

When is a Liability not a Liability?

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

May 6, 2009

Abstract: Previous research uses negative word counts as a way to measure the tone of a text. We show that word lists developed for other disciplines substantially misclassify financial text. In a large sample of 10-Ks during 1994-2007, more than half of the word count identified as negative by the commonly used Harvard-IV-4 Psychological Dictionary represents words that typically do not have negative financial meaning. Words like *tax, cost, board, foreign, vice, and liability*, simply describe company operations. We develop an alternative negative word list that better reflects the tone of financial text, and examine five other word classifications (positive, uncertain, litigious, strong modal words, and weak modal words). Our new word classification measures are significantly related to 10-K file date returns, post-file date returns, return volatility, allegations of accounting fraud, and company material weakness disclosures. We also provide evidence that word choice in 10-Ks is affected by recent legislation.

Key words: Textual analysis; Harvard Psychological Dictionary; negative word counts.

*We are indebted to Paul Tetlock for comments on a previous draft. We also thank Robert Battalio, Peter Easton, Paul Gao, Campbell Harvey (editor), Jennifer Marietta-Westberg, Paul Schultz, an anonymous referee, and seminar participants at the University of Notre Dame, for helpful comments.

When is a Liability not a Liability?

A growing body of finance and accounting research uses textual analysis to examine the tone and sentiment of corporate 10-K reports, newspaper articles, press releases, and investor message boards. Examples are Antweiler and Frank (2004), Tetlock (2007), Engelberg (2008), Li (2008), and Tetlock, Saar-Tsechansky, and Macskassy (2008). The results to date indicate that negative word classifications can be effective in measuring tone, as reflected by significant correlations with other financial variables.

A commonly used source for word classifications is the Harvard Psychological Dictionary, specifically the Harvard-IV-4 TagNeg (H4N) file. One positive feature of this list for research is that its makeup is beyond the control of the researcher. That is, the researcher cannot pick and chose which words have negative implications. Yet English words have many meanings, and a particular word categorization scheme will drive many of the conclusions from textual analysis.

In one of the first systematic examples of textual analysis, Berelson (1952) notes that "Content analysis stands or falls by its categories. Particular studies have been productive to the extent that the categories were clearly formulated and well adapted to the problem" (p. 92). In some contexts, the H4N list of negative words may work well to capture the tone of a text. The critical question, however, is whether a word list developed for psychology and sociology translates well into the realm of business.

While measuring document tone using any word classification scheme is inherently problematic, we provide evidence that the H4N list may be imprecise when gauging tone in financial applications. We use both the original Harvard negative word list (H4N) and the Harvard list extended to include inflections (H4N-Inf) to examine the tone of 10-Ks.

Inflections expand possible iterations of a root word. *Depreciate*, for example, is a negative word in the original H4N list, and its inflections include *depreciates*, *depreciated*, *depreciation*, and *depreciating* in the H4N-Inf list. There are 2,004 negative words in the H4N list and 4,183 words in H4N-Inf. Although we initially will consider both the H4N list and its inflected version (H4N-Inf), in our comparative test we focus on the inflected list.

We make several contributions in our analysis of 41,842 firm-year 10-Ks between 1994 and 2007. First, we find that more than half of the negative word counts according to the H4N-Inf list are attributable to words that are typically not negative in a financial context. Words such as *tax*, *cost*, *capital*, *board*, *liability*, *foreign*, and *vice* are on the H4N list. These words also appear in the vast majority of 10-Ks, yet often do no more than name a *board* of directors or a company's *vice*-presidents. Other words on the H4N list, such as *mine* or *capital*, are more likely to identify a specific industry than reveal negative financial events.

Our second contribution is to create a list of some 1,200 words that typically have negative implications in a financial sense. *<All word lists we generate are reported in the reviewer's appendix and are referenced in the paper to the author's website>* Polesemy – that words have multiple meanings – makes an absolute mapping of specific words into financial sentiment impossible. We can, however, develop lists based on actual usage frequency that are most likely associated with a target construct. We use the term Fin-Neg to describe our list of negative financial words. Some of these words also appear on the H4N list, but others, like *felony*, *litigation*, *misconduct*, *misstatement*, and *unanticipated* do not appear on the Harvard list.

