Using 10-K Text to Gauge Financial Constraints

Andriy Bodnaruk
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556-5646
574.631.4597 *voice*
abodnaru@nd.edu

Tim Loughran
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556-5646
574.631.8432 *voice*
Loughran.9@nd.edu

Bill McDonald
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556-5646
574.631.5137 *voice*
mcdonald.1@nd.edu

December 30, 2013

Abstract

Financial constraints — the wedge between the costs of external and internal funds — dictate the relevance of financial structure. We propose a new measure of financial constraints based on qualitative information contained in corporate disclosures. We parse 10-K disclosures filed with the Securities and Exchange Commission (SEC) to measure a document's tone as indicated by the percentage of negative words. We find that the frequency of negative words exhibits very low correlation with traditional measures of financial constraints such as size and predicts subsequent liquidity events — like dividend cuts or omissions, debt downgrades, and asset growth — better than widely-used financial constraint indexes.

Key words: Financial constraints; textual analysis; dividend omissions; debt downgrades.
JEL Classifications: G31, G32, D92.

We thank Matt Cain, Mara Faccio, Paul Gao, Gerard Hoberg, and seminar participants at the University of Mannheim and Emory University for helpful comments.

1. Introduction

Miller (1988), in a retrospective look at the Modigliani-Miller propositions, emphasizes that the complement of "irrelevance" is most important, stating that "showing what doesn't matter can also show, by implication, what does." Thus, as surmised by Hennessy and Whited (2007), the relevance of corporate finance is, to a great extent, determined by financing frictions. The nature and substance of market frictions has been considered at length (see, for example, Battacharya 1979, Townsend 1979, Myers and Majluf 1984). Whether identified market imperfections are of first or second order importance in financial decisions is an empirical question that relies critically on the ability to identify financially constrained firms — firms for which there is a wedge between the internal and external costs of funds.

Numerous methods for measuring financial constraints have been proposed. While most of them assign firms a financial constraint status based purely on a firm's accounting variables, two important papers, Kaplan and Zingales (1997) (hereafter KZ) and Hadlock and Pierce (2010) (hereafter HP) also incorporate textual disclosures in construction of their measures. KZ and HP examine 10-K text to identify cases where difficulties in obtaining external financing, liquidity problems, or forced reduction in investment are discussed and subjectively classify firms by financial constraint status on the basis of the number and severity of the disclosed constraints. They then use accounting characteristics to predict where the firm will fall within their classification. Due to the time intensive nature of their method, their analyses were limited to relatively small samples of firms.

Our paper expands on KZ and HP's approach of using qualitative information to gauge firms' financial constraints. Whereas KZ and HP use qualitative analysis of a firm's disclosures as an intermediate step in deriving accounting based indexes of financial constraints, we use qualitative information to directly construct a measure of financial constraints and use the measure to predict subsequent events traditionally associated with the deterioration of external financing conditions, events which we label "liquidity events." Our four ex post liquidity events include dividend cuts or omissions, debt downgrades, and subsequent growth in total assets. Being financially constrained typically does not have definitive endpoints. Thus, we are simply trying to capture the likelihood of being in a particular state over a reasonable time frame. Obviously, firms that appear to be in, or trending toward, the financial constraint state are more likely to experience subsequent liquidity events.

Our choice of liquidity events is deeply rooted in the financial constraints literature. Starting with Fazzari et al. (1988) and Kaplan and Zingales (1997), the literature argues that firms would pay out dividends only when their internally generated funds exceed their investment needs. Indeed, Campello, Graham, and Harvey (2010), surveying corporate CFOs, found that during the recent financial crisis, constrained firms in the U.S. planned to drastically reduce dividend payments whereas unconstrained firms did not. Similarly, company credit ratings, i.e., "an issuer's ability to meet its financial obligations in full and on time" (S&P, 2013) have been widely used to partition companies into constrained and unconstrained subsamples (e.g., Almeida et al., 2004, Faulkender and Wang, 2006, and Sufi, 2009). Finally, firms with difficulties in accessing external financing would be forced to select only high return investments; this inability to finance all the investment opportunities impedes company asset growth (e.g., Carpenter and Petersen, 2002).

We parse 10-K disclosures filed with the Securities and Exchange Commission (SEC) to measure a document's tone as indicated by the percentage of negative words from the Loughran and McDonald (2011) negative word list.[1] Commonly used negative words include, for example, *loss, limitation, adverse, impairment, failure,* and *default*. We take advantage of a particular feature of SEC disclosure regulation (Regulation S-K) requiring companies to discuss their past operating and financial results as well as to elaborate on any known trends or uncertainties that could materially affect them in the future.[2] Our conjecture is that managers anticipating financial challenges will use a more negative tone in 10-K filings to communicate their concerns to shareholders, thereby lowering their exposure to subsequent litigation. In the context of IPOs, Hanley and Hoberg (2012) find that strong disclosure in the IPO prospectus lowers the probability of being sued.

HP (2010) use a combination of total assets and firm age to measure financial constraints. Whited and Wu (2006) (hereafter WW) create a six component index; two of WW's components, total assets and dividend dummy, are directly linked with larger and older firms. Thus for both indexes, large and old firms have a lower likelihood of being financially constrained. Yet, as the financial crises over the last few decades have shown, even old and large firms can quickly become financially constrained.

As an example, consider Ford Motor Company. As of June 2001, Ford had enormous total assets (over $280 billion), positive prior annual stock returns in a down market, and was relatively old. As a result, it had extremely low values for some of the traditionally used indexes of financial constraints (i.e., the firm was not financially constrained). Yet, within 12 months, Ford's S&P debt rating fell from "A" to "BBB+", it cut dividends, and the firm experienced a decline in total assets. Interestingly, Ford's 10-K filed on March 22, 2001 contained 2.27%

[1] We use the 2011 updated version of their word list available at www.nd.edu/~mcdonald/Word_Lists.html.
[2] The guidance on requirements for presenting liquidity and capital disclosures is detailed in SEC release number 33-9144, which emphasizes the importance of discussing forward looking statements about funding and liquidity risks, including those risks not readily deduced from their financial statements.

negative words (which puts it in top 6th percentile of all firms in that year). So although some widely used financial constraint indexes would imply smooth sailing for Ford as of 2001, the high frequency of negative words in the text foreshadowed its uncertain future. Textual analysis, as a variable added to the traditional mix of finance variables that might be used to gauge the level of financial constraints, has the potential to identify inflection points not captured by variables like firm market capitalization.

We show that the tone of 10-K documents is a measure of financial constraints distinct from measures based on accounting characteristics. Though the tone of the document has very low correlations with traditional measures of financial constraints, firms with a higher percentage of negative words in their 10-Ks have higher cash flow sensitivity of cash (Almeida, Campello, and Weisbach, 2004). Further, the percent of negative words, unlike the SA and WW indexes, has a low correlation with market capitalization. Since the SA and WW indexes both use total assets as one of their components, it is unclear whether these indexes contribute beyond what is in a mix of standard control variables.

When we turn our attention to the ability of various measures of financial constraints to predict events related to the deterioration in external financing conditions, we find that a more frequent usage of negative words is strongly related to a higher likelihood of future dividend reductions (+7.1%), omissions (+8.9%), debt downgrades (+10.8%), and to lower asset growth (-5.3%).[3] The results are robust to inclusion of firm characteristics, e.g., market capitalization, book-to-market, and past performance.

In contrast, measures of financial constraints based on accounting characteristics (KZ index, SA index, and WW index) have limited success in predicting adverse liquidity events. All of

---

[3] All economic effects are estimated as marginal differences in the dependent variable (scaled down by the sample mean) related to a one standard deviation increase in the percentage of negative words.

them are statistically related in expected ways to only two (out of four) events without the presence of control variables. Moreover, the predictive power either goes down significantly or the direction of the relation flips once we include controls such as a company's market capitalization.

Besides tabulating the percentage of negative words in the entire 10-K, we also provide evidence that negative words in close proximity to the terms *liquidity* and *capital resources* are linked to some of the liquidity events. The occurrence of negative words occurring near the terms *liquidity* or *capital resources* is related to the likelihood that the firm omits a dividend or has low asset growth. This targeted indicator, however, does not appear to be as robust as scanning the entire disclosure.

Financial constraints can be thought of as a two-tail phenomenon, with some firms facing constraints due to deterioration in their cash flows, while others are unable to finance extraordinary growth. None of our tests directly identify firms that are growing, but at a slower rate than the firm desires, due to the high cost of external capital. That is, we cannot accurately measure how the inability to access reasonably priced external capital constrains a firm's ability to invest in positive NPV projects. Our best proxy for companies foregoing investments is when we examine asset growth.

Our analysis differs from earlier work on the use of qualitative information to gauge financial constraints along four key dimensions. First, our measure of financial constraints – percentage of negative words in the 10-K – is objective. That is, we do not assign the financial constraint scores ourselves, but rely on the output of the pre-specified automated parsing algorithm. Since we use the negative word list of Loughran and McDonald (2011), there is no need to read the 10-K to make subjective decisions on whether a particular sentence hints that a firm might be financially constrained. In this way, our measure is not affected by potential misinterpretations or inconsistencies of the classifier. This procedure also makes our measure easier to replicate.

