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

August 25, 2015

Key words: Financial constraints; textual analysis; dividend omissions; equity recycling; underfunded pensions.
JEL Classifications: G31, G32, D92.

We thank Matt Cain, Andreas Chouliaras, Mara Faccio, Paul Gao, Jarrad Harford (the editor), Chris Henderson, Gerard Hoberg, Fred Mittelstaedt, an anonymous referee, and seminar participants at the 2014 FMA Europe Conference, University of Lugano, University of Mannheim, University of Nebraska–Lincoln, and Emory University for helpful comments.

Using 10-K Text to Gauge Financial Constraints

Abstract

Measuring the extent to which a firm is financially constrained is critical in assessing capital structure. Extant measures of financial constraints focus on macro firm characteristics such as age and size – variables highly correlated with other firm attributes. We parse 10-K disclosures filed with the SEC using a unique lexicon based on constraining words. We find that the frequency of constraining words exhibits very low correlation with traditional measures of financial constraints and predicts subsequent liquidity events—like dividend omissions or increases, equity recycling, and underfunded pensions—better than widely-used financial constraint indexes.

I. 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 managers discuss difficulties in obtaining external financing, liquidity problems, or forced reduction in investment 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.

In manually reading the 10-K text for constraining tone, KZ and HP do not list all of the specific words used to identify the constraining sentences. Due to the time intensive nature of their method, their analyses were limited to relatively small samples of firms. With advances in textual analysis, why not have computers parse the text for constraining tone? This would greatly expand the potential sample of firms and improve the ability of others to replicate a paper's empirical results. The missing piece for researchers is a list of constraining words.

The main contribution of our paper is the creation of a constraining word list to assist other researchers in identifying whether or not a firm is financially constrained. The list contains 184 constraining words. Like the Loughran and McDonald (2011) creation of six word lists (negative, positive, uncertain, litigious, strong modal, and weak modal), we examine tens of thousands of words that appear in at least 5% of all 10-K filings. We only select words which would be most likely considered constraining in the majority of occurrences.

Specifically, 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 constraining words. Commonly used constraining words from our list are *required, obligations, requirements, permitted, comply,* and *imposed.* Our conjecture is that managers anticipating financial challenges will use a more constraining 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. Clearly, part of the use of constraining words by managers is with an eye towards lowering litigation exposure.[1]

Our paper thus 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

[1] As might be expected, there is a positive correction (0.386) between the frequency of our constraining words in the 10-K text and the frequency of litigious words from the Loughran and McDonald (2011) word list.

measure to predict subsequent events traditionally associated with either deterioration or improvement of external financing conditions, events which we label "liquidity events." Our four ex post liquidity events include dividend omissions, dividend increases, equity recycling (i.e., paying out equity proceeds to shareholders in the form of share buybacks and dividends), and underfunded pension plans. Note that the events include instances of being financially constrained (e.g., dividend omissions) and cases identifying firms that are clearly less constrained (e.g., equity recycling). 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.

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 or eliminate dividend payments whereas unconstrained firms did not. Firms with high levels of cash dividends or share repurchases relative to equity issuance are unlikely to be financially constrained. Finally, Rauh (2006) notes that capital expenditures decrease when a firm has to make mandatory contributions for its defined pension benefit plan.

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 the New York Times. As of June 2008, the New York Times had a large market value (over $2 billion), total assets of $3.5 billion, positive trailing cash flows, 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 would not be identified as financially constrained). Yet, within 12 months, the New York Times completely eliminated its dividends, did not engage in equity recycling, experienced a 63% raw decline in its stock price, and continued to have an underfunded employee pension plan.

Interestingly, the Times' 10-K filed on February 26, 2008 contained 1.05% constraining words (which put it in the top 1.5 percentile of all firms in that year). This was the firm's highest constraining word percentage of any year in our sample. Its high constraining count was caused by discussions in the New York Times 10-K concerning all the debt, legal, employee, and environmental constraints facing the firm. For example, the company notes that 47% of its workers were unionized ("As a result, we are *required* to negotiate the wages, salaries, benefits, staffing levels…"); the document also includes discussions about credit agencies ("To maintain our investment-grade ratings, the credit rating agencies *require* us to meet certain financial performance ratios"); a mandatory contract with a major paper supplier ("The contract *requires* us to purchase annually the lesser of a fixed number of tons…"); obligations ("The Company would have to perform the *obligations* of the National Edition printers under the equipment and debt guarantees if the National Edition printers defaulted under the terms of their equipment leases or debt agreements"); and underfunded defined benefit pension plans ("As of December

30, 2007, our postretirement *obligation* was approximately $229 million, representing the unfunded status of our postretirement plans") (constraining words are in italics).

So although some widely used financial constraint indexes would imply smooth sailing for the New York Times as of 2008, the high frequency of constraining 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 or age.

We show that the constraining tone of 10-K documents is a measure of financial constraints distinct from measures based on accounting characteristics. Further, the percent of constraining words, unlike the SA and WW indexes, has a low correlation with market capitalization. When we turn our attention to the ability of various measures of financial constraints to predict events related to the deterioration or improvement in external financing conditions, we find that a more frequent usage of constraining words is strongly related to a higher likelihood of future dividend omission (+10.32%), increases (-6.46%), equity recycling (-23.24%), and underfunded pensions (+2.34%).[2] The results are stronger in the cross-section than in the time-series and are also robust to inclusion of firm characteristics, e.g., market value, book-to-market, negative earnings dummy, 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 the liquidity events even without the presence of control variables.