Our third contribution is to provide several tests to see which word classification—H4N, H4N-Inf, or the Fin-Neg—is better at gauging the tone of the 10-K financial text. One obvious method is to examine a firm's abnormal returns at the time of its 10-K filing. One might expect

firms that include a high proportion of negative words in the text to have registered lower announcement returns on the filing date. If managers frequently use negative words to describe their operations, things cannot be going well.

Dividing firms into quintiles according to the proportion of H4N and H4N-Inf negative words in their 10-Ks, we find no particular pattern. Companies with a high proportion of H4N words experience approximately the same 10-K filing announcement returns as firms with a low proportion. Yet when we use our financial negative list to sort firms, we observe a strong pattern. Stocks of firms filing 10-Ks with a low proportion of negative words according to the Fin-Neg list have positive returns; companies with a high proportion of negative words experience the reverse. Regressions with multiple control variables confirm the univariate findings of no effect for the Harvard lists versus a significant impact for the Fin-Neg list. The same findings hold true if we use normalized year-to-year proportional differences to measure tone, where differencing might serve to mitigate the misclassification problem predominant in the Harvard lists.

To expand the word classification categories, we also create five additional word lists to measure their impact on 10-K file date returns, post-file date returns, and post-file date return volatility. Specifically we consider: negative, positive, uncertain, and litigious word categories, in addition to strong and weak modal words. When we look at whether our word lists actually gauge tone, we find that some of the word category proportional counts are significantly related to the various return measures. They are also significant in logit samples for periods of alleged accounting fraud and disclosures of material weaknesses in internal control.

Our last contribution is examination of the tone of 10-Ks before and after passage of the Sarbanes-Oxley Act (SOX) in 2002. Interestingly, we find that after the implementation of SOX, firms use strong modal words less often (e.g., *will*, *can*), but increase their use of weak modal

words (e.g., *might*, *could*). This result suggests that managers dilute the information content of their firms' disclosures in response to regulations that expose them to greater legal liability. This finding highlights the problem that stricter regulation may produce sterile disclosures.

When does a negative word have negative meaning? In some studies, negative words are identified by using the Harvard-IV-4 TagNeg (H4N) word list. If H4N words like *tax*, *liability*, *foreign*, or *mine* appear proportionally more frequently in the text, the document is classified as pessimistic. Our primary methodological contribution is to challenge the implicit assumption that the magnitude of misclassifications using the Harvard list is always reasonable. Our word list appears to reduce measurement error when gauging the tone of financial text, which in turn will increase the power of textual analysis methods in this field. In the case of finance and accounting, it is best to avoid use of word classifications schemes that come from outside the business domain.

I. Literature

Some of the literature using textual analysis in finance and accounting focuses on alternative techniques used to analyze text. Das and Chen (2001) develop a process based on five separate algorithms and 300 "hand-selected" words to extract investor sentiment from message postings. Antweiler and Frank (2004) use a Naïve Bayesian algorithm to classify stock board messages into buy, sell, or hold signals. Using forward looking statements in the Management Discussion and Analysis section, Li (2009) uses a Naïve Bayesian machine learning approach to gauge content. Li (2009) finds that the tone of the MD&A section using sentence level analysis has predictive power for the firm's future earnings.

Most authors use external word lists to classify text. Harvard's General Inquirer (GI) is a common source of word categories. The General Inquirer has 182 tag categories. Examples

include positive, negative, strong, weak, active, pleasure, and even pain categories. Finance and accounting researchers generally focus on the GI's Harvard IV-4 negative and positive word categories, although none seems to find much incremental value in the positive word lists.

The limitations of positive words in prior tests, as noted by others, is likely attributable to their frequent negation. It is common to see the framing of negative news using positive words ("did not benefit"); rarely do corporate communications include positive news relayed in negated negative words ("not downgraded").