Second, manual categorization, used in prior research, is extremely time consuming which imposes limits on the sample size of the analyzed firms. KZ had a sample of only 49 low-dividend paying manufacturing firms while HP used a random sample of 356 unique firms (1,848 firm-year observations in total). In contrast, in our analysis we use the entire sample of 10-K filers.

Third, both KZ and HP relied on the notion that disclosure rules force firms to reveal financial constraints, which would require them to be explicit about difficulties in obtaining financing. However, as Fazzari, Hubbard, and Petersen (2000) point out, "Regulation S-K requires the firm to reveal the inability to invest due to financial constraints only when the firm fails to act on a previously announced investment commitment." As we demonstrate, our less restrictive approach of considering a broad range of negative words appears to be better at capturing qualitative information about financial constraints than targeting negative words near Regulation S-K verbiage.

Fourth, our approach is fundamentally different in how we use qualitative information to gauge financial constraints. Prior studies, i.e., KZ and HP, use qualitative information to rank subsamples of firms according to their financial constraints status, with their subsequent measures based on accounting characteristics used to explain these rankings. In contrast, by quantifying the language of 10-Ks and using financial events to identify constrained firms, we treat qualitative information as a measure of financial constraints in its own right.

In a contemporaneous paper complementary to ours, Ball, Hoberg, and Maksimovic (2012) also use textual analysis of 10-Ks to identify financially constrained firms. The most similar construct to ours is their measure of delayed investment where they search for the words like *delay, abandon, eliminate,* or *postpone* within 12 words of investment-type words like *construction* or *expansion*. Unlike our paper which parses the entire 10-K, they focus this word search within the Liquidity and Capitalization Resource Subsection [CAP+LIQ] in the Management Discussion and Analysis (MD&A) section.[4] The authors report that only 5.5% of their sample use delay-type and investment-type words in close proximity to each other.

Due to concerns that firms might specifically avoid using *delay* and related synonyms close to *expansion*, Ball et al. create a delayed investment score to measure how similar the CAP+LIQ subsection of firms which mention postponing projects is to other firms. They use the methodology of Hanley and Hoberg (2010) to gauge similarity of text between firms.

The Ball et al. (2012) paper takes a completely different approach than ours in using 10-K text to identify financially constrained firms. Ball et al. (2012) specifically link words like *delay* and *construction* in a 10-K subsection with being constrained while we attempt to measure the level of constraints by the frequency of negative words within the entire 10-K.[5] We believe the tone of managers' words capture subtle signs that the company will face greater future financial challenges. As shown by numerous papers starting with Antweiler and Frank (2004), Tetlock (2007), and Tetlock, Saar-Tsechansky, and Macskassy (2008), document text often contains important information for investors.

[4] When measuring the tone of 10-K filings, we use the entire document whereas Regulation S-K prescribes that management's view on the company's future should be presented in the MD&A section. We find that in many 10-Ks, the MD&A section is not well defined, which inhibits accurate parsing. Many times the most dour view of liquidity is in fact presented in the risk factors section of the 10-K (see, for example, IBM's 10-K filing of 2011-02-22). Additionally, Loughran and McDonald (2011) show the MD&A section does not produce any more precise tone measures. For completeness, we will focus on words occurring around text prescribed by Regulation S-K later in the paper.

[5] Another point of differentiation is that Ball et al. (2012), beyond measuring constraints, also examine the degree to which the constraint wedge is more binding in debt markets versus equity markets.

The remainder of the paper is organized as follows. Section 2 introduces the data and variables. Section 3 reports empirical results. The impact of negative words around the terms *liquidity* and *capital resources* is examined in Section 4. A brief conclusion follows.

2. Data

2.1. The 10-K sample

We download all 10-K, 10-K405, 10KSB, 10-KSB, and 10KSB40 filings, excluding amended documents, from the SEC's EDGAR website (www.sec.gov) during 1996-2011. Table 1 shows how the original sample of 10-Ks is affected by our data filters. The two data screens having the most impact on the sample are eliminating regulated financial firms and utilities (removing 49,222 observations) and requiring the firms to have a CRSP PERMNO and be ordinary common equity (dropping 70,809 observations). Requiring Compustat information like firm age, sales, non-negative book value of equity, and total assets further reduces the sample by 3,607. The final sample is 51,533 firm-year observations during 1997-2011.

Following the methodology of Fama and French (1992, 1993), we form the sample as of June of year t. That is, each year 1997 to 2011, firms with available Compustat data from the prior fiscal year enter the sample as of the end of June. All four of the liquidity events are examined over the following year (i.e., July of year t to June of year t+1). For firms with available CRSP and Compustat information at the end of June, 1997, we investigate their growth in assets and determine whether there is a dividend cut, dividend omission, or debt downgrade during July 1997 to June 1998. Thus, we only use information available to investors as of the yearly June sample formation date. As an example, for the Eastman Kodak Company on the June 2001 formation date, we use the firm's 10-K filed on March 13, 2001. Since Kodak had lower dividends during July 2001 to June 2002 than during the prior year (i.e., July 2000 to June 2001), the firm has a *Dividend Cut Dummy* value of one for the year 2001.

2.2. Parsing the 10-K filings and deriving measures of financial constraints

We first remove all HTML and ASCII-encoded segments from each filing. Collections of text identified in HTML as tables are removed if their numeric character content is greater than 15%. After removing unambiguous proper nouns, the text is then parsed into a vector of words which are tabulated using the Loughran and McDonald (2011) dictionaries. In the Appendix, we provide a detailed discussion of how 10-Ks are parsed.

We then measure the tone of the document as the percentage of negative words (*% Negative*) from the Loughran and McDonald (2011) negative word list. This is the primary qualitative measure of financial constraints we employ in our paper. Additionally, in the later stages of our analysis we consider a more targeted parsing based on Regulation S-K verbiage, where we use *# Negative/# Liquidity*, which is defined as the number of negative words near the terms *liquidity*, *capital resource*, or *capital resources* divided by the count of how often the firm mentions the terms *liquidity* and *capital resources*.

2.3. Variable definitions and summary statistics

The important goal of our paper is to demonstrate that our textual analysis based measure has incremental explanatory power beyond that of traditional quantitative-based measures of financial constraints. We construct three measures of financial constraints widely used in the literature – the KZ index of Kaplan and Zingales (1997), the SA index of Hadlock and Pierce (2010), and the WW index of Whited and Wu (2006) – which are all based on firms' observable characteristics, and employ them alongside our measure. The variables are described in detail in the Appendix.

As has been debated in the existing literature, and as we will see in our discussion of the extant measures of financial constraints, many of the specific components proposed as proxies for financial constraints have ambiguous interpretations. Following Lamont, Polk, and Saa-Requejo (2001), the KZ index has five different components. According to the KZ index, firms with lower operating income, higher Q values, more leverage, lower dividend payouts, and less cash holdings have higher KZ index values. Higher levels of the KZ index indicate that the firm is more financially constrained. The argument for higher growth opportunities (i.e., Q values) being linked with financial constraints is that companies need to have solid future investment projects to be potentially constrained.

In contrast to KZ, other papers have argued that high cash holdings are an indication that the firm is constrained.[6] Firms being shut out of the debt markets might hoard cash in anticipation of future hardship. Clearly, having large cash holdings (scaled by prior year property, plant, and equipment) could be a sign of weakness, not financial strength.

As noted earlier, the SA index has only two parts (firm age and total assets). Higher SA index values indicate that the firm is more financially constrained. The WW index has six components (cash flow, dividend dummy, leverage, total assets, industry sales growth, and firm sales growth). As with the other measures, higher WW index values imply that a firm is more financially constrained.

We consider four liquidity events related to the deterioration of external financing conditions: (1) dividend cuts, (2) dividend omissions, (3) credit rating downgrades, and (4) low asset growth. Our dependent variables are defined as follows. *Dividend Cut Dummy* takes a value of one if the firm has a lower aggregate dividend (controlling for stock splits) during July of year t to June of

[6] For example, see Harford (1999), Opler, Pinkowitz, Stulz, and Williamson (1999), and Acharya, Davydenko, and Strebulaev (2012).

year t+1 than in the prior year, else zero. *Dividend Omission Dummy* is set to one if the firm completely omits paying a dividend during the following year, else zero. Only firms issuing a dividend in the year before the June formation date are assigned a value for the *Dividend Cut* and *Dividend Omission Dummies.*

Debt Downgrade Dummy is set to one if the firm has a lower S&P rating score in June of year t+1 than the rating as of June of year t, else zero. This variable is created only for firms with available S&P debt ratings in both years. *Change in Assets* is the difference in total assets between years t and t+1 divided by total assets in year t. See the Appendix for more detailed variable descriptions.

In Panel A of Table 2, summary statistics are reported. The first column reports values during the earlier part of the sample (1997-2003); column (2) reports for the latter part of the time period; while the last column includes the entire period. The mean percentage of negative words is higher in the later period (1.72% versus 1.47%) while the KZ, SA, and WW index mean values are all lower in the period from 2004 to 2011, a period containing the most significant economic downturn since the Great Depression. Recall that lower values of the three indexes imply that firms are less financially constrained.