The inability of the KZ index, SA index, and WW index to predict liquidity events is consistent with the findings of Farre-Mensa and Ljungqvist (2015). The two authors present

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

strong evidence that the three commonly used indexes do a poor job of identifying firms that are plausibly considered financially constrained. Surprisingly, Farre-Mensa and Ljungqvist (2015) find that 'constrained' firms identified by the three indexes are able to raise debt when it is in their best interest, continue to obtain bank borrowing after a negative shock to the supply of local bank loans, and even engage in equity recycling. Thus, these 'financially constrained' firms do not appear to face inelastic capital supply curves as would be suggested by their index values.

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 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 constraining words in the 10-K – is objective. That is, we do not assign the financial constraint scores by actually reading the document, but rely on the output of the pre-specified automated parsing algorithm. Since we use the constraining word list, 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 since we provide our entire constraining word list for other researchers to use.

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 publicly-traded 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 constraining words appears to be better at capturing qualitative information about financial constraints.

Fourth, our approach is fundamentally different in how we use qualitative information to gauge financial constraints. 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 paper complementary to ours, Hoberg and Maksimovic (2014) 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 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.[3] 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*, Hoberg and Maksimovic 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 Hoberg and Maksimovic (2014) paper takes a completely different approach than ours in using 10-K text to identify financially constrained firms. Hoberg and Maksimovic (2014) 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 constraining words within the entire 10-K.[4] We believe the tone of managers' words captures 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.

The remainder of the paper is organized as follows. Section II introduces the data and variables. Section III reports empirical results. A brief conclusion follows.

[3] 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 more precise tone measures.

[4] Another point of differentiation is that Hoberg and Maksimovic (2014), beyond measuring constraints, also examine the degree to which the constraint wedge is more binding in debt markets versus equity markets.

II. Data

A. The 10-K Sample

We download all 10-K, 10-K405, 10KSB, 10-KSB, and 10KSB40 filings, excluding amended documents, from the SEC's Electronic Data Gathering, Analysis, and Retrieval (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 Center for Research in Security Prices (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. To facilitate the ability of others to use percent constraining words as a possible measure of financial constraints, we provide the list of 184 constraining words on our website.[5]

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 whether there is a dividend omission, dividend increase, equity recycling, or underfunded pensions 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 Exxon Mobil on the June 2011 formation date, we use the firm's 10-K filed on February 25, 2011. Since Exxon Mobil had higher dividends during July

[5] http://www.nd.edu/~mcdonald/Word_Lists.html. See specifically the section labeled "Sentiment Word Lists".

2011 to June 2012 than during the prior year (i.e., July 2010 to June 2011), the firm has a *Dividend Increase Dummy* value of one for the year 2011.

Our methodology is structured to minimize potential hardwiring between the dependent and independent variables. Hadlock and Pierce (2010) criticize the endogenous nature underlining KZ's sample. Since KZ incorporate the same information into both the dependent and independent variables, HP assert that KZ's coefficients are "uninformative" and "potentially misleading."

Specifically, if the dependent variable is a financial constraint dummy variable created by examining a firm's discussion of its dividend policy, an independent variable should not be the actual scaled dividend payout of the firm. That is, if a firm reports a cut in dividends (because it is implicitly financially constrained), the inclusion of the same year payout policy as an explanatory variable will hardwire the relation. HP's Table 3 regressions illustrate the impact of this problem. To sidestep this flaw, HP specifically ignore information in the 10-K relating to cash holdings, recent dividends, and recent repurchases.

To address this issue, our paper (1) uses current year percentage constraining words on next year's numbers and (2) examines constraining word usage in the entire 10-K versus specifically around words/phrases like dividends, repurchases, equity issuance, and underfunded pensions. Even so, it is important to mention that there is a potential linkage between the percentage of constraining words and our Compustat-based variables of equity recycling and underfunded pensions due to a slight overlap in time.

As an example, we use the percentage of constraining words in Delta Air Lines' 10-K filed on February 24, 2010 as an explanatory variable for Delta's equity recycling and underfunded pension variables filed on February 16, 2011. Since Delta has a December 31 fiscal year end, there is a 55 day window (between January 1, 2010 and February 24, 2010) in which events could potentially influence manager's word selections.

We do not believe, however, that this overlap is problematic. First, much of the 10-K pertains to discussion of past operating performance. Second, we do not tabulate constraining words only around phrases like equity issuance or underfunded pensions, and instead parse the entire 10-K document for constraining words.

B. Parsing the 10-K Filings and Percentage of Constraining Words

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 constraining word list. In Appendix B, we provide a detailed discussion of how the 10-Ks are parsed.

We then measure the tone of the document as the percentage of constraining words (*% Constraining)*. The constraining word list is created by examining all words appearing in at least 5% of all 10-Ks. Our entire constraining list of 184 words is reported in Appendix C. We select words judged to be constraining if in the majority of cases the word has constraining meaning when used in the annual report (e.g., Form 10-K). This is the primary qualitative measure of financial constraints we employ in our paper.

C. 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 Appendix A.

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 (see Harford (1999), Opler, Pinkowitz, Stulz, and Williamson (1999), and Acharya, Davydenko, and Strebulaev (2012)). 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 inputs (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 or improvement of external financing conditions: (1) dividend omissions, (2) dividend increases, (3) equity recycling, and (4) underfunded pensions. Our dependent variables are defined as follows. *Dividend Omission Dummy* is set to one if the firm completely omits paying a dividend during the following year, else zero. *Dividend Increase Dummy* takes a value of one if the firm has a higher 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 issuing a dividend in the year before the June formation date are assigned a value for the *Dividend Omission* and *Dividend Increase Dummies.* We obtain the dividend information from CRSP. As should be expected, the abnormal prior year returns for firms with a dividend omission is -18.57% while the market-adjusted returns in the prior year for companies who increase their dividends is 11.50%.

