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 19: 109–131.
Plumlee, M., Y. Xie, M. Yan, and J. Yu. 2015. Bank loan spread and private information: Pending approval patents. *Review of Accounting Studies* 20 (2): 593–638.
Price, S., J. Doran, D. Peterson, and B. Bliss. 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance* 36: 992–1011.
Sethuraman, M. 2019. The effect of reputation shocks to rating agencies on corporate disclosures. *The Accounting Review* 94 (1): 299–326.
Standard & Poor's. 2007. *A guide to the loan market*. New York: Standard & Poor's.
Standard & Poor's. (2015). Form NRSRO. Available at: http://www.standardandpoors.com
Tetlock, P., M. Saar-Tsechansky, and M. Sofus. 2008. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance* 63 (3): 1437–1467.
Vashishtha, R. 2014. The role of bank monitoring in borrowers' discretionary disclosure: Evidence from covenant violations. *Journal of Accounting and Economics* 57 (2–3): 176–195.
Watts, R. 2003. Conservatism in accounting part I: Explanations and implications. *Accounting Horizons* 17 (3): 207–221.
Zhang, J. 2008. The contracting benefits of accounting conservatism to lenders and borrowers. *Journal of Accounting and Economics* 45: 27–54.
Zhang, B., H. Zhou, and H. Zhu. 2009. Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *Review of Financial Studies* 22 (12): 5099–5131.

Affiliations

John Donovan[1] **· Jared Jennings**[2] **· Kevin Koharki**[3] **· Joshua Lee**[4]

[1] Mendoza College of Business, University of Notre Dame, Notre Dame, IN, USA

[2] Olin Business School, Washington University in St. Louis, Campus Box 113, St. Louis, MO 63130-4899, USA

[3] Krannert School of Management, Purdue University, West Lafayette, IN, USA

[4] Marriott School of Business, Brigham Young University, Provo, UT, USA