For the SA index (with only age and total assets as its components), this time series improvement in financial constraints is easy to explain. As reported in Gao, Ritter, and Zhu (2013) and Doidge, Karolyi, and Stulz (2013), fewer U.S. firms have been going public in the last decade. With a scarcity of young companies entering the sample pool along with recurring delistings due to mergers or bankruptcies, existing publicly-traded firms are getting larger and older.

Fig. 1 notes the time series pattern for the mean percentage of negative words and the SA index during 1997-2011. Each year, the SA index steadily declines (i.e., firms are less financially constrained) as the remaining firms get older and have more total assets. In contrast, negative word frequencies rise from 1997 to 2004 and then have another spike after the financial crisis of 2008.

After the beginning of financial crisis in the fall of 2008, where even AAA-rated firms like General Electric (GE) had liquidity problems, we would expect that the tone of the 10-Ks would become more pessimistic. GE reported a 21% increase in the percentage of negative words from its February 20, 2008 10-K filing (1.54%) compared to its first filing after the meltdown on February 18, 2009 (1.87%). Interestingly, the firm with the second lowest WW index value over the entire time period is GE in 2009. Thus, for the WW index, GE represented an extremely unconstrained firm. Yet, GE had lost over half of its market capitalization over the prior year, its stock price was less than $12, and the firm actually cut its dividends over the next year.

As shown in Panel A of Table 2, our liquidity events are not a common occurrence. Of the firms issuing dividends, 10.74% cut dividends in the following year while only 3.65% omit dividends. On average, 7.89% of firms with an S&P debt rating have a major downgrade in the year after the June formation date.

Panel B of Table 2 reports correlations between our measure of financial constraints, other financial constraint measures, and key control variables. The panel shows that the correlations between the percentage of negative words and all other variables are quite low (with exception of the negative earnings dummy where it is still below 0.25). For example, the correlation between percent negative and excess prior returns is only 0.01. Thus, negative tone in an annual report does not merely serve as a proxy for poor prior performance. This provides the first indication that our measure captures information beyond quantitative measures of financial constraints. Low correlations are also reported by Ball et al. (2012) between their textually-determined constraint variables and the KZ and WW indexes. Using 10-K text to gauge financial constraints appears to add information beyond simple accounting variables or ratios.

It is also worth noting that some of the accounting based measures of financial constraints — SA index and WW index — exhibit very large, negative correlations with market capitalization (-0.702 and -0.839, respectively). In contrast, *% Negative* has a relatively low correlation with market capitalization (0.071). Due to the inclusion of total assets in both indices, the correlation between the SA index and WW index is very high (0.832). This value is almost identical to correlation of 0.80 reported in HP.[7] Since both the SA and WW indexes have a size variable (total assets) as one of their components, it is unclear whether these indexes add value above and beyond the information contained by market capitalization. We will show that even when market capitalization is not used as a control variable, the SA and WW indexes do not help predict all of the liquidity events.

3. Empirical findings

3.1. Cash flow sensitivity of cash

We begin our analysis by verifying that our measure is related to frictions in the allocation of internal resources. If external financing is frictionless and costless, a firm's cash retention policies should be independent of its cash flows. If, on the other hand, the firm faces constraints on raising external financing, the cash flow sensitivities of cash should be positive. To lend

[7] The negative correlation in Table 2 between the SA index and the KZ index of -0.145 differs from the value of 0.05 reported in HP. This difference appears to be caused by our separate winsorization of each of the five KZ index components at the 5% level. The correlation between the KZ and SA indexes is -0.017 if we do not winsorize the components of the KZ index.

additional support for the percentage of negative words as a measure of financial constraints, we estimate propensities to save cash out of cash flows (Almeida, Campello, and Weisbach, 2004).[8]

For comparison purposes with Almeida et al. (2004), we use bond ratings and total assets as financial constraints criteria. For bond ratings, constrained firms are defined as companies with positive long-term debt in a given year that also have no S&P bond rating. Unconstrained bond rating firms have positive long-term debt and an S&P bond rating on their long-term debt in a given year. Firms with no long-term debt in a given year are excluded from the bond rating analysis. Our procedure differs only slightly from Almeida et al. (2004) in that they place firms in the constrained category only if the firm never had rated public debt during their entire sample period (due to limited Compustat bond ratings availability during the early part of their sample period), while we only require that the firm not be rated in the given year for it to be constrained. For total assets, the bottom three deciles are defined as constrained firms while the top three deciles are defined as unconstrained.

For completeness, we also split firms into categories according to other measures of financial constraints: *KZ index*, *SA index*, *WW index*, and our measure — *% Negative*. For the three indexes and for negative words, the constrained firms are defined to be in the top three deciles of the respective measures, while unconstrained firms are in the bottom three deciles of the individual categories, determined on an annual basis.

The results are reported in Table 3. The dependent variable in all the regressions is the change in cash holdings divided by total assets between the current and prior year. The independent variables are cash flow, Tobin's Q, and size (natural log of total assets). While

[8] There is ongoing debate in the literature on whether cash flow sensitivity of cash is a valid indicator of financial constraints. Whereas Riddick and Whited (2009) argue that it is not, it has been widely used in the literature to study properties of financially constrained firms (e.g., Almeida et al., 2004, Faulkender and Wang, 2006, Sufi, 2009, and Denis and Sibilkov, 2010). In our analysis, we use cash flow sensitivity of cash to provide supportive evidence for the validity of our measure of financial constraints and its fit with prior empirical literature.

Almeida et al. (2004) focus only on manufacturers (SIC codes of 2000 to 3999) during the 1971-2000 time period, our sample includes all operating firms with available Compustat data except for financials and utilities. We also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry.

Consistent with Almeida et al. (2004) and HP (2010), we find that financially constrained firms have larger positive sensitivities of cash to cash flow.[9] This is true regardless of which measure is used to define constraints. Our evidence using the SA index to categorize constrained and unconstrained firms (item 4 in Table 3) is similar to the results of HP (2010, Table 7).

As with the other financial constraint criteria, firms constrained according to our measure (high percentage of negative words) have a positive and significant coefficient on the cash flow variable while firms with low frequencies of negative words (i.e., unconstrained firms) have a smaller coefficient value. It is interesting that the coefficient values on the cash flow variable for constrained firms is quite similar for the KZ, SA, and WW indexes and *% Negative*.

The coefficient on cash flow for high percent negative word firms (0.276) is almost identical to the coefficient values on the constrained firms defined by both the KZ index (0.269) and by the SA index (0.272). These findings suggest that the percentage of negative words in the 10-K could serve as a reasonable proxy for the financial constraints a firm is facing in a given year.

3.2. Tone of 10-Ks and liquidity events: Baseline results

A number of prior studies examine the relation between financial constraints and a firm's capital structure and payout. More constrained firms are found to have high cash holdings, keep higher leverage, and pay lower dividends. The interpretation of these results, however, is often

[9] All of the Cash Flow coefficient pairs for each financial constraint criterion are significantly different at the 1% level except the KZ index, where the difference is significant at the 5% level.

problematic due to endogeneity concerns as financial choices and constraints are determined simultaneously.

Instead, we investigate how well different measures are able to predict future developments associated with the deterioration of external financing conditions. Relating current financial constraints to future adverse events alleviates the endogeneity issues. We are interested in how our measure performs on its own and alongside other measures.

We build on the insights of Cleary (1999) and Whited (2009) who argue that firms facing financial frictions would scale down their committed dividend distributions to shareholders. Additionally, firms with high costs of external financing would also be forced to select only high return investments; this inability to finance all the investment opportunities impedes company asset growth (Carpenter and Petersen, 2002). Finally, companies that could not rely on capital markets to buffer liquidity shocks should be less able to sustain their credit rating (e.g. Whited and Wu, 2006). We therefore investigate how well measures of financial constraints predict future dividend cuts and omissions, credit rating downgrades, and asset growth.

As a first step, how well does the KZ, SA, WW indexes and *% Negative* predict liquidity events without controlling for firm characteristics? Table 4 reports summary results from 16 separate regressions. For each of the four ex post liquidity events, the KZ index, SA index, WW index, and percentage of negative words are independent variables. In each regression, an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies are included. The standard errors in all the regressions are clustered by both year and industry. The regressions for dividend cuts, dividend omissions, and debt downgrades are logits while the regression for change in assets is OLS.

These regressions provide an important perspective for assessing the usefulness of the various indexes because they do not yet include standard control variables which we know will be highly correlated with the index components (e.g., the correlation between log of total assets and log of market capitalization is 0.87).

In Table 4, an "X" represents a coefficient that is both significant at the 1% level and has the expected sign. Generally, all four of the financial constraint measures do a reasonable job of predicting the ex post liquidity events when isolated from the inclusion of standard macro-finance variables. For dividend cuts and omissions, all the variables, except for the KZ index, are significant and have the expected sign. However, even without the presence of firm characteristics, the SA and WW indexes have no link with debt downgrades or changes in assets. Overall, only the percentage of negative words is significantly linked with all four liquidity events.

3.3. Tone of 10-Ks and liquidity events: Controlling for firm characteristics

Will the predictive power of the measures of financial constraints continue to be robust once additional firm level control variables are added to the regressions? Our four additional control variables are (1) natural logarithm of market capitalization (stock price time shares outstanding); (2) natural logarithm of the book-to-market ratio; (3) excess prior year buy-and-hold returns; and (4) a dummy variable equal to one when the prior fiscal year income before extraordinary items is negative, zero otherwise. More detailed variable descriptions are provided in the Appendix. The control variables are generally statistically significant in all the regressions and represent obvious first-order factors an investor should consider when identifying financially constrained firms.