Equity Recycling is (cash dividends plus purchase of common and preferred stock in year t+1) / (sale of common and preferred stock in year t+1) divided by total assets in year t. Our use of the equity recycling variable is motivated by the work of Farre-Mensa, Michaely, and Schmalz (2015). *Underfunded Pension Dummy* is set to one if the firm has an underfunded pension plan (e.g., item PBPRO (projected pension benefit obligation) is greater than item PPLAO (pension plan assets)) during the next year, else zero. See Appendix A 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 constraining words is higher in the later period (0.74% versus 0.65%) 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.

After the beginning of financial crisis in the fall of 2008, where even AAA-rated firms like General Electric had liquidity problems, we would expect that the text of the 10-Ks would have a more constraining tone. Figure 1 reports the time series trend in *% Constraining* during our time period. Two obvious spikes in the frequency of constraining words occur.

First, in the period following the terrorist attacks of September 11, 2001 and subsequent economic slowdown, there is a sharp increase in constraining word usage. Second, for 10-Ks filed after the financial meltdown in the fall of 2008, there is a dramatic increase in the relative usage of constraining words. Some have argued that annual reports do not change much due to boiler plating of large sections of the document. However, Figure 1 provides strong evidence that manager's tone does change to reflect the economic environment in which companies operate.

As shown in Panel A of Table 2, our liquidity events vary in frequencies. Of the firms issuing dividends, 3.7% omit dividends in the following year while a substantial 57.2% increase dividends. On average, 70.9% of firms with available Compustat information on pension assets and projected obligations have underfunded pensions 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 constraining words and other variables used to measure financial constraints are quite low. For example, the correlation between percent constraining and excess prior returns is only 0.018. Thus, a constraining 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 Hoberg and Maksimovic (2014) 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, *% Constraining* has a relatively low correlation with market capitalization (0.036). 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.[6] 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

[6] 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.

when market capitalization is not used as a control variable, the SA and WW indexes do not help predict all of the liquidity events.

D. Constraining Words

When using a word list, it is important to identify which words drive most of the analysis. As noted by others, word counts from any dictionary often follow a power law probability distribution. This power law pattern for word counts is commonly referred to as Zipf's Law. Table 3 reports the fifty most frequently occurring constraining words in our 10-K sample. These fifty words represent 94.97% of the counts for the constraining words which appeared in 10-Ks. Only seven words, *required, obligations, requirements, require, impairment, obligation,* and *requires*, account for more than half of all the tabulated words. As should be expected, the constraining words *required* (16.86% of total) and *obligations* (9.72% of total) more commonly appear in 10-Ks than *permissible* and *encumbrance* (both less than 0.24% of all occurring constraining words).

III. Empirical Findings

A. 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 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 or improvement of external financing conditions. Relating current financial constraints to future liquidity 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. Farre-Mensa and Ljungqvist (2015) point out that firms facing financial constraints should be less likely to engage in equity recycling, i.e., simultaneous raising and paying out of equity. Rauh (2006) shows that financially constrained firms may have difficulty funding pension obligations to their retirees which subsequently undermines their ability to undertake investments. We therefore investigate how well measures of financial constraints predict future dividend omission and increases, equity recycling, and underfunded pension plans.

As a first step, how well do the KZ, SA, WW indexes and *% Constraining* 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 constraining 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 omission, dividend increases, and underfunded pensions are logits while the regression for equity recycling is ordinary least squares (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, traditional measures of financial constraints do a poor job of predicting the ex post liquidity events even when isolated from the inclusion of standard macro-finance variables. The SA index has a significant coefficient value for only the dividend omissions. The KZ index is significant and has the expected sign for only equity recycling and underfunded pensions. The SA and WW indexes have no link with equity recycling or underfunded pensions. As noted earlier, the poor ability of the SA and WW indexes to predict equity recycling has already been documented by Farre-Mensa and Ljungqvist (2015). The two authors sharply criticize the existing literature's measures of financial constraints. In contrast, the percentage of constraining words is significantly linked with all four liquidity events.

The fact that existing measures of constraints are quite imperfect may not be surprising for all readers. One could argue that the HP and WW papers illustrate the difficulties in consistently measuring constraints across time and samples, rather than as definitive indexes of the proper way to measure constraints. At its core, the notion of financial constraints is a delicate and nuanced concept.

B. Control Variables

Will the predictive power of the percentage of constraining words 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 times 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 Appendix A. From Table 5 we observe that 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. As before, an intercept, Fama and French 48-industry dummies, and calendar year dummies are included in all regressions. The *z*-statistics (or *t*-statistics for the OLS regressions) are in parentheses with the standard errors clustered by both year and industry.

C. Dividend Omissions

In the first two columns of Table 5, the dependent variable is the *Dividend Omission Dummy*. Since only firms with at least one dividend distribution in the prior year are included in the first two regressions, the sample size is 12,669 firm-year observations. The control variables imply that larger and better performing companies are less likely to stop paying dividends, whereas companies with negative earnings or value firms (i.e., high book-to-market value) are more likely to do so.

