Can Language Predict Bankruptcy? The Explanatory Power of Tone in 10-K Filings*

KERSTIN LOPATTA, *University of Oldenburg*

MARIO ALBERT GLOGER, *University of Oldenburg*

REEMDA JAESCHKE, *University of Oldenburg*

Received on April 20, 2016; editorial decision completed on August 1, 2017

ABSTRACT

We examine whether the language used in 10-K filings reflects a firm's risk of bankruptcy. Our sample contains 424 bankrupt U.S. companies in the period 1994–2015 and we use propensity score matching to find healthy matches. Based on a logit model of failing and vital firms, our findings indicate that firms at risk of bankruptcy use significantly more negative words in their 10-K filings than comparable vital companies. This relationship holds up until three years prior to the actual bankruptcy filing. With our investigation, we confirm the results from previous accounting and finance research. 10-K filings contain valuable information beyond the reported financials. Additionally, we show that 10-Ks filed in the year of a firm's collapse contain an increased number of litigious words relative to healthy businesses. This indicates that the management of failing firms is already dealing with legal issues when reporting financials prior to bankruptcy. Our results suggest that analysts ought to include the presentation of financials in their assessment of bankruptcy risk as it contains explanatory and predictive power beyond the financial ratios.

Keywords Language analysis; Bankruptcy; Sentiment

LE LANGAGE EMPLOYÉ PERMET-IL DE PRÉDIRE LA FAILLITE ? LE POUVOIR EXPLICATIF DU TON ADOPTÉ DANS LES DÉCLARATIONS 10-K

RÉSUMÉ

Les auteurs se demandent si le langage utilisé dans les déclarations 10-K révèle l'existence d'un risque de faillite de l'entreprise. Ils analysent un échantillon de 424 sociétés des États-Unis ayant fait faillite entre 1994 et 2015 et utilisent l'appariement des coefficients de propension pour déterminer quelles sont les correspondances d'entreprises saines. En appliquant un modèle logit aux entreprises défaillantes et aux entreprises saines, ils constatent que celles qui présentent un risque de défaillance font un usage sensiblement plus abondant de termes négatifs dans leurs déclarations 10-K que les entreprises saines comparables. Cette relation perdure jusqu'à trois ans avant le réel dépôt de bilan. L'étude confirme les

* Accepted by Pascale Lapointe-Antunes.

résultats de travaux de recherche précédents en comptabilité et en finance. Les déclarations 10-K contiennent de précieux renseignements au-delà des données financières dont elles font état. De plus, les auteurs montrent que les déclarations 10-K produites au cours de l'année de la débâcle d'une entreprise contiennent une quantité plus importante de termes litigieux par rapport à celles des entreprises saines. Cette observation révèle que la direction des entreprises défaillantes est déjà aux prises avec des problèmes d'ordre juridique lorsqu'elle produit les résultats financiers de la société avant la faillite. Ces constatations laissent croire que les analystes se doivent d'inclure la présentation des données financières dans leur évaluation du risque de défaillance, cette présentation ayant un pouvoir explicatif et prédictif qui s'ajoute à celui des ratios financiers.

Mots clés : Analyse linguistique, Faillite, Opinion

INTRODUCTION

Given the immense costs of bankruptcy (Branch, 2002; Bris, Welch, and Zhu, 2006), academics are interested in identifying the indicators of a firm's collapse, and in turn, giving investors methods to detect red flags and therefore minimize the risk of investing in failing companies. Yet, despite all efforts, no comprehensive model has been established to date. Boritz (1991) already deplores the rare disclosure of going-concern issues in failing companies, along with inappropriate accounting and auditing standards. So far few significant improvements have been made. While there are financial ratio frameworks like Altman's (1968) *Z*-score and Ohlson's (1980) O-score, or market-based models such as BSM-Prob (Hillegeist, Keating, Cram, and Lundstedt, 2004), no approach is free of disadvantages and drawbacks.

Current research links firm performance to softer, more qualitative factors. Examples include the influence of media pessimism on market prices (Tetlock, 2007) and how a CEO's personality traits relate to fraud or firm performance (Cohen, Ding, Lesage, and Stolowy, 2010; Peterson, Walumbwa, Byron, and Myrowitz, 2009). In this context, the evaluation of language of corporate filings has gained particular interest for academic researchers, who attempt to extract information about fraud and firm performance from the terminology used in corporate disclosures. Their findings are auspicious, indicating that the tone of earnings press releases is related to future firm performance (Davis, Piger, and Sedor, 2012). Management's opinion of the going concern combined with the language used in the MD&A section contains explanatory power on whether a firm can continue as a going concern (Mayew, Sethuraman, and Venkatachalam, 2015). Rogers, van Buskirk, and Zechman (2011) show that the use of optimistic language increases litigation risk. Also, the readability and language sentiment of earnings press releases influences investors' judgments, depending on their level of sophistication (Tan, Wang, and Zhou, 2014).

Still, this research is not an exclusively academic topic but in fact highly relevant for practioners, with the audit profession taking note of the additional information contained in language. For example, the Big 4 audit firm PwC cooperates with companies offering natural language processing using all sorts of unstructured data ranging from call centers, social media, or even internal documents to help their clients better understand customers and social sentiment (PwC, 2012). Another Big 4 auditor, KPMG, has identified fraud investigations as a potential field for applying language analysis using email/chat data, sentiment analysis on social media platforms, and even call center audio data to identify any red flags (KPMG, 2014). This confirms the immense relevance of the topic for both practioners and scholars. For instance, a change in customer sentiment toward a firm can only be detected if customers' language is analyzed. The results can serve as an early warning sign even before the company's financials react. Hence, language variables that enrich financial ratio models can be valuable indicators of a company's condition. This is the research subject of this study and, if our theory is proven, will raise even greater awareness of this topic among regulators, users, and preparers of financial statements and confirm the actions taken by PwC, KPMG and all providers of these techniques.

We combine traditional bankruptcy prediction models with language evaluation based on Loughran and McDonald's (2011) word lists and the comprehensive LoPucki bankruptcy data set to examine whether the language used in 10-K reports contains explanatory power with respect to corporate distress. We match bankrupt firm-year observations to non-bankrupt firm-years based on propensity score matching. Hence we incorporate bankruptcy-indicating financial ratios and language variables into logit models, with the binary dependent variable equal to one if a company has filed for bankruptcy. An additional model combining recent financial ratios with the language used in the 10-K of the previous year enables this study to demonstrate the predictive power of language.

Our study makes two major contributions. First, it promotes the inclusion of language evaluation in bankruptcy research. Davis et al. (2012) suggest a positive relationship between firm performance and optimistic tone, a relationship we assume also holds in the other direction: companies that perform so poorly that they have to file for bankruptcy protection ought to use significantly more negative words than their more vital peers. After controlling for relevant financial ratios, we find that this effect is strong until three years prior to filing for bankruptcy in our U.S. sample. A lagged language variable further indicates the predictive power of tone analysis. This contributes to the limited literature on the usefulness of this non-financial characteristic in forecasting financial distress. Second, we find that litigious terms contain explanatory power with respect to bankruptcy, even though the effect is weaker and more ambiguous than for negative terms. This has not been shown before and calls for more detailed research in this area. Our findings indicate that the presentation of financial disclosures changes in periods of financial hardship. This corresponds to a study by Holder-Webb and Cohen (2007), who finds that disclosure quality increases when facing bankruptcy. All in all, we provide strong motivation for exploring these avenues further and to include language variables in bankruptcy models, particularly as several studies show that disclosures about the potential invalidity of a going-concern paradigm are published either too late or not at all (Boritz, 1991). This once again indicates the high relevance for auditors. Language evaluation could complement ratio analyses by providing an extra measure of information; or it could be used as a cost effective way to obtain a brief overview of a firm's condition.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Literature Review

Researchers have evaluated the predictive power and influence of language in financial communication with different goals. With respect to bankruptcy, research is limited: the only studies on this subject are those of Henselmann and Scherr (2012) and Mayew et al. (2015). The former examines just a small sample of 26 bankrupt companies, matching them only to their biggest competitor. Language is considered just one of many other red flags extracted from XBRL filings. In the latter study, the authors focus on the assessment of the ability of a firm to continue as a going concern. Thus, by formulating their research question in the way they do, they rather evaluate the auditing perception of a firm. To this end the authors incorporate financial ratios and the auditor's as well as management's qualitative going-concern opinion into their model. Prior research has shown that the auditor's opinion has slightly weaker predictive power than bankruptcy prediction models with regard to actual bankruptcies (Altman and McGough, 1974; Cormier, Magnan, and Morard, 1995; Koh, 1991; Koh and Killough, 1990; McKee, 2003). It is debatable whether audit reports signal useful incremental information about the probability of bankruptcy (Lennox, 1999). Therefore, our study excludes qualified going-concern opinions and solely focuses on litigious, positive, and negative terms used in 10-Ks. Furthermore, Mayew et al. (2015) compare bankrupt firms to all available firm-year observations from COMPUSTAT. Although the authors control for financial ratios in their model, a propensity score matching based on those ratios can improve the accuracy of the results. Finally, as they do not consider a lagged language variable, Mayew et al. (2015) do not contribute to the discussion on the predictive ability of language with respect to bankruptcies.