3.3.1. Dividend cuts

In Table 5, the dependent variable is the *Dividend Cut Dummy* (equal to one if the firm lowered its dividend during the following year, else zero). Only firms having at least one dividend distribution during the prior year are included in this table. The first regression includes only the control variables. It is worth noting that company market capitalization and past performance are negatively related—as expected—to dividend cuts: larger and better performing companies are less likely to cut dividends.

As before, an intercept, Fama and French 48-industry dummies, and calendar year dummies are included in all regressions. We also include seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006) in all the remaining regressions. The *z*-statistics are in parentheses with the standard errors clustered by both year and industry.

In the next four columns of Table 5, the KZ index, the SA index, the WW index, and percentage of negative words variables are separately added as an independent variable. Note that there is a slight drop in the sample sizes when the KZ index or the WW index is included in the regressions due to slightly different data requirements.

Only in the last column does the appended variable have the expected sign. The coefficient on *% Negative* (0.234) is statistically significant at the 1% level. This implies that as the percentage of negative words in the 10-K rises, the likelihood of the firm cutting its dividend increases. The marginal effect of the coefficient is 0.0168 while the standard deviation of percent negative is 0.456. Thus, a one standard deviation increase in percentage of negative words is related to a 7.1% (=0.0168*0.456/sample mean of 0.1074) larger likelihood that a firm will cut its dividend in a subsequent year. It is encouraging that the inclusion of control variables does not have a material effect on the predictive power of our measure.

The coefficients on the KZ index and WW index are significant, but have the wrong sign, and the coefficient on the SA index is not statistically significant. The failure of the SA and WW indexes to significantly predict the dividend cut dummy variable in the expected direction is completely due to the presence of market capitalization (which is highly correlated with the components of their indexes). Without market value as an independent variable, the coefficient on SA index is 0.275 (z-statistic of 5.06) while the coefficient on the WW index is 3.804 (z-statistic of 6.79).

3.3.2. Dividend omissions

In Table 6, the dependent variable is the *Dividend Omission Dummy*. As before, only firms with at least one dividend distribution in the prior year are included in the five regressions. The control variables imply that larger and better performing companies are less likely to stop paying dividends, whereas companies with negative earnings are more likely to do so.

Of the four financial constraint measures, only the percentage of negative words has both a positive (0.417) and a statistically significant coefficient value. More negative words in a 10-K are linked with a higher likelihood of omitting dividends in the year following the 10-K filing. The marginal effect of the coefficient is 0.0071 while the standard deviation of percent negative is 0.456. Thus, a one standard deviation increase in percentage of negative words is associated with an 8.9% (=0.0071*0.456/sample mean of 0.0365) higher chance of dividend omission.

As expected, the coefficient on percent negative words is larger (with higher z-statistics) when the dependent variable is a *Dividend Omission Dummy* than when it is a *Dividend Cut Dummy*. From a financial constraint viewpoint, firms dropping dividends to zero are in much worse financial condition than companies merely decreasing their dividend payout. We find that firms subsequently omitting dividends have lower average prior year returns (-18.57%

versus -9.37%) and smaller average market values ($1.8 billion versus $4.2 billion) than companies decreasing, but still paying, dividends.

The other measures of financial constraints are either insignificant or are related to dividend omission with the wrong sign. As noted earlier, the presence of market capitalization as a control variable limits the explanatory power of the SA and WW indexes. If market value is dropped as an independent variable, the coefficient on SA index is 0.286 (z-statistic of 3.17) while the coefficient on the WW index is 2.503 (z-statistic of 5.06).

3.3.3. Debt downgrades

We now consider credit rating downgrades. Standard & Poor's downgrades company debt when it believes that there is deterioration in the firm's ability to meet its financial obligations. The first step in determining a debt downgrade is to convert the S&P long-term domestic credit rating into the assigned rating score (range: 1 to 7) of Ashbaugh-Skaife et al. (2006). Next, we compare the assigned credit rating in June of year t to the rating in June of year t+1. As an example, since Xerox Corp. had an S&P debt rating of "A" in June of 2000 and a rating of "BBB-" in June, 2001, the firm dropped from an assigned score of 5 to 4. Hence, Xerox is assigned a value of one for the *Debt Downgrade Dummy* in year 2000.

In Table 7, the coefficient values for the control variables indicate that firms whose debt is downgraded tend to be value firms with poor returns and negative trailing earnings. As might be expected, company market capitalization is negatively related to downgrades. In column (2), the coefficient on the KZ index is positive and significant. More financially constrained firms, as measured by the KZ index, have a higher likelihood of having their debt downgraded. While the SA and WW indexes are not significant or have an incorrect coefficient sign, the percent negative words variable continues to be positively and significantly linked with the dependent variable. The marginal effect for the coefficient on percent negative is 0.0200. Controlling for firm characteristics, a one standard deviation increase in percentage negative words in 10-K is related to a 10.8% larger likelihood of rating downgrade (relative to the sample mean).

3.3.4. Change in total assets

One would expect that, after controlling for other effects, financially constrained firms should have smaller growth rates in total assets than unconstrained companies. Firms with limited ability to attract funds to invest in new projects or acquisitions should have smaller changes in assets than companies who are actively growing their operations. Our last liquidity event is the change in total assets between years t and t+1. Since we require firms to have total assets in the following year, our sample size is slightly reduced.

In the Table 8 regressions, change in total assets is the dependent variable. For this liquidity event, the financial constraint measures should have a negative coefficient value. That is, more constrained firms should have lower asset growth rates.

As has generally been the case for the other liquidity events in the presence of control variables, the KZ, SA, and WW indexes are not significant or have the wrong coefficient sign for asset growth. If market capitalization is not included in the regressions, the SA and WW indexes continue to have incorrect coefficient signs. However, the coefficient on percent negative, in column (5), is statistically significant and has the anticipated sign. More negative words in the 10-K are linked with lower subsequent asset growth by the firm. In terms of the economic significance of percent negative words, a one standard deviation increase in percent negative is related to a 0.70% lower growth of assets. Given that the average change in assets from year t to year t+1 in our sample is 13.18%, this represents a 5.3% decline relative to the sample mean.

To examine firms that potentially would like to grow at a higher rate but cannot because they are financially constrained, the last column of Table 8 reports regression results when firms in the bottom and top quarter of asset growth are removed from the sample. We drop companies in the bottom quarter of asset growth (determined on a yearly basis) because these firms might be more distressed than merely constrained. Since firms in the top quarter of asset growth are successful companies which often acquire other firms, issue debt/equity, and/or experience enormous profits, they would typically not be considered financially constrained. As an example, even though Google is a large capitalization firm, it ranks in the top quarter of asset growth every year during its existence during our sample period. In the last column of Table 8, the coefficient on percent negative (-0.007) is statistically significant (t-statistic of -4.37). Thus, for the middle 50% of the sample, the higher the percentage of negative words, the lower is the firm's asset growth.

4. Negative words around the terms "liquidity" and "capital resources"

So far, our analysis has focused only on the percentage of negative words throughout the entire 10-K as a proxy for the level of financial constraints faced by the firm in a given year. Expanding on the analysis of Ball, Hoberg, and Maksimovic (2012), we now report the frequency of negative words occurring in close proximity to the terms *liquidity* and *capital resources*.[10] Specifically, we parse the entire document and count how many Loughran and McDonald (2011) negative words appear within the range of 100 characters before or 300 characters after the terms *liquidity*, *capital resource,* or *capital resources*. These are common terms used in the discussion of financial constraints required by Regulation S-K.

[10] As noted earlier, Ball et al. do not parse the 10-K for negative words. Instead, they gauge text for levels of financial constraints proxied by investment delays or firms intending to issue equity to fund R&D projects.

Thus, the variable *# Negative/# Liquidity* is defined as the number of negative words near the liquidity terms divided by how often the firm mentions *liquidity* or *capital resources*. For the entire sample, the *# Negative/# Liquidity* variable has a mean of 0.77 and a median of 0.56. Its 90^{th} percentile is 2, which implies two negative words are used in close proximity to the term *liquidity* or *capital resource(s).*

Certainly if a firm uses the terms *adverse* or *failure* a few words from the term *liquidity* in the 10-K, the managers are highlighting their potential constraints. Not surprisingly given its toxic signal, firms infrequently use negative words around the terms *liquidity* and *capital resources*. In fact, 31% of our sample never uses a negative word in close proximity to the liquidity terms.

We expect that firms using more negative words around liquidity terms should be more financially constrained. In Table 9, we report the regression results when *# Negative/# Liquidity* is used as an explanatory variable. Each of the four regressions has a different liquidity event as its dependent variable. The *# Negative/# Liquidity* variable is significantly related (with the expected sign) to dividend omissions and changes in assets. In particular, a one standard deviation increase in the number of negative words around terms "liquidity" and "capital resources" is related to a 6.6% larger likelihood of a firm missing a dividend payment (marginal effect of 0.0031*standard deviation of 0.773/sample mean of 0.0365) and is associated with 5.2% lower asset growth over the subsequent year (all estimated relative to the corresponding sample means). The relations between *# Negative/# Liquidity* and the remaining two liquidity events are neither economically nor statistically significant.