Table I summarizes word classification sources in studies using textual analysis. While not every work uses the Harvard negative word list (H4N) to gauge text tone, it is a good example of the typical word classification schemes. An advantage is that the Harvard list, unlike many others, is non-proprietary, so we can assess exactly which words contribute most to the aggregate counts.

Most researchers link the tone of newspaper articles (Tetlock, Saar-Tsechansky, and Macskassy (2008) and Kothari, Li, and Short (2008)) or company press releases (Demers and Vega (2008), Engelberg (2008), and Henry (2006)) with lower firm earnings, earnings drift, or stock returns. Other authors look at a firm's 10-K or IPO prospectus (Li (2008, 2009), Hanley and Hoberg (2008), and Feldman, Govindaraj, Livnat, and Segal (2008)). The main points of both are that the linguistic content of a document has predictive power in explaining stock returns, stock volatility, or trading volume.

Perhaps the best known study is Tetlock's (2007) paper, which links the *Wall Street Journal's* popular "Abreast of the Market" column with subsequent stock returns and trading volume. Tetlock finds high levels of pessimistic words in the column precede lower returns the next day. Pessimism is initially determined by word counts using a factor derived from 77 GI

categories from the Harvard Psychosocial Dictionary. However later in his paper, Tetlock focuses on negative words and weak words as these are most highly correlated with pessimism. A number of subsequent studies are based on the H4N list (see Table I).

In our analysis, we use the proportion of words from a given word list as our empirical measure of each construct. While other studies also use this approach, some standardize the measure by looking at changes in proportional occurrence relative to a historic benchmark (see, for example, Tetlock, Saar-Tsechansky, and Macskassy (2008) and Feldman, Govindaraj, Livnat, and Segal (2008)).

Differencing or some form of standardization might have the advantage of reducing the impact of words contextually misclassified. It assumes a reader can remember the frequency of negative words in previous news articles, columns, or 10-Ks—e.g., that today’s column or 10-K has fewer negative words than previous editions, so it may convey a bullish signal.

For our initial comparative tests of the Harvard negative word list versus the negative word list developed in this paper, we consider both simple proportions and normalized year-to-year differences. In most of our tests we use simple proportions as our tone measure, which assumes readers interpret content relative to a cross-section of firms and in the context of today’s contemporary economic environment. Since differencing could potentially mitigate the degree of misclassification, which would favor the Harvard lists, we also use differencing in the initial comparisons.

II. Data and Variables

A. *The 10-K sample*

We download all forms 10-K and 10-K405, excluding amended documents, from the EDGAR website (www.sec.gov) over 1994-2007.¹ Table II shows how the original sample of 10-Ks is impacted by our data filters and data requirements. Most notably, the original sample of 113,154 10-Ks is reduced to 70,925 forms by the requirement of a CRSP PERMNO match.² Many of the firms with missing PERMNOs are real estate, non-operating income, or asset-backed partnerships/trusts that are required to file with the SEC.

We require data to be available for regression control variables (e.g., size, book-to-market, institutional ownership), and firms to be listed on the NYSE, Amex, or Nasdaq, with a reported stock price immediately before the file date of at least \$5. Eliminating low-priced firms reduces the role of bid/ask bounce when we examine the market reaction to the 10-K filing. We require the firm to have sixty days of trading in the year before and after the filing date. We also require the 10-K document to include more than 5,000 words (some filings simply reference other filings). We include only one filing per firm in each calendar year, with at least 180 days between filings. These sample selection criteria produce a sample of 41,842 observations consisting of 7,852 unique firms with an average of 5.3 yearly observations per firm.

Figure 1 depicts the 10-K sample size and median market capitalization by year. As electronic filing was not required of all firms until May 1996, the first few years reveal a smaller

¹ A 10-K405 is a 10-K where a box on the first page is checked indicating that a “disclosure of delinquent filers pursuant to Item 405” was not included in the current filing. Until this distinction was eliminated in 2003, a substantial portion of 10-Ks were categorized as 10-K405. The SEC eliminated the 405 classification due to confusion and inconsistency in its application. The choice does not impact our study, so we include both form types in our sample and will simply refer to them as 10-Ks.