Besides being one factor of many that can explain corporate distress in the study of Mayew et al. (2015), language has been linked to other research areas. Such studies are often based on business-specific word lists (Loughran and McDonald, 2011). First attempts to define disclosure quality with respect to narratives are made by Beattie, McInnes, and Fearnley (2004). Furthermore, the use of plain, concise language increases investors' willingness to invest in projects, particularly when they feel psychologically more distant to a firm (Elliott, Rennekamp, and White, 2015). The effect of sentiment on asset prices, proxied by the fraction of negative and positive words in financial news from the *New York Times*, is reported by García (2013). Schleicher and Walker (2010) show that companies bias the narratives in forward-looking sections of their annual reports with respect to firm characteristics.

Also, scholars use language to detect fraud in annual reports (Chen, 2014; Goel, Gangolly, Faerman, and Uzuner, 2010). Moreover, the tone of disclosure can hint at litigation risks (Rogers et al., 2011). In addition, studies suggest that qualitative verbal information can be used to predict firms' accounting earnings and stock returns, for example, a prevailing negative tone is said to forecast low earnings (Tetlock, Saar-Tsechansky, and Macskassy, 2008). Tan et al. (2014) bring together both language and readability of annual reports, stating that tone in disclosures influences investors particularly when readability is simultaneously low. Huang, Teoh, and Zhang (2014) go beyond that to show that managers strategically use tone management to mislead investors. For all mentioned reasons, incorporating language can help to extract more information from 10-Ks beyond the financials.

Traditionally, only bankruptcy prediction models based on financial ratios have been developed. At this point, language variables could be included to improve the accuracy of the models. A very important contribution in this context is the study by Beaver (1966), which analyzes ratios in six categories: profitability, liquidity, leverage, cash flow, cash and liquid positions, and turnover. Those categories are common in bankruptcy research—almost all scholars incorporate at least one ratio from three or four groups into their models (e.g., Altman, Haldeman, and Narayanan, 1977; Blum, 1974; Meyer and Pifer, 1970). Altman (1968) bridges the gap between traditional ratio analysis and rather sophisticated methods by developing one of the first functioning MDAs for use in this context. Based on five ratios, Altman (1968) calculates a firm-specific Z-score that describes a firm's probability of bankruptcy. In further research, Altman et al. (1977) refines the model to calculate the Z-score from an extended set of seven ratios incorporated into a linear framework. If only two groups (e.g., non-failed and failed corporations) are important for the study, the MDA can be extended to a logit system by constraining the dependent variable to values between zero and one (Cramer, 2002), which meets statistical assumptions better than MDAs.

Nonetheless, evaluating language in official company documents not only increases the accuracy of financial ratio-based bankruptcy prediction but can also help to interpret respective financial ratios better, as it reveals the specific business context in which the ratios are calculated. In other words, the ratios change according to the business environment, which is in turn explained by the use of language in reports. This has been illustrated by Magnusson et al. (2005); they also state that language rather reflects the future plans of the firm, while ratios reflect the current state. This relationship ultimately hints at the importance of language when evaluating a firm's financials and condition.

Shumway (2001) adds to the literature by arguing when to consider a firm bankrupt or healthy. He applies hazard methodology by only calling the last available set of accounting data of a failing firm "bankrupt," while earlier observations remain in the sample as healthy firm-year observations. Other authors like Wu, Gaunt, and Gray (2010) follow his approach. Our paper applies the traditional understanding of a healthy firm and evaluates whether the overall results change due to a change to hazard methodology in the robustness test section. We can say that the relationship between negative language and bankruptcy is robust to this change; by contrast, under Shumway's (2001) hazard assumptions, litigious words do not show explanatory power.

Still, neither versions of Altman's Z-score and none of the other models constitute a comprehensive bankruptcy prediction model. Regardless of its relatively cost-efficient nature, Altman's (1968) model is rather backward-looking and offers no predictive value. Also, using accounting-based measures from financial statements which are prepared with a going-concern assumption to predict firm failures can be considered a paradox from the beginning (Li, 2012). Despite all disadvantages, several accounting-based bankruptcy prediction models have been developed that are distinguishable mainly by the ratios used (Altman, 1968; Altman et al., 1977; Beaver, 1966; Libby, 1975). To improve the accuracy of bankruptcy prediction models, we enrich traditional financial ratios with modern language analysis.

Hypothesis Development

The management of a company files for bankruptcy when debts due cannot be repaid or in cases of excessive indebtedness. Management literature identifies three typical consecutive phases of corporate crisis that lead to bankruptcy: strategy crisis, performance crisis, and liquidity crisis (e.g., Bickhoff, Blatz, Eilenberger, Haghani, and Kraus, 2004; Thießen, 2013). The risk of bankruptcy increases from one stage to the next. A strategy crisis, if recognized in time, can be resolved if turnaround efforts start early enough (Sheppard, 1994). During the performance stage, the crisis starts to affect traditional performance goals like profitability, sales, or EBIT. The probability of a successful turnaround decreases. The last stage before bankruptcy is reached when serious liquidity concerns become obvious (Thießen, 2013). So, except for significant shocks such as fraud detection[1] or uninsured natural disasters that can destroy almost all of a firm's capital, it can be assumed that filing for bankruptcy is the result of a long-term process which should be reflected in the firm's disclosure.

From an auditor standpoint, Boritz (1991, 27) provides an important statement in this respect that obviously relates to financial indicators: "Except in rare circumstances, the going concern assumption becomes questionable some time before its questionability becomes apparent by virtue of obvious and concrete indicators." Together with the finding that language predicts fraud and litigation risks (Chen, 2014; Goel et al., 2010; Purda and Skillicorn, 2015; Rogers et al., 2011), this backs up our assumption that language could also be used to predict bankruptcies. Furthermore, Tetlock et al. (2008) and García (2013) state that a greater number of negative words in the financial press forecasts low firm earnings. Low earnings can be considered a major component of poor firm performance. Thus, when the company's performance weakens to the extent that the firm's earnings have been too low to support operations for multiple years, the probability of bankruptcy increases. As a result, negative language can be hypothesized to have a positive relationship with bankruptcy probability.

1. Consider, for example, the Enron fraud scandal.

In this context, Mayew et al. (2015) suggest that negative wording in the MD&A section is indicative of a potential going-concern problem and therefore of an increased likelihood of bankruptcy. They find that this effect holds until 3 years prior to bankruptcy. Also, annual reports help to minimize information asymmetries between the management of the company and its owners and therefore tackle the agency issues described by Marino and Ozbas (2014), Jensen and Meckling (1976), Grossman and Hart (1983), and Gaston (1997). The SEC's 1998 Plain English Initiative regarding corporate disclosures was an even stronger expression of investors' call for official filings to accurately reflect the company's asset, financial, and earnings situation. Summing up, this study expects the language in 10-Ks to be more negative immediately before a bankruptcy filing than in regular business periods. This leads to our first hypothesis:

Hypothesis 1. *Failing companies use more negative language than non-failing companies.*

Furthermore, bankruptcies lead to court cases and routines (White, 1989). Courts appoint insolvency administrators to handle the cases and prepare the ground for a fair course of justice. Management literature argues that legal confrontations contain strong potential for triggering corporate crises (Thießen, 2013). In addition, litigation-specific public relations (PR) strategies are recommended. One major element of a PR strategy is the publication of messages to defend the company's reputation (Thießen, 2013). One can assume that the tone and language used is very important in this context. Occasionally criminal charges are filed against management or other individuals associated with the failing company. A firm in this situation will have to plan its next steps carefully and is legally obliged to inform its stakeholders about the charges. No study has tried to link litigious terminology to financial distress yet. Therefore, based on the same data as Hypothesis 1, our second hypothesis is as follows:

Hypothesis 2. *Failing companies use more litigious language than non-failing companies.*

METHODOLOGY

Research Design and Variable Measurement

Research Design

We apply propensity score matching to select appropriate control firm-year observations. This is more accurate than matching only on industry and size and also reduces concerns about dissimilar firms by controlling for bankruptcy-indicating factors (Dehaan, Hodge, and Shevlin, 2013). Established financial ratios, together with time and industry dummies, ensure that we find control firms that are similar to failing firms with respect to these variables.