The results for this alternative textual analysis based measure of financial constraints provide additional evidence that negative words are related to future deterioration of external financing. However, they also demonstrate that targeted indicators are potentially less robust and have lower explanatory power than measures based on a broad analysis of the disclosure text. This is consistent with the Fazzari et al. (2000) observation that since regulation does not require firms to be specific about the nature of their current or potential financial difficulties (unless they renege on a previously announced investment commitment), targeted analysis of disclosures could yield poor financial constraint classifications. Our less restrictive approach of considering the entire 10-K document appears to be better at capturing qualitative information about financial constraints.

5. Conclusions

We extend the earlier work of Kaplan and Zingales (1997) and Hadlock and Pierce (2010) who use 10-K text to segment firms by their financial constraint status. KZ and HP then use variables like cash holdings, firm age, or total assets to explain where firms lie across their financial constraint classifications.

This paper differs by using the tone of 10-K text, and not solely accounting characteristics, to help gauge which firms will become financially constrained. Unlike other measures, our variable, the frequency of negative words from the Loughran and McDonald (2011) word list, is more likely to identify inflection points. For example, everyone agrees that as firms get larger, they typically become less financially constrained. Our variable, percentage of negative 10-K words, helps indicate when a large company might suddenly slip into the realm of being financially constrained. As the financial meltdowns of the last few decades have shown, even large and mature firms can quickly become financially constrained.

Beyond the obvious control variables, like market capitalization, prior returns, and negative earnings indicator, percent negative words adds more value in identifying which firm will be financially constrained than widely used constraint indexes. The SA and WW indexes, in univariate regressions, do a reasonable job at predicting two of the four liquidity events. Yet, in the presence of market capitalization, neither the SA or WW indexes have explanatory power in the expected direction. The important contribution of percent negative words is that it adds value beyond the standard list of control variables.

The more managers believe the firm will face a more uncertain future, the more the text of the 10-K will reflect this negative outlook. Our measure has several important advantages: (1) since no subjective reading of text is required, our measure is easy to replicate; (2) the entire CRSP/Compustat universe of 10-K filers can be included in the analysis instead of small hand-collected samples; and (3) negative word frequency captures subtle cues from managers who may not be required to issue explicit liquidity warnings to investors.

We test the ability of the KZ, SA, WW indexes and percentage of negative 10-K words to predict four liquidity events after controlling for standard firm characteristics. The ex post liquidity events are dividend cuts or omissions, debt downgrades, and total asset growth. With or without the presence of firm characteristic control variables, percentage of negative words outperforms the other widely used financial constraint measures when explaining subsequent liquidity events. The percentage of negative words also has a reasonable economic impact. For example, a one standard deviation increase in negative words increases the likelihood of a dividend omission by 8.9% and a debt downgrade by 10.8%.

The explicit use of negative words around the terms *liquidity* and *capital resources* is also examined. We find that more than 30% of the sample never uses negative words in close proximity to *liquidity* and *capital resources*. We report that more negative words around the liquidity terms is linked with a higher likelihood of a dividend omission or lower total asset growth after controlling for other variables. The targeted word search does not, however, work as well as the aggregate textual analysis.

The gauging of firm level financial constraints is a critically important research area. We extend the literature by recommending researchers use the percentage of negative 10-K words as an indicator of financial constraints beyond the usual macro-finance control variables. A higher frequency of negative words in the language used by managers to describe current and subsequent operations helps predict a more financially constrained future for the company. The percent of negative words measure is relatively easy to calculate and available for all firms filing annual 10-Ks with the SEC. The application of textual analysis as an additional measure of financial constraints provides an example of how qualitative information can provide a differentiated contribution to the usual mix of financial and accounting variables.

Appendix

A.1 Definitions of the variables used in the paper

% Negative	Percentage of words in the 10-K that are on Loughran and McDonald's (2011) negative word list. Examples of negative words include *losses, impairment, decline, volatility,* and *doubtful*
Age	Following Hadlock and Pierce (2010), age is defined as the number of years the firm is listed with a non-missing stock price on Compustat at the time of the 10-K filing.
Total Assets	Compustat data item AT.
SA Index	Following Hadlock and Pierce (2010), the SA index is defined as [-0.737*log(Total Assets)]+[0.043*log(Total Assets)^2]-(0.040*Age). Total Assets are winsorized at $9,531 million (95% percentile of our sample) while Age is winsorized at 42 years (our 95% percentile). Higher values of the SA index imply greater levels of financial constraint.
Dividend Dummy	Dummy variable set to one if Compustat reports a positive preferred (item DVP) or common dividend (item DVC), else zero.
Cash Flow	Cash flow is defined as income before extraordinary items (item IB) plus depreciation (item DP).
Cash	Cash and short-term investments (item CHE).
KZ Index	Following Lamont, Polk, Saa-Requejo (2001), the KZ index is defined as -1.001909*[Income before extraordinary items (item IB) + Depreciation (item DP)]/lagged Property, Plant, and Equipment (item PPENT) + 0.2826389*[Total Assets (item AT) + Market Value as of December year t-1 – Common Equity (item CEQ) – Deferred Taxes (TXDB)]/Total Assets + 3.139193*[Long-term Debt (item DLTT) + Short-term Debt (item DLC)]/[Long-term Debt + Short-term Debt + Shareholder Equity (item SEQ)] – 39.3678*[Common Dividends (DVC) + Preferred Dividends (DVP)]/ lagged Property, Plant, and Equipment – 1.314759*(Cash (item CHE)/lagged Property, Plant, and Equipment. Each of the individual components of the KZ index are winsorized at the 5% level. Higher levels of the KZ index

imply that the firm is more financial constrained.

Cash/Total Assets	The ratio of cash (item CHE)/total assets (item AT).
Sales Growth	Firm sales growth is the firm's most recent annual percentage change in sales (item SALE). Thus, sales growth is (sales in year t minus sales in year t-1) / (sales in year t-1). The sales growth variable is winsorized at the 1% level.
Industry Sales Growth	Industry sales growth is defined as the most recent annual percentage change in aggregate industry sales. Firms within the same three-digit SIC industry are aggregated to calculate sales growth for the industry.
WW Index	The Whited-Wu index is defined as (−0.091*CF) – (0.062*Dividend Dummy) + (0.021*TLTD) – (0.044*LNTA) + (0.102*ISG) – (0.035*SG) where CF is a ratio of cash flow divided by total assets (item AT); dividend dummy is equal to one if the firm pays a dividend, else zero; TLTD – long-term debt to total assets; LNTA – logarithm of total assets; ISG – three-digit SIC industry sales growth; and SG – firm sales growth. All of the individual components of the WW index are winsorized at the 5% level except for dividend dummy, long-term debt to assets, and log of total assets. Higher values of the WW index imply greater levels of financial constraint.
# Negative/# Liquidity	This variable is defined as the number of negative words near the terms *liquidity, capital resource,* or *capital resources* divided by the count of how often the firm mentions the terms *liquidity* and *capital resources.*

<u>Control Variables</u>

Market Capitalization	The variable is stock price multiplied by shares outstanding (in millions of dollars) as of June of year t.
Book-to-Market	This variable is defined as the prior year's book value of equity (Compustat data item CEQ plus balance sheet deferred taxes and investment tax credit (item TXDITC)) divided by the firm's market value as of December of year t-1. Firms with negative or missing values of CEQ are dropped. The variable is winsorized at the 1% level.

Excess Prior Returns	The buy-and-hold firm returns during the prior year minus the buy-and-hold returns of the CRSP value-weighted index over an identical period.
Negative Earnings Dummy	Dummy variable set to one if the firm has a negative income before extraordinary items (item IB), else zero.
Debt Rating Dummies	The seven debt rating dummies are from Ashbaugh-Skaife, Collins, and LaFond (2006, Table 1).
<u>Liquidity Events</u>	
Dividend Cut Dummy	Dummy variable set to one if the firm has a lower aggregate dividend (controlling for stock splits) during July of year t to June of year t+1 than in the prior year, else zero. Only firms which issued a dividend in the year before the June formation date are assigned a value for this variable.
Dividend Omission Dummy	Dummy variable set to one if the firm completely omits paying a dividend during July of year t to June of year t+1, else zero. Only firms issuing a dividend in the year before the June formation date are assigned a value for this variable.
Debt Downgrade Dummy	Dummy variable set to one if the firm has a lower S&P rating score in June of year t+1 compared to its rating as of June of year t, else zero. Only firms with an S&P debt rating in both June of year t and June of year t+1 are assigned a value for this variable. The S&P long-term domestic credit rating (item SPLTICRM) is converted into the assigned rating score (range of 1 to 7) of Ashbaugh-Skaife, Collins, and LaFond (2006, Table 1). For example, AAA debt ratings are assigned a value of 7; AA+, AA, AA- are assigned a value of 6; A+, A, and A- have a rating score of 5; etc. A decrease in this integer classification is considered a downgrade.
Change in Assets	The ratio of (total assets in year t+1 – total assets in year t)/total assets in year t. This variable is winsorized at the 5% level.