When *% Constraining words* is added as an independent variable in column (2), its coefficient is positive (1.318) with a statistically significant coefficient value (*z*-statistic of 3.80). More constraining 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 0.023 while the standard deviation of percent constraining is 0.166. Thus, a one standard deviation increase in percentage of constraining words is associated with a 10.32% (=0.023*0.166 / dividend omission sample mean of 0.037) higher chance of a dividend omission.

D. Dividend Increases

In columns (3) and (4) of Table 5, the dependent variable is the *Dividend Increase Dummy* (equal to one if the firm increased its dividend during the following year, else zero). As before, only firms with at least one dividend distribution in the prior year are included in the regressions. The column (3) regression includes only the control variables. It is worth noting that company market capitalization and past performance are related—as expected—to dividend increase: larger and better performing companies are more likely to increase dividends. Value firms and companies with negative trailing earnings are less likely to increase their dividends.

In column (4), the coefficient on *% Constraining* (-0.733) is statistically significant at the 1% level. This implies that as the percentage of constraining words in the 10-K rises, the likelihood of the firm increasing its dividend decreases. The marginal effect of the coefficient is -0.178 while the standard deviation of percent constraining is 0.166. Thus, a one standard deviation increase in percentage of constraining words is related to a -6.46% (=-0.178*0.166 / dividend increase sample mean of 0.4572) smaller likelihood that a firm will increase its dividend in the subsequent year.

E. Equity Recycling

Our third liquidity event is equity recycling. Farre-Mensa and Ljungqvist (2015) also use this variable to determine whether or not a firm is financially constrained. One would expect that, after controlling for other effects, financially constrained firms should be less likely than unconstrained firms to pay dividends or repurchase shares with the capital obtained from issuing equity. This variable is winsorized at 1% and is scaled by total assets. Since we require firms to have a non-missing value for the sale of common and preferred stock in the following year, our sample size is 36,477 firm-year observations for the column (5) and (6) regressions.

In the last two columns of Table 5 regressions, equity recycling is the dependent variable. For this liquidity event, the financial constraint measures should have a negative coefficient value. That is, more constrained firms should be engaging in equity recycling to a lesser degree. With only the control variables in the column (5) regression, smaller firms and companies with positive trailing earnings are more likely to engage in equity recycling.

In column (6), the coefficient on equity recycling is negative (-0.084) with a significant *t*-statistic of -4.70. More relative usage of words like *required, obligations, commitment,* and *restricted* significantly lowers the amount of equity recycling done by the company. This is consistent with the percentage of constraining words serving as a direct proxy for the firm's level of financial constraints. A one standard deviation increase in constraining words is associated with a 23.24% (=-0.084 * 0.166 / equity recycling sample mean of 0.060) lower equity recycling value.

F. Underfunded Pension Plans

Using the IRS Form 5500 filing by firms from the Department of Labor during 1990-1998, Rauh (2006) finds that required pension contributions have a negative effect on a firm's capital expenditures. About 25% of the firms in his sample experience at least one year when required pension contributions are at least 10% of its annual capital expenditures. For our paper, we will use available Compustat pension data to proxy for the level of financial constraints faced by firms with underfunded pension benefit plans.

In Table 6, the dependent variable, *Underfunded Pension Dummy*, is set to one if the firm has an underfunded pension plan (e.g., item PBPRO (pension projected benefit obligation) is greater than item PPLAO (pension plan assets)) during the next year, else zero. Each regression in Table 6 has 11,915 firm-year observations since we require non-missing values for pensions plan assets and pension projected benefit obligations in year t and t+1. Many large, old, established companies, like Delta Air Lines, Eastman Kodak, Black & Decker, Caterpillar, Unisys Corp., Bethlehem Steel, and Maytag Corp., have underfunded pension plans for numerous years during our sample period. Having an underfunded pension benefit plan should restrict the ability of these firms to invest in positive net present value projects.

From the descriptive statistics in Table 2 we observe that pension underfunding occurs in about 70.93% of firm-year observations. While this number may appear to be large, it is consistent with estimates obtained in other post-World War II studies (e.g., Ippolito (1986)) and is likely to be significantly affected by firms' strategic behavior to increase their bargaining power vis-à-vis labor unions. Ippolito (1985) argues that underfunded pension plans help to resolve the hold-up problem between the firm and its employees. Benmelech, Bergman, and Enriquez (2012) show that airlines obtain wage concessions from employees whose pension plans are underfunded. In addition, the falling discount rates used to determine the projected benefit obligations have pushed the majority of companies with defined benefit plans into an underfunded status.

Given that many firms have underfunded pensions for several years in a row, in the Table 6 regressions, we add a dummy variable set to one if the firm has an underfunded pension in year t. In column (1), the only control variable that is significant is the *Underfunded Pension Dummy* in year t. As expected, firms with underfunded pension benefit plans in year t will likely continue to have underfunded plan in the following year. When *% Constraining* is added to the column (2) regression, its coefficient is positive and statistically significant at the 1% level. More constraining text devoted in the 10-K increases the likelihood of the firm having underfunded pensions in the next year after controlling for whether the company currently has underfunded pensions. A one standard deviation larger usage of constraining words is related to 2.34% (= marginal effect of the coefficient is 0.101 * 0.164 / Underfunded Pension Dummy sample mean of 0.709) higher instance of underfunded pension liabilities in the next year.

Is this economic impact large or small? Since pension funding status is highly correlated over time (first order autocorrelation coefficient of 0.66) the best predictor of firm's future pension funding status is its current funding status – firms which have underfunded pension obligations in the current year are 76.33% more likely to have underfunded obligation in the next year than firms with currently fully funded pension obligations.[7] However, if we assess the economic impact of other controls, e.g., *Negative Earnings Dummy* (-1.14%), they all are smaller than the effect of the constrained word usage and mostly have the wrong sign.