We further choose logit methodology over linear MDAs to test for Hypotheses 1 and 2. Although traditional bankruptcy research, for example, Altman (1968) and Beaver (1966), is mainly conducted based on MDAs, Eisenbeis (1977) identifies three main disadvantages of this methodology. First, the assumption of equal group dispersion matrices over all groups is not realistic. In the likely case of unequal dispersions (due to differences in group size or an increasing number of variables), significance levels are likely to exceed hypothesized levels and therefore produce biased results. Second, MDAs are based on the normal distribution assumption. Intuitively, a bell-shaped curve for a binary dependent variable that is constrained between 0 and 1 is not imaginable. Third, it is not possible to interpret the relative importance of variables in an MDA since there is no test for the absolute value of each individual variable. There is no unique coefficient emerging from all ratio values taken together—it is still the ratio which is unique.

By contrast, a standard conditional logit model is not based on a certain distribution and equal group dispersion and therefore provides more accurate results than MDAs (Ohlson, 1980). Furthermore, coefficients can be interpreted in terms of their contribution to the probability of the dependent variable being "1" (bankrupt) or "0" (not bankrupt). More importantly, their relative values indicate their relative influence on the binary variable (Roncek, 1991). The final, and clearest, reason to apply logit methodology to our propensity score-matched samples is provided by Khalili Araghi, and Makvandi (2013): Using the same sample, the authors compare the accuracy of logit, probit, and MDA models and give evidence for the superiority of the first when predicting bankruptcies. All in all, the combination of propensity score matching and logit methodology ensures that the results are not biased by self-selection, as described by Heckman (1979), and therefore increases their accuracy and verifiability compared to solely matching on industry and size.

Variable Measurement

We include two measures to capture information stemming from language in 10-K disclosure based on the dictionary of Loughran and McDonald (2011). We use complete 10-Ks (excluding tables) instead of annual reports because they are more lengthy and comprehensive, are based on very strict SEC guidelines and therefore offer a higher degree of standardization across all companies. This should reduce management's impact on presentation. Annual reports, on the other hand, are designed according to management's preferences and are ultimately intended to attract investors rather than fulfill the information needs of the SEC. Additionally, Guay, Samuels, and Taylor (2016) demonstrate that firms show different voluntary disclosure behaviors specifically before and after publishing a 10-K, from which we conclude the strong importance of a 10-K filing for each company.

Our first measure is *NEGATIVE*, calculated as the number of negative words minus the number of positive words divided by the total number of words in a 10-K. To show the values in percent, we multiply the value by a factor of 100. Examples of negative words are *breakdown*, *catastrophe*, and *demolish*. By contrast, words like *achievement*, *advantage*, and *reward* belong to the positive word list. The computation of the variable combines previously separate word lists into one comprehensive measure of overall language and therefore avoids probable collinearity between both groups and narrows the analysis down to one joint variable. The second measure, *LITIGIOUS,* calculated as the number of litigious words divided by the total number of words and multiplied by a factor of 100, captures the explanatory power of litigious terms. Examples are *attorney*, *felony*, or *justice*. If language contains explanatory power, we would expect both variables' coefficients to be negative, as reflected in our hypotheses.

Furthermore, we apply five control variables to our propensity score matching model and regression models. All of them have been proven to discriminate between bankrupt and non-bankrupt firms. First, we incorporate the profitability measure EBIT divided by total assets into our model, following, for example, Altman (1968), Altman et al. (1977), Ohlson (1980), and Beaver, McNichols, and Rhie (2005), which is referred to as *PROFITABILITY*. This measure is part of Altman's (1968) *Z*-score, which assumes that more profitable firms are less likely to file for bankruptcy.

Second, to control for differences in the liquidity situation of bankrupt and non-bankrupt firms, we include the current ratio in our model, following, inter alia, Altman's (1968) *Z*-score framework, Libby (1975), Altman et al. (1977), and Zmijewski (1984). Hence, more liquid companies are considered less likely to file for bankruptcy. This variable is referred to as *LIQUIDITY*.

Third, we control for size by including the natural logarithm of market value of equity in our model. This variable is referred to as *SIZE*. Size has been shown to discriminate between bankrupt and vital firms in prior studies (e.g., Altman et al., 1977; Darrat, Gray, Park, and Wu, 2014; Ohlson, 1980; Tykvová and Borell, 2012). Larger firms are therefore assumed to have a lower bankruptcy probability. By choosing the natural logarithm of market value of equity instead of total assets which is a commonly used indicator for size, we include another market-based factor. This responds to a criticism that is commonly leveled at Altman's *Z*-score and traditional ratio frameworks in general, namely that such models are backward-looking and do not have predictive power for the future without incorporating market-based measures (Li, 2012). Furthermore, the natural logarithm of the market value of equity is highly correlated with the natural logarithm of total assets (almost 90 percent), although the variance inflation factor (VIF) is significantly lower.

Fourth, we incorporate a cash ratio into our model, calculated as cash and other short-term investment positions divided by total assets, assuming that firms with more cash are less likely to file for bankruptcy. In so doing, we follow Beaver (1966), Meyer and Pifer (1970), and Libby (1975). This variable is referred to as *CASH*.

Fifth, we include a market-related variable. Certain different market-based variables can be found in the bankruptcy literature, for example, in Altman (1968), Altman et al. (1977), and Griffin and Lemmon (2002). Following Darrat et al. (2014), we opt for market-to-book ratio because of its frequent practical use and because it combines market-based with accounting-based data. Hereafter, the market-to-book ratio is referred to as *MARKET*. We deliberately exclude stock return-based market measures because of their unclear explanatory power with respect to bankruptcy in general (Aharony, Jones, and Swary, 1980) and the fact that they only have a short-term effect (up to seven months), as demonstrated in studies that find a significant correlation between stock return measures and bankruptcy (Frino, Jones, and Wong, 2007). Since we focus on the long term with up to five years before bankruptcy, stock-based measures are not appropriate variables.

Sample

Our sample is based on data from three sources. Bankruptcy data are obtained from the UCLA LoPucki Bankruptcy Research Database (BRD). This database provides information on 1,013 large public U.S. firms that filed for bankruptcy between 1980 and 2015 (LoPucki, 2015). It is the basis for several other studies such as Cao and Narayanamoorthy (2014), Correia, Richardson, and Tuna (2012), and Jiang, Li, and Wang (2012). In the unedited database, 226 CIK codes are missing. We obtain this information manually from the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) company lookup mask on the SEC's website, where possible.[2] After eliminating all except the first bankruptcy of the same firm, 840 bankruptcies remain in the sample.

The second source is the database by Loughran and McDonald (2011) on language in corporate disclosure. This database includes language information for more than 260,000 10-Ks. For bankrupt firms, we restrict the sample to one 10-K per year, published between six years before the bankruptcy filing and the bankruptcy filing date. We choose this time span as earlier bankruptcy research based on financial ratios reports the explanatory power of their models up to five years before bankruptcy (e.g., Altman et al., 1977; Beaver, 1966). Although we also restrict the number of 10-Ks per year to one, this time span restriction does not apply for 10-Ks filed by healthy companies. They constitute the pool from which the control observations based on propensity score matching are drawn.[3]

Finally, we obtain data for the computation of our financial ratios from the COMPUSTAT North America Financial Database. We drop all firm-year

2. http://www.sec.gov/edgar/searchedgar/webusers.htm.

3. Under some circumstances, firms are obliged to publish multiple annual filings, for example, to adjust previously filed reports or if their fiscal year changes. For consistency, we establish a ranking of form types with respect to the relevance and extent of the publication. If a company filed more than one report per year, the lower number is kept in the sample, the other report is eliminated. If two similar reports are issued the same year, the earlier one is kept in the sample, the other is eliminated. The established ranking is as follows: (1)10-K; (2)10-K405 (no substantial difference to the regular 10-K form; this only indicates that the company failed to report its insider-trading activities. Also included by Loughran/McDonald, 2011); (3) 10-KSB (the 10-K for smaller businesses and an abbreviated version of the 10-K); (4) 10-KSB40 (for 10-KSB issuers that failed to file forms 3, 4, and 5 on time); (5) 10-K/A (contains an amendment to a previously filed 10-K); (6) 10-K405/A (contains an amendment to a previously filed 10-K405); (7) 10-KSB/A (contains an amendment to a previously filed 10-KSB); (8) 10-KSB40/A (contains an amendment to a previously filed 10-KSB40).

observations without a complete set of financials. The final sample including language data and financials as well as bankruptcy information consists of 93,134 firm-year observations of which 1,826 are associated with firms that failed over the subsequent six years. This corresponds to approximately 2 percent of the total reports. These 1,826 failing firm-year observations are again divided into subsamples according to the difference in years between filing of the 10-K and filing of bankruptcy. Therefore, if a company filed for bankruptcy in 2005, the 10-K for the fiscal year 2005 (if available) is assigned with a 0. Hence, the 10-K for 2004 is assigned with a 1, respectively. These subsamples are described in more detail in the descriptive statistics section of this paper. Table 1 provides an overview of the sample creation steps.