A.2. Parsing the 10-K filings

A.2.1. Stage one parsing

All 10-K SEC complete text document filings are downloaded for each year/quarter. We use "10-K" to represent any SEC filing that is a 10-K variant, i.e., 10-K/A, 10-K405, 10KSB, and 10KSB40. We do not include amended filings. The text version of the filings provided on the SEC server is an aggregation of all information provided in the browser-friendly files also listed on EDGAR for a specific filing. For example, IBM's 10-K filing on 20120228 lists the core 10-K document in html format, ten exhibits, four jpg (graphics) files, and six XBRL files.[11] All of these files are also contained in a single text file with the embedded HTML, XBRL, exhibits, and the ASCII-encoded graphic.[12] In the IBM example, of the 48,253,491 characters contained in the file, only about 7.6% account for the 10-K text including the exhibits and tables. The HTML coding accounts for about 55% of the file. The XBRL tables have a very high ratio of tags to data and account for about 33% of the text file. The remaining 27% of the file is attributable to the ASCII-encoded graphics. In many cases, ASCII-encoded pdfs, graphics, xls, or other binary files that have been encoded can account for more than 90% of the document.

Because most textual analysis studies focus on the textual content of the document, we have created files where all of the 10-K documents have been parsed to exclude markup tags, ASCII-encoded graphics, and tables. We exclude tables, because they are not the focus of our textual analysis.

Each of these raw text files downloaded from EDGAR is parsed using the following sequence (Relevant Regular Expression code is provided in parentheses.):

1. Remove ASCII-Encoded segments – All document segment <TYPE> tags of GRAPHIC, ZIP, EXCEL and PDF are deleted from the file. ASCII-encoding is a means of converting binary-type files into standard ASCII characters to facilitate transfer across various hardware platforms. A relative small graphic can create a substantial ASCII segment. Filings containing multiple graphics can be orders of magnitude larger than those containing only textual information.

[11] XBRL (eXtensible Business Reporting Language) is a markup language. A variant of XML and related to HTML, it provides semantic context for data reported within a 10-K. For example, one line in Google's 20111231 10-K filing contains "<us-gaap:StockholdersEquity contextRef="eol_PE633170--1110-K0018_STD_0_20081231_0" unitRef="iso4217_USD" decimals="-6">28239000000</us-gaap:StockholdersEquity>". The "eol …" segment defines the XBRL implementation, the data are in US dollars and the "-6" indicates the number is rounded to millions. See http://xbrl.sec.gov. A few firms began including XBRL in their filings in 2005 with the number expanding substantially in 2010.

[12] ASCII-encoding converts binary data files to plain ASCII-printable characters, thus ensuring cross platform conformity. The conversion from binary to plain text increases the size of the original file by orders of magnitude.

2. Remove <DIV>,<TR>,<TD> and <FONT> tags – Although we require some HTML information for subsequent parsing, the files are so large (and processed as a single string) that we initially simply strip out some of the formatting HTML.
3. Remove all XBRL – all characters between <XBRL …> … </XBRL> are deleted.
4. Remove SEC Header/Footer – All characters from the beginning of the original file thru </SEC-HEADER> (or </IMS-HEADER> in some older documents) are deleted from the file after identifying the SIC classification. In addition the footer “-----END PRIVACY-ENHANCED MESSAGE-----” appearing at the end of each document is deleted.
5. Remove tables – all characters appearing between <TABLE> and </TABLE> tags are removed.
    a. Note that some filers use table tags to demark paragraphs of text, so each potential table string is first stripped of all HTML and then the number of numeric versus alphabetic characters is compared. For this parsing, only table encapsulated strings where *numeric chars/(alphabetic+numeric chars)* > 15%.
    b. In some instances, Item 7 and/or Item 8 of the filings begins with a table of data where the Item 7 or 8 demarcation appears as a line within the table string. Thus, any table string containing “Item 7” or “Item 8” (case insensitive) is *not* deleted.
6. Remove Markup Tags – remove all remaining markup tags (i.e., <…>).
7. Re-encode reserved HTML characters (character entity references)—In order to encode a broad set of universal characters within the limitations of ASCII coding many characters are encoded. For example, the “&” symbol can be encoded as “&” or “&”. For items listed below we replace the encode items with a character(s). The remaining encoded items are deleted.
    a. “<” or “&#60” -> “ LT “ - note we use LT instead of “<” to avoid any confusion with markup tags.
    b. “>” or “&#62” -> “ GT “
    c. “ ” or “ ” -> “ “
    d. “"” or “&#34” -> “””
    e. “&APOS;” or “&#39” -> “’”
    f. “&” or &#38” -> “&”
    g. All Regular Expression \t and \v items are deleted.
    h. All remaining ISO 8859-1 symbols and characters are deleted.
8. Finally some remaining idiosyncratic anomalies are parsed out:
    a. Linefeeds (\n) following hyphens are removed.
    b. Hyphens preceded and followed by a blank space are removed.
    c. The token “and/or” (case insensitive) is replaced by “and or”.
    d. Sequences of two or more hyphens, periods or equal signs possibly followed by spaces (e.g., REGEX = “(-|\.|=)\s*”) are removed.
    e. All underscore characters (“_”) are removed.
    f. All sequences of three or more blanks are replaced by a single blank.
    g. All sequences of three or more linefeeds possibly separated by spaces (REGEX = “(\n\s*){3,}”) are replaced by two linefeeds.
    h. All linefeeds not preceded by a linefeed and not followed by a blank or linefeed are replaced by a blank.
9. Delete SEC header.

10. Delete hyphens preceding a linefeed.
11. Replace hyphens preceding a capitalized letter with a space.
12. Delete names and unambiguous proper nouns.
13. Delete capitalized or all capitals for March, May, and August.
14. Delete possessive “s”.
15. Remove phrase “Table of Contents” (which can occur as a link at the top of each page).
16. Remove page numbers.

The remaining text in each filing is then parsed into words and counts are created for the various tests.

A.2.2 Stage two parsing

In addition to parsing the full document, each occurrence of word “liquidity” or phrase “capital resources” is identified and the segment of text 100 characters preceding the word/phrase and 300 characters following the word phrase is isolated. Negative word counts are then performed for each of these segments.

References

Acharya, V., Davydenko, S., Strebulaev, I., 2012. Cash holdings and credit risk. Review of Financial Studies 25, 3572–3609.

Almeida, H., Campello, M., Weisbach, M., 2004. The cash flow sensitivity of cash. Journal of Finance 59, 1777–1804.

Antweiler, W., Frank, M., 2004. Is all that talk just noise? The information content of Internet stock message boards. Journal of Finance 59, 1259-1293.

Ashbaugh-Skaife, H., Collins, D., LaFond, R., 2006. The effects of corporate governance on firms' credit ratings. Journal of Accounting and Economics 42, 203-243.

Ball, C., Hoberg, G., Maksimovic, V., 2012. Redefining financial constraints: a text-based analysis. Unpublished working paper, University of Maryland.

Battacharya, S., 1979. Imperfect information, dividend policy and the 'bird in the hand' fallacy. Bell Journal of Economics 10, 259-270.

Campello, M., Graham, J., Harvey, C., 2010. The real effects of financial constraints: Evidence from a financial crisis. Journal of Financial Economics 97, 470-487.

Carpenter, R., Petersen, B., 2002. Is the growth of small firms constrained by internal finance? The Review of Economics and Statistics 84, 298-309.

Cleary, S., 1999. The relationship between firm investment and financial status. Journal of Finance 54, 673-692.

Denis, D., Sibilkov, V., 2010. Financial constraints, investment, and the value of cash holdings. Review of Financial Studies 23, 247-269.

Doidge, C., Karolyi, A., Stulz, R., 2013. The U.S. left behind? Financial globalization and the rise of IPOs outside the U.S. Journal of Financial Economics 110, 546-573.

Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Fama, E., French, K., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Faulkender, M., Wang, R. 2006. Corporate financial policy and the value of cash, Journal of Finance 61, 1957-1990.

Fazzari, S., Hubbard, R., Petersen, B., 1988. Financing constraints and corporate investment, Brookings Papers on Economic Activity 1, 141-206.

Fazzari, S., Hubbard, R., Petersen, B., 2000. Investment-cash flow sensitivities are useful: A comment on Kaplan and Zingales. Quarterly Journal of Economics 115, 695–705.

Gao, X., Ritter, J., Zhu, Z., 2013. Where have all the IPOs gone? Journal of Financial and Quantitative Analysis, forthcoming.

Hadlock, C., Pierce, J., 2010. New evidence on measuring financial constraints: Moving beyond the KZ Index. Review of Financial Studies 23, 1909–1940.

Hanley, K., Hoberg, G., 2010. The information content of IPO prospectuses. Review of Financial Studies 23, 2821-2864.

Hanley, K., Hoberg, G., 2012. Litigation risk, strategic disclosure and the underpricing of Initial Public Offerings. Journal of Financial Economics 103, 235–254.

Harford, J., 1999. Corporate cash reserves and acquisitions. Journal of Finance 54, 1969-1997.

Hennessy, C., Whited, T., 2007. How costly is external financing? Evidence from a structural estimation. Journal of Finance 62, 1705-1745.