² We use the Wharton Data Services CIK file to link SEC CIK numbers to the CRSP PERMNOs.

number of filings, predominantly by larger firms. The median market value has generally risen each year since 1997, consistent with the overall upward trend in the US equity market.

B. Parsing the 10-Ks

As in most studies using textual analysis in finance, we use a “bag of words” method that requires us to parse the 10-K documents into a vector of words. First we eliminate all non-alphabetic characters. Then we create a vector of tokens by splitting the document into alphabetic collections of characters delineated by spaces. Each token is then compared with a dictionary to determine if it is a word.³

Parsing requires care as there are a number of tripwires in the 10-K data. For example, documents increasingly include embedded graphics, which are jpg or gif formatted pictures encoded as ASCII characters. The ASCII encoding of graphics increases by orders of magnitude the character count and file size of the document. Thus, all encoded graphics must be purged, or the use of document size-related variables can be severely affected.

The same is true for HTML, which increased substantially in the late 1990s; some 10-Ks include more HTML than actual content. Once again, without careful parsing, character counts or certain word counts (i.e., tag words) can be grossly misstated.

It is not clear that tone is dictated primarily in a single section of the document. We do not limit our analysis to the Management Discussion and Analysis (MD&A) section of the 10-Ks for four reasons. First, while in our preliminary tests we capture the majority of MD&A sections correctly, in a significant minority of documents the parsing fails to correctly extract this

³ We use release 4.0 of the 2of12inf dictionary from <http://wordlist.sourceforge.net/12dicts-readme.html>, which includes word inflections but does not include abbreviations, acronyms, or names. Extensive dictionaries (actually word lists) are generally available on the internet due to their usefulness in hacking where they are used for lookups to crack passwords. We have added 11 words not identified in the original list of more than 80,000 words including Scholes (since Black and Scholes is a valuation technique and not simply a person’s name). Our final master dictionary used to determine if a token is a word is available at <http://<author website>>.

section.⁴ Second, because we also develop word lists focusing on litigiousness and uncertainty, we want to include segments discussing risks and litigation. Third, the MD&A section is filtered by lawyers and accountants more so than most of the other textual content. Finally, and most importantly, a substantial portion of the MD&A sections are “included by reference” to the annual report, thus reducing the sample and possibly biasing it toward more recent years or larger firms.⁵ For these reasons, we consider all the textual material in the 10-K documents, not just the MD&A section.

C. 10-K Subsamples

To evaluate the economic relevance of our word lists, we also consider two samples documenting negative financial events in other studies. First, we consider a sample of firms subject to shareholder litigation under Rule 10b-5.⁶ All the firms in this limited sample have been accused of accounting fraud. The sample was created by a keyword search for “GAAP” and “Restatement” in 10b-5 class action suits. The Rule 10b-5 allegations argue that material omissions by managers lead to inflated stock prices.

The dataset identifies the filing date for the case and the time interval over which a firm’s management was accused of accounting improprieties. The available Rule 10b-5 data start in 1990 while our 10-K sample begins in 1994. Thus, the sample overlap is between January 1, 1994 and September 22, 2004.

⁴ Most 10-K forms are not currently filed using XBRL markup language that allows specific identification of document components. Lacking section tags, the parser essentially must rely on an artificial intelligence algorithm to target the desired section of the document. Inconsistencies in the structure of these documents, especially in the earlier years of the sample, make the process very imprecise.

⁵ We examined a random sample of 50 10-Ks for each of the years 1995, 2000, and 2005 taken from our final sample. The number of MD&A sections that were not included in the 10-K and were instead incorporated by reference to the annual report were, respectively, 40%, 30%, and 15%.

⁶ We thank Peter Easton, Greg Sommers, Mark Zmijewski, and Chicago Partners LLC for providing us with the 10b-5 data.