Propensity Score Matching

The matching procedure is important for this study. We apply one to one propensity score matching, whereas other papers only rely on industry, size and/or year (Altman, 1968; Beaver, 1966; Goel et al., 2010; Purda and Skillicorn, 2015). To ensure that the best possible firm-year observation from the control sample is matched to each bankrupt firm, we allow replacement, so that two or more bankrupt firms can be matched to the same control firm (Table 2 shows that this effect is minimal). In those cases, the control firms' weights increase according to how many bankrupt firms they match. The matching model can be described as a logit model with a binary dependent variable *BANKRUPT* that is equal to one for firm-years associated with a bankruptcy or zero if otherwise.

The next step involves selecting appropriate propensity model variables. The purpose of matching is to combine a group of the same property (in this case, companies that are going bankrupt) with a group without this property—but with equal characteristics with respect to the propensity model variables. Because we aim to show whether or not the language in 10-Ks has explanatory power beyond financials with respect to bankruptcy prediction, we choose the five financial ratios that have been proven to contain valuable information on bankruptcy prediction and which are discussed in the variable measurement section.[4] This implies that both bankrupt and non-bankrupt control firms are similar in terms of these financial ratios. Finally, year (*TIME*) and Fama-French 12 industry codes (*FFINDS12*)[5] are included in the propensity model as binary independent variables. In doing so, we also control for time and industry effects as other authors do, although these characteristics can be overruled in the propensity score. We choose *FFINDS12* over Fama-French 48 industry codes or other industry classification schemes, since applying a more detailed matrix with more categories is likely to cause problems

4. The five chosen variables are EBIT divided by total assets (*PROFITABILITY*), current ratio (*LIQUIDITY*), logarithm of market value of equity (*SIZE*), cash and other short-term investment positions divided by total assets (*CASH*), and the market-to-book ratio (*MARKET*).
5. The Fama-French 12 industry codes are defined on Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html).

TABLE 1
Sample creation steps

Steps to achieve the final sample	
Step 1: UCLA LoPucki Bankruptcy Database	
Bankruptcy cases from UCLA LoPucki Database	**1,013**
Drop firms without CIK identifier	−114
Drop firms that did not file for bankruptcy the first time	−59
Drop firms without evaluated 10-Ks in Loughran/McDonald	−31
Number of firms	**809**
Step 2: Loughran/McDonald's language database	
Evaluated 10-K and 10-Q variants in Loughran/McDonald database	**910,645**
Drop 10-Q variants	−646,633
Useable 10-K filings	**264,012**
Merged useable bankruptcies with useable 10-K filings	**264,012**
10-Ks associated with failing firms	8,474
Drop 10-Ks filed after bankruptcy	−2,579
Drop 10-Ks filed more than 2,190 days before bankruptcy	−1,870
Useable 10-Ks associated with failing firms	4,025
10-Ks associated with control firms	255,538
Subtotal	**259,563**
Keep only one 10-K per year	−51,267
Drop control firms without fiscal year end data	−1,139
Drop reports with overlapping fiscal years	−1,967
Subtotal	**205,190**
Step 3: COMPUSTAT North America Financial Database	
Merged with COMPUSTAT Financial Data	**205,190**
Drop all 10-Ks without words	−16
Drop firm-year observations denoted in Canadian dollars	−223
Drop all observations with missing data to compute ratios	−111,817
Total	**93,134**
Of which associated with failing firms	1,826
Of which associated with control firms	91,308

Notes:

The sample is based on three databases: UCLA LoPucki Bankruptcy Database, Loughran and McDonald's (2011) language database, and COMPUSTAT. Only U.S. firms are considered.

with finding appropriate matches for all industry classes due to the rather small sample size when using the two-step matching approach described in the robustness tests. Furthermore, heteroscedasticity-robust standard errors are used and

TABLE 2
Definition of subsamples

Sample	Description	Unique bankrupt firm-year observations	Unique control firm-year observations
(1)	All 10-Ks associated with bankruptcy and their matches	1,826	1,759
(2)	10-K is filed in the same calendar year as the bankruptcy	215	211
(3)	10-K is filed one calendar year prior to bankruptcy	387	363
(4)	10-K is filed two calendar years prior to bankruptcy	362	347
(5)	10-K is filed three calendar years prior to bankruptcy	309	299
(6)	10-K is filed four calendar years prior to bankruptcy	258	250
(7)	10-K is filed five calendar years prior to bankruptcy	212	207
(8)	10-K is filed six calendar years prior to bankruptcy	83	82

Notes:

Sample (1) includes all 10-Ks associated with bankruptcy and their matches. Sample (2) includes all 10-Ks issued by failing firms in the same year they filed for bankruptcy and their matches. Sample (3) includes all 10-Ks issued by failing firms in the calendar year prior to filing for bankruptcy and their matches. Sample (4) includes all 10-Ks which are issued by failing firms two calendar years before bankruptcy and their matches. Sample (5) includes all 10-Ks which are issued by failing firms three calendar years before bankruptcy and their matches. Sample (6) includes all 10-Ks which are issued by failing firms four calendar years before bankruptcy and their matches. Sample (7) includes all 10-Ks which are issued by failing firms five calendar years before bankruptcy and their matches. Sample (8) includes all 10-Ks which are issued by failing firms six calendar years before bankruptcy and their matches.

variables are winsorized to the first and 99th percentile. The propensity matching model reads as follows:

$$\begin{aligned} BANKRUPT_{it} = {} & \alpha + \beta_1 PROFITABILITY_{it} + \beta_2 LIQUIDITY_{it} + \beta_3 SIZE_{it} \\ & + \beta_4 CASH_{it} + \beta_5 MARKET_{it} + \sum TIME_t \\ & + \sum INDUSTRY_i + \varepsilon_{it}. \end{aligned} \tag{1}$$

The results of this model are shown in Table 3.

All coefficients are highly significant at the 1 percent level and indicate that bankrupt firm-year observations in the sample are less liquid, have smaller market values of equity, and less cash. Interestingly, they are also larger and more profitable. The latter has been shown by Mayew et al. (2015), too. The Pseudo-R^2 equals about 10 percent and is therefore close to the Pseudo-R^2 of comparable financial variable models before including language variables (Mayew et al., 2015).

The distribution of the propensity score (p-score) of the sample is provided in panel B of Table 3. The distribution is positively skewed. A total of six different

TABLE 3

Propensity score matching

Panel A: Results of propensity score matching

Variables	Propensity score matching model *BANKRUPT*
PROFITABILITY	0.328***
	(0.000)
LIQUIDITY	−0.050*
	(0.072)
SIZE	0.108***
	(0.000)
CASH	−1.603***
	(0.000)
MARKET	−0.016***
	(0.000)
Constant	−6.182***
	(0.000)
Observations	92,570
Pseudo R^2	0.096
Time	Yes
Industry	Yes
Model	Logit

Panel B: Distribution of the propensity score over the complete sample.

Variable	*N*	Mean	SD	Min.	P25	Median	P75	Max.
p-score	92,570	0.020	0.021	0.000	0.005	0.013	0.030	0.281

Notes:

Panel A: Our propensity score matching model consists of the five variables *PROFITABILITY* (EBIT divided by total assets), *LIQUIDITY* (current ratio), *SIZE* (natural logarithm of market value of equity), *CASH* (cash divided by total assets), and *MARKET* (market-to-book ratio). Propensity score matching is conducted with replacement. All variables are winsorized to the first and 99th percentile to control for extreme values. Clustered and heteroscedasticity-robust standard errors are applied to compute *t*-statistics. * and *** represent significance levels of 10 percent and 1 percent, respectively. Thus, all five factors are statistically significant. The underlying COMPUSTAT data items to compute the financial ratios are: *SIZE* = ln (PRCC_F×CSHO); *CASH* = CHE/AT; *LIQUIDITY* = ACT/LCT; *PROFITABILITY* = EBIT/AT; *MARKET* = (PRCC_F×CSHO)/SEQ. Time as well as industry dummies are included. The model fit is reported by 9.6 percent.

Panel B: The *p*-score, computed for 92,570 firm-year observations, is positively skewed. This score is the basis for matching bankrupt firms to non-bankrupt companies.

subsamples are then created additionally, depending on the difference in years between filing for bankruptcy and the filing of the 10-K in question. All 10-Ks that are published in the same year as the company's bankruptcy filing are assigned to one sample. If evaluated 10-Ks are filed one fiscal year prior to bankruptcy, they are assigned to another sample. A list of the subsamples' compositions is provided in Table 2. We also conduct other, traditional matching routines that only focus on year and industry and lead to similar results. They are discussed in the robustness test section together with the results of the applied hazard methodology.