Kaplan, S., Zingales, L., 1997. Do financing constraints explain why investment is correlated with cash flow? Quarterly Journal of Economics 112, 169–216.

Lamont, O., Polk, C., Saa´-Requejo, J., 2001. Financial constraints and stock returns. Review of Financial Studies 14, 529–544.

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

Miller, M., 1988. The Modigliani-Miller propositions after thirty years. Journal of Economic Perspectives 2, 99-120.

Myers, S., Majluf, N., 1984. Corporate financing and investment decisions when firms have information that investors do not have. Journal of Financial Economics 13, 187–221.

Opler T., Pinkowitz, L., Stulz, R., Williamson, R., 1999. The determinants and implications of corporate cash holdings. Journal of Financial Economics 52, 3-46.

Riddick, L., Whited, T., 2009. The corporate propensity to save. Journal of Finance 64, 1729–1766.

Standard and Poor, 2013, Credit Ratings Definitions and FAQ, http://www.standardandpoors.com/ratings/definitions-and-faqs/en/us.

Sufi, A., 2009. Bank lines of credit in corporate finance: An empirical analysis. Review of Financial Studies 22, 1057-1088.

Tetlock, P., 2007. Giving content to investor sentiment: The role of media in the stock market. Journal of Finance 62, 1139–1168.

Tetlock, P., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. Journal of Finance 63, 1437-1467.

Townsend, R., 1979. Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory 21, 265–293.

Whited, T., 2009. What can cash shortfalls and windfalls tell us about finance constraints? In Giorgio Calcagnini and Enrico Saltari (eds.) The Economics of Imperfect Markets, pp. 17–32.

Whited, T., Wu, G., 2006. Financial constraints risk. Review of Financial Studies 19, 531–559.

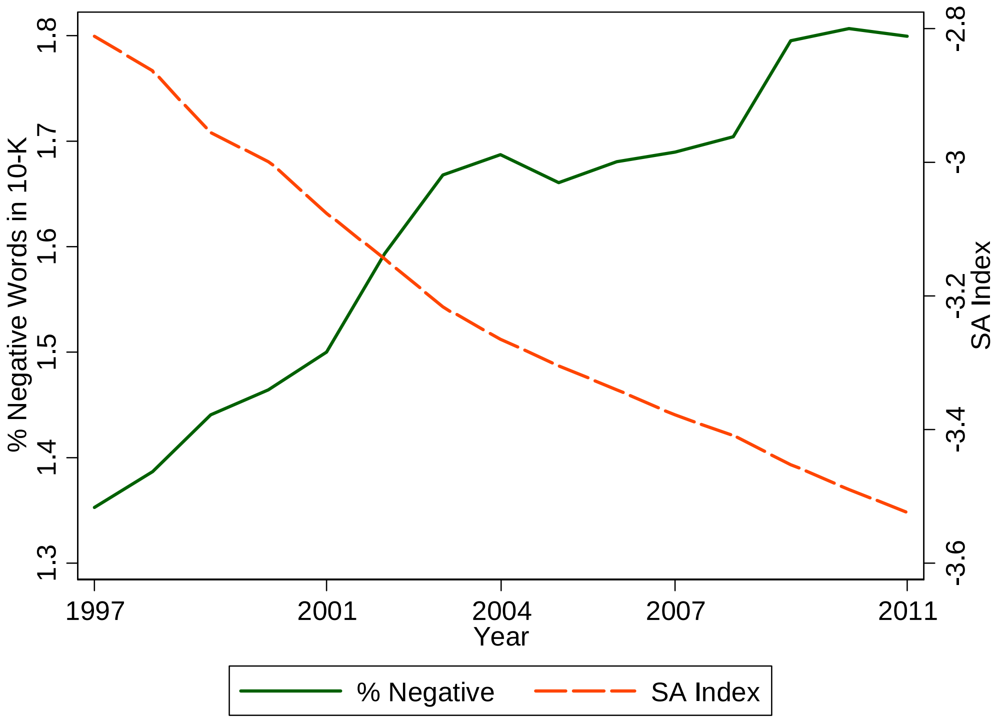

Fig. 1. We plot the time series of mean % negative words in 10-K and mean SA Index by year, 1997-2011. The Loughran-McDonald (2011) word dictionary is used to define negative words. Following Hadlock and Pierce (2010), the SA Index is defined as [-0.737 × log(Total Assets)] + [0.043 × log(Total Assets)^2] - (0.040*Age).

Table 1
Sample creation.
This table reports the impact of various data filters on the initial 10-K sample.

	Dropped	**Sample Size**
SEC 10-K files 1996–2011		183,214
Drop financial firms and utilities	49,222	133,992
Eliminate duplicates within year/CIK	3,542	130,450
Drop if file date < 180 days from prior	464	129,986
CRSP PERMNO match and ordinary common equity	70,809	59,177
Drop if number of 10-K words is < 2,000	40	59,137
Drop if required Compustat data is missing	3,607	55,530
Market capitalization data available on CRSP	3,997	51,533

Table 2

Summary statistics and correlations.

We present descriptive statistics by time period in Panel A; corrections between key variables of interest are reported in Panel B. In Panel A, the sample sizes vary by the availability of data. For % negative, SA index, book-to-market, market capitalization, excess prior returns, and negative earnings dummy, the sample is 51,533. The sample size is 12,806 for dividend cut and dividend omit since firms are required to be distributing dividends in the prior year to be included. The overall sample size for debt downgrade dummy is 11,223 due to the requirement of having an available S&P long-term domestic credit rating in both June of year t and June of year t+1. See Appendix for detailed variable definitions.

Panel A: Mean Summary Statistics

Variable	(1) 1997 to 2003	(2) 2004 to 2011	(3) 1997 to 2011
% Negative	1.47	1.72	1.59
KZ Index	-4.34	-6.05	-5.12
SA Index	-2.99	-3.39	-3.17
WW Index	-0.24	-0.28	-0.26
Dividend Cut Dummy	12.05%	9.57%	10.74%
Dividend Omission Dummy	3.96%	3.38%	3.65%
Debt Downgrade Dummy	9.06%	6.80%	7.89%
Change in Assets	14.41%	11.70%	13.18%
Book-to-Market	0.79	0.65	0.73
Market Capitalization	$2,131.1	$3,464.1	$2,736.0
Excess Prior Returns	0.72%	6.74%	3.45%
Negative Earnings Dummy	38.52%	33.88%	36.42%

Panel B: Correlations

	% Negative	*KZ Index*	*SA Index*	*WW Index*	*Log(Mkt Cap)*	*Log (Book-to-Market)*	*Excess Prior Returns*
KZ Index	-0.114						
SA Index	-0.010	-0.145					
WW Index	0.014	-0.055	0.832				
Log(Mkt Cap)	0.071	-0.028	-0.702	-0.839			
Log(Book-to-Market)	-0.021	0.150	-0.123	-0.041	-0.357		
Excess Prior Returns	0.013	-0.007	-0.028	-0.019	0.176	-0.338	
Negative Earnings Dummy	0.233	0.010	0.361	0.426	-0.343	0.015	-0.070

Table 3

Regressions of cash flow sensitivity on different measures of financial constraints, 1997–2011.

The dependent variable is the change in cash and marketable securities scaled by total assets. Cash flow is the ratio of income plus depreciation divided by total assets. Q is the market value of equity divided by total assets. The natural log of total assets is the measure of firm size. Constrained bond rating firms have positive long-term debt in a given year with no S&P bond rating. Unconstrained bond rating firms have positive long-term debt and an S&P bond rating on their long-term debt in a given year. For total assets, the bottom three deciles are defined as constrained firms while the top three deciles are defined as unconstrained. For KZ index, SA index, WW index, and % negative words, the constrained firms are defined to be in the top three deciles of the respective measures while unconstrained firms are in the bottom three deciles of the individual categories, determined on an annual basis. The Loughran and McDonald (2011) word lists are used to define negative 10-K words. All financial and utility firms are excluded from the sample. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry.

	Independent variables			
Dependent variable: *Change in Cash Holdings*	Cash Flow	Q	Size	R^2
Financial Constraints Criteria				
1. *Bond Ratings*				
Constrained firms	0.190	0.028	0.003	8.14%
	(5.46)	(5.18)	(1.27)	
Unconstrained firms	0.017	0.013	-0.002	3.51%
	(0.60)	(4.67)	(-2.70)	
2. *Total Assets*				
Constrained firms	0.263	0.024	0.029	10.16%
	(7.34)	(6.84)	(3.07)	
Unconstrained firms	0.029	0.015	-0.003	5.86%
	(0.65)	(4.61)	(-2.71)	
3. *KZ Index*				
Constrained firms	0.269	-0.005	0.004	23.57%
	(8.82)	(-2.17)	(4.38)	
Unconstrained firms	0.086	0.033	-0.003	10.26%
	(1.33)	(7.54)	(-1.65)	
4. *SA Index*				
Constrained firms	0.272	0.029	0.038	12.69%
	(7.08)	(7.27)	(3.51)	
Unconstrained firms	0.098	0.008	-0.001	4.04%
	(3.45)	(3.25)	(-1.21)	
5. *WW Index*				
Constrained firms	0.301	0.023	0.015	10.98%
	(9.16)	(6.93)	(2.07)	
Unconstrained firms	-0.119	0.020	-0.006	9.08%
	(-1.59)	(4.20)	(-2.98)	
6. *% Negative Words*				
Constrained firms	0.276	0.030	0.001	11.41%
	(8.24)	(5.63)	(0.66)	
Unconstrained firms	0.139	0.015	-0.002	4.69%
	(3.44)	(5.30)	(-2.03)	

Table 4
Preliminary tests.