As a robustness check, the last column of Table 6 reports that *% Constraining* remains significant if market value of equity and prior year *Underfunded Pension Dummy* are not included as control variables in the regression. In unreported results we also performed our analysis by setting *Underfunded Pension Dummy* to one if predicted pension obligations exceed pension plan assets by 5% or 10%. Our results remain essentially unchanged.

As noted before, the autocorrelation of *Underfunded Pension Dummy* is 0.66. With the exception of *Dividend Omission* (-0.02), the autocorrelations of the dependent variables in Table

[7] This economic impact corresponds to approximately 2.13 standard deviations variation in the *Underfunded Pension Dummy*.

5 are also relatively high (*Dividend Increase*=0.55*, Equity Recycling*=0.36). The *% Constraining* variable has an autocorrelation of 0.58. These levels would suggest that the significance of the *% Constraining* variable is most likely influenced by cross-sectional variation. Some evidence of this is provided by re-estimating the regressions in Tables 5 and 6 with firm fixed effects (and clustered standard errors). When using this specification, the *% Constraining* variable is not significant in any of the models. This result supports the contention that most of the relevant variation in our measure, like other measures of financial constraints, is in the cross-section.[8]

It is also important to underscore the explanatory ability of the *% Constraining* variable. In all the Table 5 and 6 regressions, the r-squared values, while relatively high by corporate finance standards, are still low in absolute terms. Although *% Constraining* adds a unique and significant contribution to the measurement of financial constraints, we are certainly far from done in creating a measure that accurately captures financial constraints with a high degree of precision.

IV. 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 percentage of constraining words in the 10-K text, and not solely accounting characteristics, to help gauge which firms will become financially constrained. To measure the level of financial constraints faced by the firm,

[8] Hadlock and Pierce (2010) note that their SA index will be more effective in the cross-section than in detecting time-series variation in constraints.

we create a list of 184 constraining words. Examples of constraining words include *required, obligations, impairment,* and *covenants*.

Unlike other measures, the frequency of constraining words is more likely to identify inflection points. For example, it is commonly believed that as firms get larger, they typically become less financially constrained. Our variable, percentage of constraining 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.

We test the ability of percentage of constraining 10-K words to predict four liquidity events after controlling for standard firm characteristics. The ex post liquidity events are dividend omissions or increases, equity recycling, and underfunded pension plans. With or without the presence of firm characteristic control variables, percentage of constraining words helps explain the subsequent liquidity events. The percentage of constraining words also has a nontrivial economic impact. For example, a one standard deviation increase in constraining words increases the likelihood of a dividend omission by 10.32% and decreases the probability of a dividend increase by 6.46%.

In contrast to percent constraining words, the KZ, SA, and WW indexes, in univariate regressions, do an imperfect job at predicting the four liquidity events. The limited explanatory power of the KZ, SA, and WW indexes in predicting liquidity events is consistent with the evidence of Farre-Mensa and Ljungqvist (2015), who argue that the measures poorly identify firms who are actually financially constrained.

The more managers believe the firm will face constraints in the future, the more the text of the 10-K will reflect this 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) constraining word frequencies capture subtle cues from managers who may not be required to issue explicit liquidity warnings to investors. Although we focus on 10-Ks for our analysis, other possible areas where the constraining word list could be used by researchers include newspaper articles, conference calls, or press releases.

The gauging of firm level financial constraints is a critically important research area. We extend the literature by creating a constraining word list for other researchers to use as an indicator of financial constraints beyond the usual macro-finance control variables. A higher frequency of constraining 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 constraining 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. Definitions of the variables used in the paper

% Constraining	Percentage of words in the 10-K that are constraining. Examples of constraining words include *required, obligations, requirements, comply,* and *require*. See Appendix C for a complete list of our 184 constraining words.
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 Standard Industrial Classification (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.

Control Variables

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.

Liquidity Events

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. We obtain the dividend information from CRSP.
Dividend Increase Dummy	Dummy variable set to one if the firm has a higher 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 issuing a dividend in the year before the June formation date are assigned a value for this variable. We obtain the dividend information from CRSP.
Equity Recycling	The ratio of (cash dividends (item DV) + purchase of common and preferred stock (item PRSTKC))/ sale of common and preferred stock (item SSTK) during July of year t to June of year t+1 scaled by total assets in year t. This variable is winsorized at the 1% level.

Underfunded Pension Dummy	Dummy variable set to one if the firm has an underfunded pension plan (item PBPRO (pension projected benefit obligation) is greater than item PPLAO (pension plan assets)) during the next year, else zero. Only firms with non-missing pension plan assets and projected pension benefit obligations are assigned a value for this variable.

Appendix B. Parsing the 10-K filings

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.[9] All of these files are also contained in a single text file with the embedded HTML, XBRL, exhibits, and the ASCII-encoded graphic.[10] 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.

---

[9] 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.

[10] 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.

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.
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.