Model Specifications

Based on the matched samples, we estimate logit models with the binary dependent variable *BANKRUPT* equal to one if a firm files for bankruptcy or zero if it does not. Now, *NEGATIVE* and *LITIGIOUS* are included in the model to evaluate whether or not they have predictive power beyond the variables from propensity score matching. To assess if the propensity score matching is successfully applied, we consequently use all propensity matching ratios as control variables. They ought to no longer indicate any significant differences between failed and non-failed firms of the sample, because they are matched on those ratios. Again, using heteroscedasticity-robust and clustered standard errors as well as winsorized variables, the generalized formalized logit model reads as follows:

$$\begin{aligned} BANKRUPT_{it} = {} & \alpha + \beta_1 NEGATIVE_{it} + \beta_2 LITIGIOUS_{it} + \beta_3 PROFITABILITY_{it} \\ & + \beta_4 LIQUIDITY_{it} + \beta_5 SIZE + \beta_6 CASH_{it} + \beta_7 MARKET_{it} \\ & + \sum TIME_t + \sum INDUSTRY_i + \varepsilon_{it}. \end{aligned} \tag{2}$$

This general formula is then applied to the complete sample as well as to all subsamples, so that each model's dependent variable only contains defaulting firm-year observations with the same difference in years between the 10-K and the bankruptcy filing. The independent side of the model remains the same for all samples. Furthermore, to evaluate whether *NEGATIVE* contains predictive power with respect to bankruptcy, we replace $NEGATIVE_{it}$ with the lagged variable $LAG\_NEGATIVE_{it-1}$. t denotes the year of bankruptcy, $t-1$ the year before. By this we show whether the amount of *NEGATIVE* one year before bankruptcy already explains the firm failure in the subsequent year. The model changes to:

$$\begin{aligned} BANKRUPT_{it} = {} & \alpha + \beta_1 LAG\_NEGATIVE_{it-1} + \beta_2 LITIGIOUS_{it} \\ & + \beta_3 PROFITABILITY_{it} + \beta_4 LIQUIDITY_{it} + \beta_5 SIZE \\ & + \beta_6 CASH_{it} + \beta_7 MARKET_{it} + \sum TIME_t \\ & + \sum INDUSTRY_i + \varepsilon_{it}. \end{aligned} \tag{3}$$

DESCRIPTIVE STATISTICS AND LANGUAGE IN 10-KS

Descriptive Statistics

To begin with, we describe the complete sample of 1,826 10-Ks associated with a bankruptcy within the subsequent six years before we divide it into subsamples. The 1,826 10-Ks of the sample are filed by 424 unique companies. This equals an average of 4.3 reports per company. A 10-K is considered filed by a "bankrupt" firm if the company files for bankruptcy within the following six years. Therefore, each bankrupt entity can have a maximum of seven different filings in the sample. Moreover, all 10-K reports from the sample were filed between fiscal years 1993 and 2013.

Figure 1 provides the distribution of bankruptcy filings of the 424 unique sample firms. We observe a significant peak around the turn of the millennium with more than 50 percent of all sample firms filing for bankruptcy between 1999 and 2003. The likely explanation is the dotcom collapse, in the course of which a large number of companies experienced significant economic problems and eventually had to file for bankruptcy.[6]

In addition to that, we show that the amount of negative language increases the closer the bankruptcy approaches. This relationship is extremely stringent.

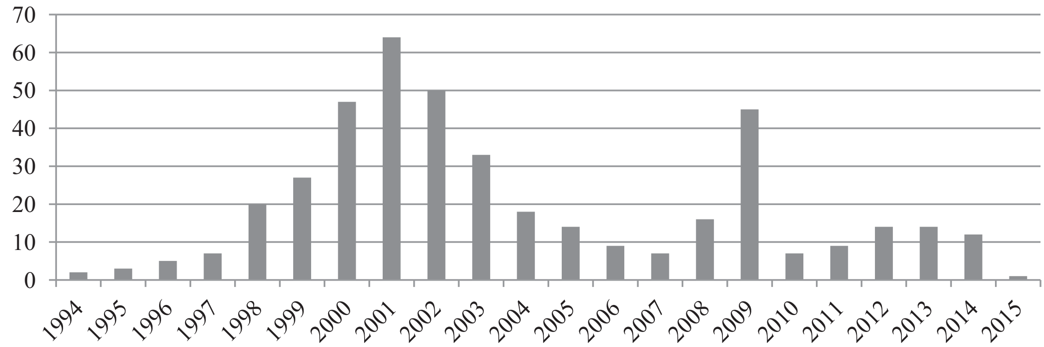
The left-hand side of Table 4 shows the summary statistics of all firm-year observations linked to bankruptcy with respect to all important variables, including control as well as language variables. The same statistics for control firm-year observations are shown in the middle of Table 4. Although the differences in those variables are small between both groups, they are as suggested by other studies: bankrupt firms tend to be smaller, slightly less profitable and liquid (indicated by the medians rather than the means) and have a lower market-to-book value. By contrast, the means of the language variables *NEGATIVE* and *LITIGIOUS* show significant differences.

Insignificant differences in means regarding all control variables provide assurance that the propensity score matching has been applied properly. High *t*-values, indicating the significant different means of both language variables at the 1 percent level, already suggest the important role that textual analysis can play when it comes to predicting bankruptcy. Finally, the small correlation coefficients and VIFs below 10 as suggested by Neter, Wasserman, and Kutner (1983) are demonstrated by Table 5 for all employed variables and are a strong statement against collinearity concerns.

To examine whether corporate language changes when approaching bankruptcy, we divide all 1,826 bankrupt firm-year observations into subsamples according to the time difference between filing for bankruptcy and the actual filing of the 10-K (e.g., if the report was filed in the same year as the bankruptcy, the difference in years equals zero). We again match them on the nearest propensity

6. The small number of bankruptcies in 2015 is the result of data constraints. For this study, data were only available until the end of February 2015.

FIGURE 1 Bankruptcies of sample firms by year



Notes:

Our final sample consists of 424 unique bankruptcy cases. The most bankruptcies were filed in 2001, the fewest in 1994.

score. Because replacement is allowed, less than 5 percent of control firms are closest neighbors for more than one bankrupt firm.

Language in 10-K Filings

As we examine whether the language used in 10-K filings contains explanatory power beyond financial ratios with respect to bankruptcies, our dependent variable is the binary variable *BANKRUPT* which identifies bankrupt companies. The results based on equation (2) are provided in Table 6.

The first column of Table 6 shows the results for all 1,826 firm-year observations matched to their nearest propensity score neighbor, without any classification into subsamples. The second column includes 215 10-Ks that were filed in the same year as the bankruptcy. One notable result is that *NEGATIVE* is significant across the first five samples. This strongly supports Hypothesis 1 because it indicates that failing firms use more negative language in 10-Ks than non-failing firms for filings published up to a maximum of 36 months before the bankruptcy filing. After that, no effect can be detected. This is at the lowest edge of what previous findings from traditional ratio analyses suggest, since scholars have stated that a firm failure can be predicted based on data from three to six years prior to bankruptcy (e.g., Altman et al., 1977; Beaver, 1966). Mayew et al. (2015) report the same time period, however. As the coefficients are positive, we conclude that the more *NEGATIVE* the language in a 10-K (the higher the number of negative words minus positive words divided by the total number of words), the higher the probability of a bankruptcy. This finding indicates that companies adjust their language according to the firm's condition. As assumed, except for *PROFITABILITY* across all firm-year observations (since we propensity-match on several variables, *PROFITABILITY* is overruled and shows significant differences in that particular sample), the whole set of control variables is insignificant. This proves that the propensity score matching has worked well.

TABLE 4

Summary statistics of bankrupt and control sample's firm-year observations

	Bankrupt firm sample							Control sample							Test statistics	
	N = 1,826							N = 1,826								
Variable	Mean	SD	Min.	P25	Med.	P75	Max.	Mean	SD	Min.	P25	Med.	P75	Max.	t-value	*p*-value
Propensity score variables																
PROFITABILITY	0.00	0.18	−3.68	−0.03	0.04	0.08	0.37	−0.01	0.42	−6.08	0.00	0.07	0.12	0.37	−1.24	0.21
LIQUIDITY	1.99	2.12	0.04	1.06	1.53	2.22	25.24	1.99	1.74	0.01	1.07	1.59	2.36	25.24	0.00	1.00
SIZE	5.49	1.57	−1.07	4.53	5.53	6.44	10.56	5.44	2.46	−1.07	3.63	5.56	7.11	10.56	−0.76	0.44
CASH	0.10	0.14	0.00	0.01	0.04	0.13	0.91	0.10	0.14	0.00	0.01	0.04	0.12	0.97	−0.89	0.38
MARKET	1.87	6.83	−50.52	0.50	1.22	2.46	59.26	1.97	8.22	−50.52	1.02	1.84	3.30	59.26	0.39	0.70
Language variables																
NEGATIVE	0.95	0.47	−0.27	0.66	0.96	1.26	2.05	0.77	0.48	−0.27	0.46	0.76	1.10	2.05	−11.00***	0.00***
LITIGIOUS	2.08	1.09	0.40	1.13	1.88	2.96	4.27	1.85	1.06	0.40	0.94	1.62	2.67	4.27	−6.41***	0.00***

Notes:

This table shows the descriptive statistics for all bankrupt and non-bankrupt firm-year observations of the sample. The propensity score variables' means are statistically not significantly different. By contrast, language variables are highly significant. Thus, these results provide assurance that the propensity score matching has worked well. *** represents significance level of 1 percent. All variables are winsorized to the first and 99th percentile.