Each cell represents a separate regression, with each column representing the four ex post liquidity events as a separate dependent variable. The independent variables are the three indexes and % negative words. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. An "X" signifies a 1% significance level on the constraint index or negative word frequency variables in the predicted direction. The standard errors in all the regressions are clustered by both year and industry. In total, the results from 16 different regressions (4 x 4) are reported.

	Ex Post Liquidity Events			
Independent variables	*Dividend Cut Dummy*	*Dividend Omission Dummy*	*Debt Downgrade Dummy*	*Change in Assets*
KZ Index			X	X
SA Index	X	X		
WW Index	X	X		
% Negative Words	X	X	X	X

Table 5

Logit regressions with dividend cut dummy as the dependent variable, 1997-2011.

The dependent variable is *Dividend Cut Dummy*, set to one if the firm has lower aggregate dividends (controlling for stock splits) during July of year t to June of year t+1 than in the prior year, else zero. *% Negative* is defined as the percentage of words in the 10-K that are on Loughran and McDonald's (2011) negative word list. All regressions include an intercept, seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006), Fama and French 48-industry dummies, and calendar year dummies. See Appendix for definitions of all other variables. The z-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)	(4)	(5)
KZ Index		-0.024			
		(-5.06)			
SA Index			-0.003		
			(-0.04)		
WW Index				-7.944	
				(-5.82)	
% Negative					0.234
					(2.47)
Log(Mkt Cap)	-0.253	-0.244	-0.254	-0.618	-0.264
	(-6.86)	(-6.83)	(-5.36)	(-7.78)	(-7.11)
Log(Book-to-Market)	0.090	0.120	0.089	-0.187	0.080
	(1.01)	(1.31)	(0.89)	(-2.37)	(0.89)
Excess Prior Returns	-0.008	-0.008	-0.008	-0.008	-0.008
	(-3.97)	(-3.85)	(-3.96)	(-3.86)	(-4.01)
Negative Earnings Dummy	0.994	1.041	0.994	1.026	0.946
	(12.37)	(13.85)	(12.30)	(11.53)	(12.06)
Pseudo R^2	12.67%	13.21%	12.67%	13.72%	12.78%
Sample size	12,806	12,687	12,806	11,777	12,806

Table 6

Logit regressions with dividend omission dummy as the dependent variable, 1997-2011.

The dependent variable, *Dividend Omission Dummy*, is set to one if the firm completely omits paying a dividend during July of year t to June of year t+1, else zero. Only firms paying a dividend in the prior year are included in these regressions. *% Negative* is defined as the percentage of words in the 10-K from Loughran and McDonald's (2011) negative word list. All regressions include an intercept, seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006), Fama and French 48-industry dummies, and calendar year dummies. See Appendix for definitions of all other variables. The *z*-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)	(4)	(5)
KZ Index		-0.001			
		(-0.11)			
SA Index			0.010		
			(0.08)		
WW Index				-14.372	
				(-8.15)	
% Negative					0.417
					(3.60)
Log(Mkt Cap)	-0.254	-0.277	-0.251	-0.890	-0.277
	(-10.57)	(-9.97)	(-5.53)	(-17.16)	(-13.53)
Log(Book-to-Market)	0.293	0.277	0.295	-0.201	0.278
	(2.50)	(2.29)	(2.12)	(-17.19)	(2.36)
Excess Prior Returns	-0.014	-0.014	-0.014	-0.012	-0.013
	(-4.84)	(-4.75)	(-4.85)	(-3.94)	(-4.86)
Negative Earnings Dummy	1.023	1.043	1.023	1.102	0.940
	(10.65)	(10.29)	(10.97)	(13.55)	(10.42)
Pseudo R^2	19.15%	19.46%	19.15%	20.93%	19.43%
Sample size	12,669	12,551	12,669	11,653	12,669

Table 7

Logit regressions with debt downgrade dummy as the dependent variable, 1997-2011.

The dependent variable, *Debt Downgrade Dummy*, is set to one if a firm has a lower S&P rating score in June of year t+1 compared to its rating as of June of year t, else zero. Only firms with an S&P debt rating in both June of year t and June of year t+1 are assigned a value for this variable. *% Negative* is defined as the percentage of words in the 10-K from Loughran and McDonald's (2011) negative word list. All regressions include an intercept, seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006), Fama and French 48-industry dummies, and calendar year dummies. See Appendix for definitions of all other variables. The *z*-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)	(4)	(5)
KZ Index		0.027			
		(1.96)			
SA Index			0.066		
			(0.90)		
WW Index				-4.504	
				(-3.83)	
% Negative					0.490
					(4.23)
Log(Mkt Cap)	-0.373	-0.370	-0.364	-0.530	-0.392
	(-5.92)	(-5.93)	(-5.53)	(-6.15)	(-6.20)
Log(Book-to-Market)	0.250	0.237	0.255	0.155	0.243
	(2.18)	(2.10)	(2.25)	(1.59)	(2.19)
Excess Prior Returns	-0.019	-0.019	-0.019	-0.018	-0.019
	(-9.35)	(-9.14)	(-9.35)	(-9.19)	(-9.20)
Negative Earnings Dummy	1.006	0.992	1.014	1.028	0.936
	(8.62)	(8.61)	(8.95)	(8.14)	(7.78)
Pseudo R^2	17.24%	17.42%	17.25%	17.34%	17.61%
Sample size	11,042	10,966	11,042	10,207	11,042

Table 8

Regressions with change in assets as the dependent variable, 1997-2011.

The dependent variable, *Change in Assets*, is defined as the ratio of (Total Assets in year t+1 – Total Assets in year t)/Total Assets in year t. *% Negative* is defined as the percentage of words in the 10-K from Loughran and McDonald's (2011) negative word list. All regressions include an intercept, seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006), Fama and French 48-industry dummies, and calendar year dummies. The last column removes firms in the bottom quarter or top quarter in terms of asset growth (determined on an annual basis). See Appendix for definitions of all other variables. The *t*-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)	(4)	(5)	(6)
KZ Index		-0.001				
		(-1.28)				
SA Index			0.098			
			(7.16)			
WW Index				0.716		
				(8.13)		
% Negative					-0.015	-0.007
					(-2.26)	(-4.37)
Log(Mkt Cap)	0.002	0.002	0.030	0.034	0.002	0.002
	(0.52)	(0.60)	(5.15)	(6.26)	(0.67)	(2.73)
Log(Book-to-Market)	-0.153	-0.150	-0.124	-0.128	-0.153	-0.015
	(-10.61)	(-10.35)	(-10.57)	(-9.21)	(-10.55)	(-7.95)
Excess Prior Returns	0.001	0.001	0.001	0.001	0.001	0.001
	(8.89)	(9.24)	(8.81)	(8.49)	(8.89)	(3.99)
Negative Earnings Dummy	-0.096	-0.096	-0.107	-0.112	-0.093	-0.021
	(-11.38)	(-10.94)	(-12.04)	(-12.36)	(-10.27)	(-10.31)
R^2	21.71%	21.78%	22.63%	22.74%	21.73%	30.85%
Sample size	50,892	50,328	50,892	45,532	50,892	25,441

Table 9

Regressions with the four liquidity events as the dependent variables and an alternative measure of negative words, 1997-2011.

The dependent variables are *Dividend Cut Dummy* (column (1)); *Dividend Omission Dummy* (column (2)); *Debt Downgrade Dummy* (column (3)); and *Change in Assets* (column (4)). # Negative/# Liquidity is defined as the number of negative words near the terms *liquidity* and *capital resources* divided by the count of how often the firm mentions the terms *liquidity* and *capital resources.* All regressions include an intercept, seven debt rating dummies from Ashbaugh-Skaife, Collins, and LaFond (2006), Fama and French 48-industry dummies, and calendar year dummies. See Appendix for definitions of all other variables. The *z*-statistics or *t*-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)	(4)
	Dividend Cut Dummy	Dividend Omission Dummy	Debt Downgrade Dummy	Change in Assets
# Negative / # Liquidity	0.068	0.184	0.026	-0.009
	(1.10)	(2.55)	(0.51)	(-3.41)
Log(Mkt Cap)	-0.256	-0.261	-0.374	0.002
	(-6.81)	(-12.16)	(-5.95)	(0.49)
Log(Book-to-Market)	0.087	0.286	0.249	-0.153
	(0.98)	(2.51)	(2.17)	(-10.60)
Excess Prior Returns	-0.008	-0.014	-0.019	0.001
	(-4.01)	(-4.93)	(-9.35)	(8.88)
Negative Earnings Dummy	0.982	0.989	1.003	-0.094
	(11.46)	(9.86)	(8.25)	(-11.52)
Pseudo R^2 / Adj. R^2	12.70%	19.34%	17.25%	21.73%
Sample size	12,806	12,669	11,173	50,892