Our logit analysis uses a dummy variable set equal to one if a Rule 10b-5 related lawsuit is filed against the firm any time in the 365 days after its 10-K file date or if the firm's 10-K file date is within the period when alleged accounting improprieties might have taken place. In 507 observations the dummy variable is set equal to one, from a total of 31,570 10-Ks filed over the same time period. Firms in the 10b-5 sample include Enron, Boston Chicken, and Cardinal Health.

Our second sample considers Doyle, Ge, and McVay's (2007) 705 firms that disclosed at least one material weakness in internal control between August 2002 and November 2005.⁷ These disclosures are an artifact of Sections 302 and 404 of the Sarbanes-Oxley Act. A material weakness in internal control is described by the Public Company Accounting Oversight Board (2004) as "more than a remote likelihood that a material misstatement of the annual or interim financial statements will not be prevented or detected."

We use the internal control data set as another litmus test of word categories' ability to identify problematic 10-K disclosures. We create a dummy variable set equal to one if a firm discloses a material weakness in a 10-K, 10-Q, or 8-K any time within approximately one and a half years (540 days) from a given 10-K file date. In 598 cases the dummy variable equals one from a total of 12,878 10-Ks filed during that sample period.

D. Variables

We use various stock return measures relative to the 10-K filing date as both control variables and dependent variables. The file date return is measured as the log of one plus the cumulative excess return from day 0 through day 3. In all cases, excess return refers to the stock return minus the CRSP value-weighted market index return. The event window for the file date

⁷ The data were downloaded from <http://faculty.washington.edu/geweili/ICdata.html> in August 2008.

return is based on Table 2 in Griffin (2003), who documents that 10-Ks' "elevated response extends to day 3" (p. 447).⁸

We also use post-file date returns as a dependent variable, defined as one plus the cumulative excess return for the year after the file date, excluding the five days immediately after the file date. A firm must have a minimum of 60 days with complete returns in the one year interval. We also use the post-file date mean-square-error from a market model regression, with the CRSP value-weighted market index as the market index, to measure return volatility. The return interval and data requirements for the market model estimates are the same as for the post-file date excess return variable.

We include as control variables firm size, book-to-market, institutional ownership, and a dummy variable for Nasdaq listing. Given the prior evidence by Fama and French (1992) that small firms and value firms have higher subsequent returns, it is important to control for these variables. To adjust for microstructure effects and different stock trading behaviors, a Nasdaq dummy and the proportion of the firm owned by institutions are also added.

Firm size is the number of shares outstanding in millions times the price of the stock as reported by CRSP on the day before the file date. Book-to-market data are derived from the Compustat data items as specified in Fama and French (2001), and the market value is taken from CRSP. The variable is based on the most recent Compustat data no more than one year before the file date. After eliminating observations with negative book-to-market, we winsorize the book-to-market variable at 0.005 and 0.995. Institutional ownership is taken from the

⁸ The exact point at which the 10-K becomes public is confounded by changes in the SEC's dissemination process during the sample period and the assumption that the public has completely read and assessed the document. Under the dissemination system established in November 1998, a filing is checked for simple perfunctory errors and then a private contractor makes the document available to the public. Subscribers to the private service can immediately access the documents. Free public access to the filings occurs after a 24-hour delay. For a sample of 10-Ks ending in the mid-1980s, Easton and Zmijewski (1993) find a mean delay of five days in making 10-K documents publicly available.

CDA/Spectrum database for the most recent quarter before the file date. The variable is considered missing for negative values and winsorized to 100% on the positive side.

We include year and industry dummy variables to control for the cross-sectional and time-series effects in the panel of data. We use the 48-industry classification scheme of Fama and French (1997) except for the logit regressions based on the 10b-5 and material weakness samples. In these cases, there are too few observations of a logit variable equal to one, creating situations where a given observation is uniquely identified within one of the 48 industries. Thus for the logit analysis of these two samples we use the Fama and French five industry classifications.