TABLE 5
Correlation matrix and variance inflation factors (VIFs) for the complete sample

Variable	VIF	1	2	3	4	5	6	7
1	7.08	1.00						
2	7.43	0.42	1.00					
3	1.22	−0.13	0.05	1.00				
4	3.15	−0.07	−0.06	0.04	1.00			
5	9.63	−0.07	0.08	0.28	0.03	1.00		
6	2.65	0.08	−0.09	−0.15	0.55	0.05	1.00	
7	1.13	−0.03	0.01	0.14	0.10	0.12	0.11	1.00

Notes:

The table shows correlation coefficients and variance inflation factors for all variables used in the models as defined in equation (2). 1: *NEGATIVE,* 2: *LITIGIOUS,* 3: *PROFITABILITY,* 4: *LIQUIDITY,* 5: *SIZE,* 6: *CASH,* 7: *MARKET.*

Generally, we find a strong positive correlation between firm failure and the use of negative words in 10-K reports. Furthermore, the right-hand side of Table 6 indicates the predictive ability of negative language. Instead of incorporating the language of the 10-K in the year of bankruptcy, we include *LAG_NEGATIVE* and therefore the language used one year earlier. A highly significant value at the 1 percent level shows that the language of the 10-K one year before bankruptcy already contains information beyond recent financial data. Although this study is designed to explain bankruptcy, this hints at the predictive power of language with respect to bankruptcies, even though we cannot calculate confidence intervals as in Ben-David, Graham, and Harvey (2013) with this lagged approach and logit methodology to evaluate how language affects the confidence intervals beyond increasing R^2.

LITIGIOUS shows explanatory power for 10-Ks filed in the same year as the bankruptcy, which comprise sample (2). The results indicate that the management of firms about to fail already consider the legal issues associated with a bankruptcy during the short time span before the bankruptcy and disclose them to shareholders via the 10-K. A highly significant coefficient for sample (6) which contains firm-year observations of around four years between a bankruptcy filing and filing of a 10-K, shown in Table 6, is indicative of another intriguing factor. Since most court cases are time-consuming and expensive, they have the capacity to precipitate a bankruptcy. Firms that are involved in litigation have to form provisions and will sustain major legal expenses for a few years. They also have to develop a public relations strategy to protect their reputation. Based on the results, it can be assumed that around four years later they have used up all their reserves, have been fined a significant amount, or lost too much of their reputation to retain customers and keep the business running. In the end they are forced to file for bankruptcy.

TABLE 6

Results of the logit estimation of bankruptcy probability

Sample DV = *BANKRUPT*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Forecast DV = *BANKRUPT*	
Years to bankruptcy	0–6	0	1	2	3	4	5	6	Years to bankruptcy	0
NEGATIVE	0.804***	1.690***	2.031***	1.061***	0.450**	–0.047	0.098	–0.196	*LAG_NEGATIVE*	0.621**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.048)	(0.862)	(0.740)	(0.686)		(0.026)
LITIGIOUS	0.041	0.268**	0.056	–0.069	0.037	0.408***	0.047	0.060	*LITIGIOUS*	0.525***
	(0.397)	(0.043)	(0.569)	(0.484)	(0.716)	(0.001)	(0.710)	(0.760)		(0.000)
PROFITABILITY	0.293**	0.174	0.375	0.379	–0.444	–0.027	0.123	–1.160	*PROFITABILITY*	0.153
	(0.039)	(0.591)	(0.300)	(0.148)	(0.514)	(0.957)	(0.706)	(0.282)		(0.670)
LIQUIDITY	0.012	0.112	0.007	0.017	–0.001	–0.051	0.019	–0.020	*LIQUIDITY*	–0.015
	(0.692)	(0.364)	(0.914)	(0.708)	(0.982)	(0.341)	(0.733)	(0.779)		(0.907)
SIZE	0.008	–0.046	–0.026	–0.002	–0.005	0.048	0.033	0.011	*SIZE*	–0.113**
	(0.773)	(0.399)	(0.535)	(0.962)	(0.916)	(0.417)	(0.582)	(0.918)		(0.037)
CASH	–0.050	–0.318	–0.375	0.028	–0.310	1.251	–0.132	–0.857	*CASH*	1.192
	(0.909)	(0.784)	(0.654)	(0.969)	(0.697)	(0.160)	(0.876)	(0.653)		(0.320)
MARKET	–0.003	0.018	–0.001	–0.001	0.001	0.002	–0.002	–0.028	*MARKET*	0.015
	(0.502)	(0.223)	(0.953)	(0.933)	(0.920)	(0.860)	(0.900)	(0.456)		(0.337)

(The table is continued on the next page.)

TABLE 6 (contiuned)

Sample DV = *BANKRUPT*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Forecast DV = *BANKRUPT*	
Years to bankruptcy	0–6	0	1	2	3	4	5	6	Years to bankruptcy	0
Constant	−1.389**	−2.603***	−2.082***	−0.945*	−0.356	−1.609	−1.071	0.089	Constant	−1.272
	(0.012)	(0.001)	(0.001)	(0.099)	(0.812)	(0.249)	(0.445)	(0.959)		(0.113)
Observations	3,652	428	773	723	616	514	424	165	Observations	384
Pseudo R^2	0.030	0.127	0.122	0.050	0.027	0.053	0.041	0.052	Pseudo R^2	0.083
Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Time	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Industry	Yes
Model	Logit	Logit	Logit	Logit	Logit	Logit	Logit	Logit	Model	Logit

Notes:

This table shows logit language models as described in equation (2): $BANKRUPT_{it} = \alpha + \beta_1 NEGATIVE_{it} + \beta_2 LITIGIOUS_{it} + \beta_3 PROFITABILITY_{it} + \beta_4 LIQUIDITY_{it} + \beta_5 SIZE + \beta_6 CASH_{it} + \beta_7 MARKET_{it} + \sum TIME_t + \sum INDUSTRY_i + \varepsilon_{it}$. Control variables for models (1) to (8) include profitability, liquidity, size, cash position, and a market variable. The subsamples are described in more detail in Table 2. Furthermore, time and industry dummies are included. Winsorized data as well as clustered and heteroscedasticity-robust standard errors are applied. The forecast is calculated on basis of equation (3): $BANKRUPT_{it} = \alpha + \beta_1 NEGATIVE_{it-1} + \beta_2 LITIGIOUS_{it} + \beta_3 PROFITABILITY_{it} + \beta_4 LIQUIDITY_{it} + \beta_5 SIZE + \beta_6 CASH_{it} + \beta_7 MARKET_{it} + \sum TIME_t + \sum INDUSTRY_i + \varepsilon_{it}$. *, **, and *** represent significance levels of 10 percent, 5 percent, and 1 percent, respectively.

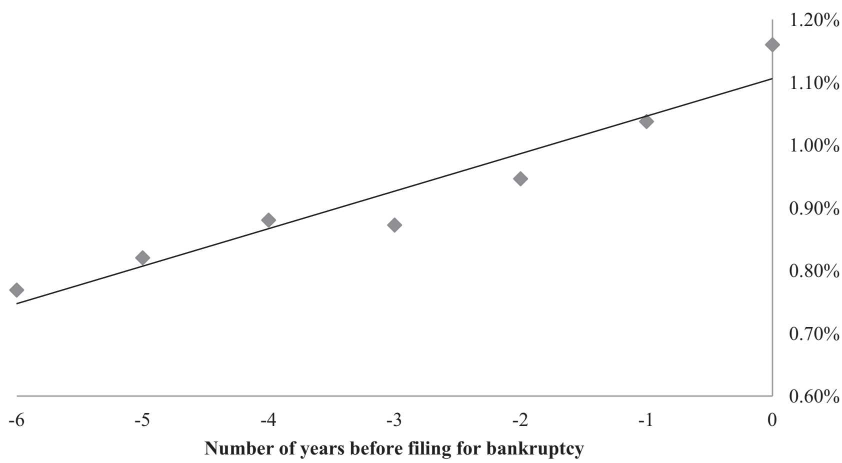
In summary, this study shows that litigious language contains explanatory power in the very short run in addition to *NEGATIVE* and beyond financial ratios, which supports Hypothesis 2. Additionally, we find evidence that this relationship holds also for a longer time frame. Furthermore, the Pseudo-R^2 increases for samples (2) and (3), which reflect 10-Ks filed either one year before bankruptcy or in the same year, compared to the propensity score matching model. The values between 12 and 13 percent are quite close to the results of Mayew et al. (2015), who report a Pseudo-R^2 of 14.58 percent for their language models. They also report an incremental predictive ability of language used in the MD&A section of around 5 percent, close to the roughly 3 percent in our study. This value is calculated as the model fit of sample (2) which includes all control and language variables minus the model fit of the propensity score matching model. Hence, by including language variables, the model fit measured as Pseudo-R^2 increases from 0.096 to 0.127.