Appendix C. List of Constraining Words

ABIDE	DICTATING	INHIBITS	PRECLUDED	STIPULATE
ABIDING	DIRECTIVE	INSIST	PRECLUDES	STIPULATED
BOUND	DIRECTIVES	INSISTED	PRECLUDING	STIPULATES
BOUNDED	EARMARK	INSISTENCE	PRECONDITION	STIPULATING
COMMIT	EARMARKED	INSISTING	PRECONDITIONS	STIPULATION
COMMITMENT	EARMARKING	INSISTS	PRESET	STIPULATIONS
COMMITMENTS	EARMARKS	IRREVOCABLE	PREVENT	STRICT
COMMITS	ENCUMBER	IRREVOCABLY	PREVENTED	STRICTER
COMMITTED	ENCUMBERED	LIMIT	PREVENTING	STRICTEST
COMMITTING	ENCUMBERING	LIMITING	PREVENTS	STRICTLY
COMPEL	ENCUMBERS	LIMITS	PROHIBIT	UNAVAILABILITY
COMPELLED	ENCUMBRANCE	MANDATE	PROHIBITED	UNAVAILABLE
COMPELLING	ENCUMBRANCES	MANDATED	PROHIBITING	
COMPELS	ENTAIL	MANDATES	PROHIBITION	
COMPLY	ENTAILED	MANDATING	PROHIBITIONS	
COMPULSION	ENTAILING	MANDATORY	PROHIBITIVE	
COMPULSORY	ENTAILS	MANDITORILY	PROHIBITIVELY	
CONFINE	ENTRENCH	NECESSITATE	PROHIBITORY	
CONFINED	ENTRENCHED	NECESSITATED	PROHIBITS	
CONFINEMENT	ESCROW	NECESSITATES	REFRAIN	
CONFINES	ESCROWED	NECESSITATING	REFRAINING	
CONFINING	ESCROWS	NONCANCELABLE	REFRAINS	
CONSTRAIN	FORBADE	NONCANCELLABLE	REQUIRE	
CONSTRAINED	FORBID	OBLIGATE	REQUIRED	
CONSTRAINING	FORBIDDEN	OBLIGATED	REQUIREMENT	
CONSTRAINS	FORBIDDING	OBLIGATES	REQUIREMENTS	
CONSTRAINT	FORBIDS	OBLIGATING	REQUIRES	
CONSTRAINTS	IMPAIR	OBLIGATION	REQUIRING	
COVENANT	IMPAIRED	OBLIGATIONS	RESTRAIN	
COVENANTED	IMPAIRING	OBLIGATORY	RESTRAINED	
COVENANTING	IMPAIRMENT	OBLIGE	RESTRAINING	
COVENANTS	IMPAIRMENTS	OBLIGED	RESTRAINS	
DEPEND	IMPAIRS	OBLIGES	RESTRAINT	
DEPENDANCE	IMPOSE	PERMISSIBLE	RESTRAINTS	
DEPENDANCES	IMPOSED	PERMISSION	RESTRICT	
DEPENDANT	IMPOSES	PERMISSIONS	RESTRICTED	
DEPENDENCIES	IMPOSING	PERMITTED	RESTRICTING	
DEPENDENT	IMPOSITION	PERMITTING	RESTRICTION	
DEPENDING	IMPOSITIONS	PLEDGE	RESTRICTIONS	
DEPENDS	INDEBTED	PLEDGED	RESTRICTIVE	
DICTATE	INHIBIT	PLEDGES	RESTRICTIVELY	
DICTATED	INHIBITED	PLEDGING	RESTRICTIVENESS	
DICTATES	INHIBITING	PRECLUDE	RESTRICTS	

References

Acharya, V.; S. Davydenko; and I. Strebulaev. "Cash Holdings and Credit Risk." *Review of Financial Studies*, 25 (2012), 3572–3609.

Antweiler, W., and M. Frank. "Is all that Talk just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance*, 59 (2004), 1259-1293.

Battacharya, S. "Imperfect Information, Dividend Policy and the 'Bird in the Hand' Fallacy." *Bell Journal of Economics*, 10 (1979), 259-270.

Benmelech, E.; N. K. Bergman; and R. J. Enriquez. "Negotiating with Labor under Financial Distress." *Review of Corporate Finance Studies*, 4 (2012), 28-67.

Campello, M.; J. Graham; and C. Harvey. "The Real Effects of Financial Constraints: Evidence from a Financial Crisis." *Journal of Financial Economics*, 97 (2010), 470-487.

Cleary, S. "The Relationship between Firm Investment and Financial Status." *Journal of Finance*, 54 (1999), 673-692.

Doidge, C.; A. Karolyi; and R. Stulz. "The U.S. left Behind? Financial Globalization and the Rise of IPOs Outside the U.S." *Journal of Financial Economics*, 110 (2013), 546-573.

Fama, E., and K. French. “The Cross-Section of Expected Stock Returns.” *Journal of Finance*, 47 (1992), 427-465.

Fama, E., and K. French. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics*, 33 (1993), 3-56.

Fama, E., and K. French. “Industry Costs of Equity.” *Journal of Financial Economics*, 43 (1997), 153–193.

Farre-Mensa, J., and A. Ljungqvist. “Do Measures of Financial Constraints Measure Financial Constraints?” Working Paper, Harvard University (2015).

Farre-Mensa, J.; R. Michaely; M. Schmalz. “Financing Payouts.” Working Paper, Harvard University (2015).

Fazzari, S.; R. Hubbard; and B. Petersen. “Financing Constraints and Corporate Investment.” *Brookings Papers on Economic Activity*, 1 (1988), 141-206.

Fazzari, S.; R. Hubbard; and B. Petersen. “Investment-Cash Flow Sensitivities are Useful: A Comment on Kaplan and Zingales.” *Quarterly Journal of Economics*, 115 (2000), 695–705.