In our initial comparisons of the Harvard negative word lists and the negative word list developed in this study we consider both the simple proportion of words and the normalized differences. Tetlock (2007) uses negative word proportions that are normalized based on the mean and standard deviation for all firms and all news stories from the prior year. We normalize using the mean and standard deviation of the negative word proportions for the Fama-French (1997) 48 industries taken from the prior year. In our subsequent regressions we focus on the simple proportion, which prevents us from losing observations (due to differencing) in the small subsamples focusing on fraud and material weakness.⁹

III. Textual Analysis and Word Lists

There are other labels for textual analysis in different disciplines; terms such as content analysis, natural language processing, information retrieval, or computational linguistics describe

⁹ The natural language processing literature will sometimes consider a weighted measure of word frequency. A common weight is tfidf (term frequency-inverse document frequency), where each word's frequency is adjusted for its frequency of use across all documents (see Manning and Schütze (2003)). Using this adjustment produces slightly higher t-statistics for our Fin-Neg word list. Because term weighting schemes are ad hoc and did not substantively impact our results, we chose not to use them in the current study.

a similar set of text-based methods. Dovring (1954) documents the use of textual analysis as far back as the early 1600s, when hymns were examined for word choices that threatened a particular religious group. Many different disciplines use textual analysis, including psychology, anthropology, linguistics, political science, journalism, and computer science.

As computers became more accessible, textual analysis attracted more attention beginning in the 1950s. The initial General Inquirer group that produced the early versions of the Harvard word lists became most active in the 1960s. The arrival of the internet and the development of search engines caused new interest and prompted more sophisticated techniques.¹⁰

The *Harvard Third Psychosociological Dictionary*, released in March 1963, is the third major revision of a dictionary including numerous categories of words. The current version is available through the General Inquirer website (see <http://www.wjh.harvard.edu/~inquirer/>). We focus on the TAGNeg file used in Tetlock, Saar-Tsechansky, and Macskassy (2008) because it is a non-proprietary list used elsewhere that should be representative of negative word lists developed in disciplines other than finance and accounting.

To use a list to classify words in a document, one must account for inflections, or different forms of the same word. There are two choices. Stemming attempts to programmatically extract a root word from a token in order to account for simple variants. For example, if we consider *accident* a negative word, we would probably also want to include words such as *accidental*, *accidentally*, and *accidents*. A popular stemming algorithm by van Rijsbergen, Robertson, and Porter (1980) is widely available in a variety of programming languages and is used in some prior studies in finance.

¹⁰ For a description of the early research in textual analysis see Stone, Dunphy, Smith, and Ogilvie (1966).

The problem with stemming is that many times a word meaning changes when common prefixes or suffixes are added. From the H4N list, for example, *odd* and *bitter*, take on different meanings when made plural: *odds* and *bitters*.

For this reason, we choose a second approach to account for word inflections: conjugation. We expand the H4N list by inflecting each word to conjugations that retain the original meaning of the root word. Thus, the inflection process targets the same objective as stemming. Although this inflection approach is more labor intensive, it is also more likely to preserve the intention and the tone of the root word. The original H4N list includes 2,004 words; our inflected version increases the number to 4,183.

Tetlock (2007), Engleberg (2008), and Kothari, Li, and Short (2008) find little incremental information in the Harvard positive word list, and we do not include this GI category in the analysis. Engleberg (2008) argues that the Harvard positive word list may fail to correlate with financial disclosures because of erroneous classification (he cites *company*, *shares*, and *outstanding* from the Harvard list). For completeness, however, when we propose additional word lists we include a positive category even though our primary focus is comparison of the Harvard negative word list with our Fin-Neg list.

We propose five other word lists: positive (Fin-Pos); uncertain (Fin-Unc); litigious (Fin-Lit); and strong modal words (MW-Strong) and weak modal words (MW-Weak). All these lists are available at <http://<author website>>.

To create the Fin-Neg, Fin-Pos, Fin-Unc, and Fin-Lit word lists, we first develop a dictionary of words and word counts from all 10-Ks filed during 1994-2007. To be included, a word must appear in at least 5% of the 10-K documents. This filter excludes rarely occurring words that might be in 10-Ks of unique firms or industries (an unusual negative word used only

by a specific firm or industry essentially becomes a dummy variable identifying that case). We are careful in examining each word, to consider their most likely usage in financial documents (including inflections). Our word-based measures will be the proportional occurrence of the words relative to a document's total word count.