Figure 2 finally provides a sense of the distribution of *NEGATIVE* in 10-Ks over the full sample, classifying the means of *NEGATIVE* according to the difference in years between bankruptcy filing and 10-K filing. The relationship is exceptionally stringent: the closer a firm is to bankruptcy, the higher the number of negative words in its 10-Ks.

Robustness Tests

To verify whether the results hold true when the parameters of the model are changed, we conduct several robustness tests. Because matching is a critical factor

FIGURE 2 The evolution of *NEGATIVE* before filing for bankruptcy



Notes:

This figure shows the evolution of *NEGATIVE* before filing for bankruptcy for the full sample. The values are calculated as the means over the complete sample. The trend line indicates a stable upward trend: the closer the bankruptcy, the more negative the terminology.

when it comes to logit regressions, all calculations are repeated on differently matched samples.

First, we use propensity score matching with the same propensity-match variables but no longer allow replacement. In other words, one unique control firm observation cannot match more than one bankrupt firm-year observation. This does not change the results with regard to its content. Second, we create a two-step match approach. In the first step, matching is conducted on the same firm-year and industry. In a second step, the nearest propensity score neighbor within the same year and industry is chosen. This is again conducted with replacement. *NEGATIVE* is then highly significant up to six years prior to bankruptcy. *LITIGIOUS* shows no explanatory power at all. On top of that, several control variables show significance in some samples, which is evidence of greater differences between bankrupt and non-bankrupt firms with respect to those propensity matching ratios. This is a result of the altered matching procedure and indicates why traditional ratio analyses may have found discriminative power of their ratios from data longer before the bankruptcy filing than ours. For this reason, scholars have to make a choice when conducting future research. To answer the research question, is it more important to match on firm fundamentals and ratios, or on industry and year? We strongly argue for the former to obtain more accurate and less biased results. In our study, the effect of *NEGATIVE* on bankruptcies is strong for both matching routines.

Third, we exclude the firm-years of the two major economic crises in 2000–2002 (the dotcom collapse) and 2008–2009 (the financial crisis) from the sample to ensure the results are not patterns evolving from extreme situations but rather applicable to all periods. Again, no significantly different results for *NEGATIVE* emerge even though the sample size decreases by almost 30 percent from 93,134 to 68,232 firm-year observations. By contrast, excluding the years of the financial crisis, *LITIGIOUS* no longer shows any explanatory power for bankruptcies. Finally, financial institutions (all firm-year observations with the Fama-French 12 industry code 11) are excluded from the sample. Only 3,568 out of 93,134 firm-year observations are dropped. Again, the overall results are robust to this adjustment and do not change, except for *LITIGIOUS*: the exclusion of financial institutions yields additional significant results at the 1 percent level for the complete sample and five years before bankruptcy. By contrast, the explanatory power of *LITIGIOUS* in the very short run diminishes.

Moreover, we evaluate whether the results change when applying Shumway's (2001) hazard methodology as described above. The sample size does not alter due to this change. Regarding *NEGATIVE*, no difference can be detected. The coefficient is still positive and highly significant at the 1 percent level. The forecast *LAG_NEGATIVE* also remains highly significant. By contrast, *LITIGIOUS* is not robust to this alteration: the statistical significance and therefore the explanatory power disappears under hazard assumptions. All regression tables regarding the robustness tests are available upon request.

CONCLUSION

This study shows that failing firms' 10-Ks contain significantly more negative language than those of healthy firms. This effect is strong up to three years prior to the actual bankruptcy filing and does not change due to changes in the matching routines. The results are robust to the exclusion of both financial institutions and of periods of global economic crisis. Using a lagged variable, we can unequivocally confirm that negative language has predictive power with respect to bankruptcies. In times of financial hardship, companies make reference to their problems in their 10-Ks and accordingly use more negative words. Some research suggested that the management strategically manages tone (e.g., Huang, Teoh et al., 2014), while other studies emphasize the capacity of language to reveal true firm performance (e.g., Huang, Zang, and Zheng, 2014; Huang, Teoh et al., 2014; Li, 2008). Our study supports the latter, also rebutting the prejudice arising from general agency theory that management keeps the problems a secret from their shareholders (Grossman and Hart, 1983; Jensen and Meckling, 1976).

Moreover, 10-K reports by defaulting firms contain significantly more litigious words than reports by healthy firms immediately prior to bankruptcy. Additionally, we show that prospectively bankrupt firms use significantly more litigious terms around four years before filing for bankruptcy, whereas the effect diminishes between one and three years before bankruptcy. This pattern can be explained by a limitation of the study: at this point we cannot distinguish whether a company is sued, is suing or is engaged in some other legal action—we only provide an overall view on litigious language, so the effect is rather unstable. For example, one can imagine that a company which is sued for antitrust violations is more likely to enter a precarious financial situation than a company that sues. Both effects can offset each other in the data. This is a promising follow-up research question: to gather more detailed litigation data and better explain the effect of litigious language.

Another possible explanation for the instability of the results regarding *LITIGIOUS* is the general time frame of court cases, since it can take several years to obtain a final ruling. In particular, legal proceedings that determine the future of a firm are very likely to drag on since all parties are likely to appeal and fight. The reason for a firm failure can therefore be attributed partly to legal proceedings that eat into a firm's financial reserves, result in significant court-imposed fines, and oblige the company to form provisions for legal expenses instead of investing in operations. Hence, from the results of this study, we conclude that new legal proceedings are likely to be discussed in more detail in the 10-K than an ongoing lawsuit. This should be examined in further studies that focus more on the use of litigious terms. Another possible explanation is that a financial crisis that occurs between one and four years prior to bankruptcy leads to a shift toward more negative words. In such a crisis a firm is likely to put a higher focus on the negative macroeconomic situation than on a lawsuit, however threatening it may be. This holds for the authors of a 10-K, too. We warmly encourage further research on this aspect.

By matching on financial ratios, we show that language provides more information beyond the financials of a firm. Both analyses can be combined to form a more comprehensive model. Nonetheless, language evaluation in itself offers an interesting new perspective. Since it is an easy and cost-effective way of discriminating between bankrupt and non-bankrupt firms, it can be used when information about financials cannot be easily obtained or if there is a lack of confidence in the disclosed financials. For instance, if a stakeholder of a company suspects aggressive earnings management or hidden risks in the firm's balance sheet or income statement, they could use textual analysis to either confirm or disprove their theory. Therefore, the topic is highly important for regulators as well.

Various potential areas of application for language evaluation are imaginable. For instance, it could serve as a cost-effective indicator of general firm performance, earnings management, or even fraud. KPMG and PwC as major audit firms already pursue this approach and have confirmed the relevance of language analysis in practice. Potential other users include banks, financial analysts, and rating agencies, since—unlike traditional ratio analyses—the evaluation of qualitative characteristics not only provides uncommented numbers and ratios, it is also capable of creating a clearer picture of the characteristics and traits of companies and their representatives. For some stakeholders it can be equally important to get a sense of these factors and of a firm's culture than to focus exclusively on numbers. This study can be extended to incorporate various languages. It would be of great interest to compare European or Asian samples to North American firms and evaluate whether corporate terminology differs between continents and cultures. Finally, in the longer run, a set of qualitative characteristics composed of language and other variables can be used alongside financial statement analysis to evaluate and forecast firm performances. Our results prove that language contains much inherent information so we strongly advocate exploring these avenues. This would produce a more comprehensive insight into business valuation. As these qualitative characteristics are either in the process of being developed or already applicable, omitting them deliberately would be wrong and careless.

REFERENCES

Aharony, J., C. P. Jones, and I. Swary. 1980. An analysis of risk and return characteristics of corporate bankruptcy using capital market data. *Journal of Finance* 35 (4): 1001–16.

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23 (4): 589–609.

Altman, E. I., and T. P. McGough. 1974. Evaluation of a company as a going concern. *Journal of Accountancy* 138 (6): 50–57.

Altman, E. I., R. G. Haldeman, and P. Narayanan. 1977. ZETA™ analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance* 1 (1): 29–54.

Beattie, V., B. McInnes, and S. Fearnley. 2004. A methodology for analysing and evaluating narratives in annual reports: A comprehensive descriptive profile and metrics for disclosure quality attributes. *Accounting Forum* 28 (3): 205–36.

Beaver, W. H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4 (1): 71–111.

Beaver, W. H., M. F. McNichols, and J. Rhie. 2005. Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies* 10 (1): 93–122.