Gao, X.; J. Ritter; and Z. Zhu. “Where have all the IPOs Gone?” *Journal of Financial and Quantitative Analysis*, 48 (2013), 1663-1692.

Hadlock, C., and J. Pierce. "New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index." *Review of Financial Studies*, 23 (2010), 1909–1940.

Hanley, K., and G. Hoberg. "The Information Content of IPO Prospectuses." *Review of Financial Studies*, 23 (2010), 2821-2864.

Hanley, K., and G. Hoberg. "Litigation Risk, Strategic Disclosure and the Underpricing of Initial Public Offerings." *Journal of Financial Economics*, 103 (2012), 235–254.

Harford, J. "Corporate Cash Reserves and Acquisitions." *Journal of Finance*, 54 (1999), 1969-1997.

Hennessy, C., and T. Whited. "How Costly is External Financing? Evidence from a Structural Estimation." *Journal of Finance*, 62 (2007), 1705-1745.

Hoberg, G., and V. Maksimovic. "Redefining Financial Constraints: A Text-Based Analysis." Forthcoming in *Review of Financial Studies*, (2014).

Ippolito, R.A. "The Economic Function of Underfunded Pension Plans." *Journal of Law and Economics*, 28 (1985), 611–651.

Ippolito, R.A. “The Economic Burden of Corporate Pension Liabilities.” *Financial Analysts Journal*, 42 (1986), 22–34.

Kaplan, S., and L. Zingales. “Do Financing Constraints Explain why Investment is Correlated with Cash Flow?” *Quarterly Journal of Economics*, 112 (1997), 169–216.

Lamont, O.; C. Polk; and J. Saa´-Requejo. “Financial Constraints and Stock Returns.” *Review of Financial Studies*, 14 (2001), 529–544.

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

Miller, M. “The Modigliani-Miller Propositions after Thirty Years.” *Journal of Economic Perspectives*, 2 (1988), 99-120.

Myers, S., and N. Majluf. “Corporate Financing and Investment Decisions When Firms have Information that Investors do not have.” *Journal of Financial Economics*, 13 (1984), 187–221.

Opler T.; L. Pinkowitz; R. Stulz; and R. Williamson. “The Determinants and Implications of Corporate Cash Holdings.” *Journal of Financial Economics*, 52 (1999), 3-46.

Rauh, J. “Investment and Financing Constraints: Evidence from the Funding of Corporate Pension Plans.” *Journal of Finance*, 61 (2006), 33-71.

Tetlock, P. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance*, 62 (2007), 1139–1168.

Tetlock, P.; M. Saar-Tsechansky; and S. Macskassy. "More than Words: Quantifying Language to Measure Firms' Fundamentals." *Journal of Finance*, 63 (2008), 1437-1467.

Townsend, R. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, 21 (1979), 265–293.

Whited, T. "What can Cash Shortfalls and Windfalls tell us about Finance Constraints?" In Giorgio Calcagnini and Enrico Saltari (eds.) The Economics of Imperfect Markets, (2009), 17–32.

Whited, T., and G. Wu. "Financial Constraints Risk." *Review of Financial Studies*, 19 (2006), 531–559.

FIGURE 1

Time Series of the % Constraining Words in 10-Ks

The graph in Figure 1 plots the average % Constraining words in 10-Ks for each year in our sample, 1997-2011.

TABLE 1
Sample Creation

Table 1 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

Table 2 reports descriptive statistics by time period in Panel A; correlations between key variables of interest are reported in Panel B. In Panel A, the sample sizes vary by the availability of data. For *SA Index, % Constraining, 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 increase* since firms are required to be distributing dividends in the prior year to be included. The sample is 36,477 for the equity recycling variable due to the requirement of subsequent, non-zero sale of common and preferred stock data item. See Appendix A for detailed variable definitions.

Panel A: Mean Summary Statistics

Variable	(1) 1997 to 2003	(2) 2004 to 2011	(3) 1997 to 2011
% Constraining	0.65	0.74	0.69
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 Omission Dummy	3.96%	3.38%	3.65%
Dividend Increase Dummy	49.81%	63.77%	57.17%
Equity Recycling	6.89%	5.05%	6.00%
Underfunded Pension Dummy	53.44%	88.45%	70.93%
Market Capitalization	$2,131.1	$3,464.1	$2,736.0
Book-to-Market	0.79	0.65	0.73
Excess Prior Returns	0.72%	6.74%	3.45%
Negative Earnings Dummy	38.52%	33.88%	36.42%

Panel B: Correlations

	% Constraining	*KZ Index*	*SA Index*	*WW Index*	*Log(Mkt Cap)*	*Log (Book-to-Market)*	*Excess Prior Returns*
KZ Index	0.072						
SA Index	-0.063	-0.145					
WW Index	-0.050	-0.055	0.832				
Log(Mkt Cap)	0.036	-0.028	-0.702	-0.839			
Log(Book-to-Market)	0.068	0.150	-0.123	-0.041	-0.357		
Excess Prior Returns	0.018	-0.007	-0.028	-0.019	0.176	-0.338	
Negative Earnings Dummy	0.098	0.010	0.361	0.426	-0.343	0.015	-0.070