Ben-David, I., J. R. Graham, and C. R. Harvey. 2013. Managerial miscalibration. *Quarterly Journal of Economics* 128 (4): 1547–84.

Bickhoff, N., M. Blatz, G. Eilenberger, S. Haghani, and K. Kraus. 2004. *Die Unternehmenskrise als Chance: Innovative Ansätze zur Sanierung und Restrukturierung*. Berlin, Heidelberg: Springer.

Blum, M. 1974. Failing company discriminant analysis. *Journal of Accounting Research* 12 (1): 1–25.

Boritz, J. E. 1991. Going...going...gone!. *CA Magazine* 124 (4): 22–29.

Branch, B. 2002. The costs of bankruptcy: A review. *International Review of Financial Analysis* 11 (1): 39–57.

Bris, A., I. Welch, and N. Zhu. 2006. The costs of bankruptcy: Chapter 7 liquidation versus chapter 11 reorganization. *Journal of Finance* 61 (3): 1253–303.

Cao, Z., and G. S. Narayanamoorthy. 2014. Accounting and litigation risk: Evidence from directors' and officers' insurance pricing. *Review of Accounting Studies* 19 (1): 1–42.

Chen, Y. 2014. Detecting fraud in narrative annual reports. Unpublished manuscript. http://ssrn.com/abstract=2511629, retrieved October 18, 2014.

Cohen, J., Y. Ding, C. Lesage, and H. Stolowy. 2010. Corporate fraud and managers' behavior: Evidence from the Press. *Journal of Business Ethics* 95 (S2): 271–315.

Cormier, D., M. Magnan, and B. Morard. 1995. The auditor's consideration of the going concern assumption: A diagnostic model. *Journal of Accounting, Auditing & Finance* 10 (2): 201–22.

Correia, M., S. Richardson, and İ. Tuna. 2012. Value investing in credit markets. *Review of Accounting Studies* 17 (3): 572–609.

Cramer, J. S. 2002. The origins of logistic regression. Unpublished manuscript. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=360300, retrieved November 01, 2002.

Darrat, A. F., S. Gray, J. C. Park, and Y. Wu. 2014. Corporate governance and bankruptcy risk. *Journal of Accounting, Auditing & Finance* 31 (2): 1–40.

Davis, A. K., J. M. Piger, and L. M. Sedor. 2012. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research* 29 (3): 845–68.

Dehaan, E., F. Hodge, and T. Shevlin. 2013. Does voluntary adoption of a clawback provision improve financial reporting quality? *Contemporary Accounting Research* 30 (3): 1027–62.

Eisenbeis, R. A. 1977. Pitfalls in the application of discriminant analysis in business, finance, and economics. *Journal of Finance* 32 (3): 875–900.

Elliott, W. B., K. M. Rennekamp, and B. J. White. 2015. Does concrete language in disclosures increase willingness to invest? *Review of Accounting Studies* 20 (2): 839–65.

Frino, A., S. Jones, and J. B. Wong. 2007. Market behavior around bankruptcy announcements: Evidence from the Australian stock exchange. *Accounting and Finance* 47 (1): 713–30.

García, D. 2013. Sentiment during recessions. *Journal of Finance* 68 (3): 1267–300.

Gaston, N. 1997. Efficiency wages, managerial discretion, and the fear of bankruptcy. *Journal of Economic Behavior & Organization* 33 (1): 41–59.

Goel, S., J. Gangolly, S. R. Faerman, and O. Uzuner. 2010. Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting* 7 (1): 25–46.

Griffin, J. M., and M. L. Lemmon. 2002. Book-to-market equity, distress risk, and stock returns. *Journal of Finance* 57 (5): 2317–36.

Grossman, S. J., and O. D. Hart. 1983. An analysis of the principal-agent problem. *Econometrica* 51 (1): 7–45.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62 (2): 1–36.

Heckman, J. T. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153–61.

Henselmann, K., and E. Scherr. 2012. Content analysis of XBRL filings as an efficient supplement of bankruptcy prediction? Empirical evidence based on US GAAP annual reports. Unpublished manuscript. https://www.econstor.eu/dspace/bitstream/10419/58246/1/716238845.pdf, retrieved November 06, 2015.

Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9 (1): 5–34.

Holder-Webb, L., and J. Cohen. 2007. The association between disclosure, distress, and failure. *Journal of Business Ethics* 75 (3): 301–14.

Huang, X., S. H. Teoh, and Y. Zhang. 2014. Tone management. *The Accounting Review* 89 (3): 1083–113.

Huang, A. H., A. Y. Zang, and R. Zheng. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review* 89 (6): 2151–80.

Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–60.

Jiang, W., K. Li, and W. Wang. 2012. Hedge funds and chapter 11. *Journal of Finance* 67 (2): 513–60.

Khalili Araghi, M., and S. Makvandi. 2013. Comparing logit, probit and multiple discriminant analysis models in predicting bankruptcy of companies. *Asian Journal of Finance & Accounting* 5 (1): 48–59.

Koh, H. C. 1991. Model predictions and auditor assessments of going concern status. *Accounting and Business Research* 21 (84): 331–8.

Koh, H. C., and L. N. Killough. 1990. The use of multiple discriminant analysis in the assessment of the going-concern status of an audit client. *Journal of Business Finance & Accounting* 17 (2): 179–92.

KPMG. 2014. *Data analytics—How they can help your business*. https://www.kpmg.com/Ca/en/topics/KPMG-Ignite/Documents/S11-Dataanalytics.pdf, retrieved December 2, 2016.

Lennox, C. S. 1999. The accuracy and incremental information content of audit reports in predicting bankruptcy. *Journal of Business Finance & Accounting* 26 (5&6): 757–78.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2): 221–47.

Li, J. 2012. Prediction of corporate bankruptcy from 2008 through 2011. *Journal of Accounting and Finance* 12 (1): 31–41.

Libby, R. 1975. Accounting ratios and the prediction of failure: Some behavioral evidence. *Journal of Accounting Research* 13 (1): 150–61.

LoPucki, L. M. 2015. *UCLA-LoPucki bankruptcy research database user's manual*. Unpublished manuscript. http://lopucki.law.ucla.edu/index.htm, retrieved October 01, 2015.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66 (1): 35–65.

Magnusson, C., A. Arppe, T. Eklund, B. Back, H. Vanharanta, and A. Visa. 2005. The language of quarterly reports as an indicator of change in the company's financial status. *Information & Management* 42 (4): 561–74.

Marino, A. M., and O. Ozbas. 2014. Disclosure of status in an agency setting. *Journal of Economic Behavior & Organization* 105 (1): 191–207.

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): 1621–51.

McKee, T. E. 2003. Rough sets bankruptcy prediction models versus auditor signalling rates. *Journal of Forecasting* 22 (8): 569–86.

Meyer, P. A., and H. W. Pifer. 1970. Prediction of bank failures. *Journal of Finance* 25 (4): 853–68.

Neter, J., W. Wasserman, and M. Kutner. 1983. *Applied Linear Regression Models*. Homewood, AL: Richard Irwin Inc.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18 (1): 109–31.

Peterson, S. J., F. O. Walumbwa, K. Byron, and J. Myrowitz. 2009. CEO positive psychological traits, transformational leadership, and firm performance in high-technology start-up and established firms. *Journal of Management* 35 (2): 348–68.

Purda, L., and D. Skillicorn. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research* 32 (3): 1193–223.

PwC. 2012. *Natural language processing and social media intelligence*. http://www.pwc.com/us/en/technology-forecast/2012/issue1/features/feature-mining-social-media-intelligence.html, retrieved December 2, 2016.

Rogers, J. L., A. van Buskirk, and S. L. C. Zechman. 2011. Disclosure tone and shareholder litigation. *The Accounting Review* 86 (6): 2155–83.

Roncek, D. W. 1991. Using logit coefficients to obtain the effects of independent variables on changes in probabilities. *Social Forces* 70 (2): 509–18.

Schleicher, T., and M. Walker. 2010. Bias in the tone of forward-looking narratives. *Accounting and Business Research* 40 (4): 371–90.

Sheppard, J. P. 1994. Strategy and bankruptcy: An exploration into organizational death. *Journal of Management* 20 (4): 795–883.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74 (1): 101–24.

Tan, H., E. Y. Wang, and B. Zhou. 2014. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research* 52 (1): 273–302.

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62 (3): 1139–68.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63 (3): 1437–67.

Thießen, A. 2013. *Handbuch Krisenmanagement*. Wiesbaden: Springer Fachmedien Wiesbaden.

Tykvová, T., and M. Borell. 2012. Do private equity owners increase risk of financial distress and bankruptcy? *Journal of Corporate Finance* 18 (1): 138–50.

White, M. J. 1989. The corporate bankruptcy decision. *Journal of Economic Perspectives* 3 (2): 129–51.

Wu, Y., C. Gaunt, and S. Gray. 2010. A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics* 6 (1): 34–45.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22 (1): 59–82.