TABLE 3
Fifty Most Frequently Occurring *Constraining* Words in 10-Ks

Word	% of total	Cumulative %	Word	% of total	Cumulative %
REQUIRED	16.86%	16.86%	REQUIRING	0.86%	84.48%
OBLIGATIONS	9.72%	26.57%	DEPENDS	0.67%	85.15%
REQUIREMENTS	9.21%	35.78%	ESCROW	0.64%	85.79%
REQUIRE	5.37%	41.16%	PLEDGE	0.61%	86.40%
IMPAIRMENT	4.45%	45.61%	COMMITTED	0.61%	87.01%
OBLIGATION	4.16%	49.77%	PROHIBITED	0.61%	87.62%
REQUIRES	3.88%	53.64%	IMPOSE	0.60%	88.22%
PERMITTED	3.24%	56.89%	BOUND	0.59%	88.81%
RESTRICTED	2.83%	59.72%	LIMITING	0.58%	89.39%
COVENANTS	2.79%	62.51%	PLEDGED	0.55%	89.94%
COMMITMENTS	2.45%	64.97%	RESTRICT	0.48%	90.42%
RESTRICTIONS	2.44%	67.41%	IMPAIR	0.45%	90.88%
COMPLY	2.30%	69.71%	MANDATORY	0.45%	91.33%
LIMIT	1.85%	71.55%	IRREVOCABLY	0.45%	91.78%
DEPENDENT	1.36%	72.91%	RESTRICTION	0.37%	92.15%
COMMITMENT	1.35%	74.26%	PROHIBIT	0.37%	92.52%
IMPAIRED	1.22%	75.48%	RESTRICTIVE	0.36%	92.88%
IMPOSED	1.15%	76.63%	IRREVOCABLE	0.36%	93.24%
PREVENT	1.09%	77.73%	IMPAIRMENTS	0.30%	93.54%
REQUIREMENT	1.08%	78.80%	IMPOSITION	0.28%	93.82%
LIMITS	1.07%	79.87%	PRECLUDE	0.24%	94.06%
OBLIGATED	1.05%	80.92%	PROHIBITS	0.23%	94.29%
DEPEND	0.97%	81.89%	ENCUMBRANCES	0.23%	94.52%
COVENANT	0.87%	82.76%	PERMISSIBLE	0.23%	94.75%
DEPENDING	0.86%	83.62%	ENCUMBRANCE	0.22%	94.97%

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 *% Constraining* 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 word list 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 Omission Dummy*	*Dividend Increase Dummy*	*Equity Recycling*	*Underfunded Pension*
KZ Index			X	X
SA Index	X			
WW Index	X	X		
% Constraining Words	X	X	X	X

TABLE 5

Logit Regressions with Three Different Liquidity Events as the Dependent Variable, 1997-2011

The column (1) and (2) 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. The column (3) and (4) dependent variable, *Dividend Increase Dummy*, is set to one if the firm increases its 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. The dependent variable in columns (5) and (6), *Equity Recycling*, is defined as cash dividends plus purchase of common and preferred stock/ sale of common and preferred stock scaled by trailing total assets. *% Constraining* is the percentage of constraining words in the 10-K. All regressions include an intercept, Fama and French 48-industry dummies, and calendar year dummies. See Appendix A for definitions of all other variables. The *z*-statistics (or *t*-statistic) are in parentheses with standard errors clustered by year and industry.

	Dividend Omission		*Dividend Increase*		*Equity Recycling*	
	(1)	(2)	(3)	(4)	(5)	(6)
% Constraining		1.318		-0.733		-0.084
		(3.80)		(-4.30)		(-4.70)
Log(Mkt Cap)	-0.202	-0.205	0.139	0.137	-0.028	-0.028
	(-5.69)	(-5.70)	(5.19)	(5.03)	(-11.27)	(-11.20)
Log(Book-to-Market)	0.368	0.358	-0.404	-0.397	0.004	0.005
	(3.49)	(3.50)	(-5.49)	(-5.36)	(1.06)	(1.38)
Excess Prior Returns	-1.454	-1.413	0.615	0.630	-0.004	-0.004
	(-5.27)	(-5.16)	(5.89)	(6.10)	(-1.50)	(-1.39)
Negative Earnings Dummy	1.144	1.100	-0.992	-0.960	-0.068	-0.065
	(10.59)	(9.76)	(-11.20)	(-10.18)	(-7.44)	(-7.55)
Pseudo R^2	18.06%	18.48%	11.83%	12.01%	5.40%	5.63%
Sample size	12,669	12,669	12,806	12,806	36,477	36,477

TABLE 6
Logit Regressions with Underfunded Pension Dummy as the Dependent Variable, 1997-2011

The dependent variable, *Underfunded Pension Dummy*, is set to one if the firm has an underfunded defined benefit pension plan during July of year t to June of year t+1, else zero. Only firms with non-missing pension assets and projected pension obligations are included in these regressions. *% Constraining* is the percentage of constraining words in the 10-K. All regressions include an intercept, Fama and French 48-industry dummies, and calendar year dummies. See Appendix A for definitions of all other variables. The *z*-statistics are in parentheses with standard errors clustered by year and industry.

	(1)	(2)	(3)
% Constraining		0.709	0.914
		(4.26)	(3.74)
Underfunded Pension Dummy(t)	3.788	3.780	
	(10.43)	(10.42)	
Log(Mkt Cap)	0.070	0.077	
	(1.45)	(1.60)	
Log(Book-to-Market)	0.072	0.067	-0.008
	(1.43)	(1.33)	(-0.11)
Excess Prior Returns	-0.024	-0.029	-0.021
	(-0.64)	(-0.76)	(-0.47)
Negative Earnings Dummy	-0.116	-0.146	-0.221
	(-1.13)	(-1.43)	(-1.87)
Pseudo R^2	51.15%	51.23%	27.90%
Sample size	11,915	11,915	11,915