We account for simple negation only for Fin-Pos words. Simple negation is observation of one of six words (*no, not, none, neither, never, nobody*) immediately before a positive word. We would not expect to see phrases like “not terrible earnings” in a report, so we do not consider negation for the negative word lists.

Unlike the H4N list, the Fin-Neg list is specific to business terminology. In the language of business, words like *increase* or *decrease* become tonally ambiguous. In this case, what these words imply depends on whether they precede words like revenues or costs. Words from the Harvard lists, such as *liability, tax, or loss* are expected to appear in both positive and negative contexts simply as a structural artifact of accounting language. The critical empirical question is whether such words appear often enough to invalidate the use of word lists derived from other disciplines. Any non-trivial word list applied to as many documents as in our 10-K sample will misclassify—the issue is to what extent. A discipline specific word list can reduce measurement error, thus increasing power and reducing the associated attenuation bias in parameter estimates.

Of the 1,202 words in our Fin-Neg list, 346 overlap with the H4N list, and 772 overlap with the H4N-Inf list. Frequently occurring words in our list that are not on the H4N or H4N-Inf lists include: *litigation, discontinued, unpaid, investigation, misstatement, misconduct, forfeiture, serious, allegedly, deterioration, and felony*.

Our Fin-Pos word list consists of 264 words including inflections, substantially fewer words than in the negative word list. The list also includes fewer words subject to

misclassification as noted by Engleberg (2008). In attempting to select positive words, one quickly realizes that there are few positive words that are not easily compromised. Knowing that readers are using a document to evaluate the value of a firm, writers are likely to be circumspect and avoid negative language, instead qualifying positive words, often in ways that are not easily detected by a parsing program. The tone of negative words has a much more pervasive effect.

For example, *felony*, even if appearing in the phrase “whose *felony* conviction was overturned,” is still negative. Words in our Fin-Pos list such as *achieve*, *attain*, *efficient*, *improve*, *profitable*, or *upturn* are more unilateral in potential tone. Thus we include a positive word list more in the interest of symmetry rather than in an expectation of discerning an impact on tone identification.

The Fin-Unc list includes words denoting uncertainty, with the emphasis on the general notion of imprecision rather than risk. Eight of the 123 words in our original Fin-Unc list overlap with the Fin-Neg list and their inflections, and we exclude them here. Words such as *approximate*, *contingency*, *depend*, *fluctuate*, *indefinite*, *uncertain*, and *variability* are included in the Fin-Unc word list.

The Fin-Lit list categorizes words reflecting a propensity for legal contest or, per our label, litigiousness. Sixteen of the 95 words appearing in our original Fin-Lit list overlap with the Fin-Neg list (again excluded). Words such as *claimant*, *deposition*, *interlocutory*, *testimony*, and *tort* (and their inflections) appear in this list. We also include words like *legislation* and *regulation*, which do not necessarily imply a legal contest but may reflect a more litigious environment.

Words can be classified using many different taxonomies. Pinker (2007) discusses word classifications that go well beyond traditional grammatical categories to identify commonalities

based on the physical world and causality. We look specifically at the modality of words because we believe this categorization is a defining one with respect to the growing regulatory environment for financial disclosures.

We extend Jordan's (1999) categories of strong and weak modal words to include other terms expressing levels of confidence. Examples of strong modal words (MW-Strong) are words such as *always*, *highest*, *must*, and *will*. Examples of weak modal words (MW-Weak) are *could*, *depending*, *might*, and *possibly*.

There are 12 MW-Strong words in our list and 15 MW-Weak words. Four of the MW-Weak words overlap with the Fin-Neg list, but we do not consider this problematic. This dichotomy of strong and weak modal words facilitates our hypothesis that increased regulation and accountability will cause managers to issue vague opinions or pronouncements.

How generalizable are our word lists to other documents like newspaper articles or press releases? Since our list is generated by examination of a large collection of words used in 10-Ks, we believe that the Fin-Neg list could be applied successfully to other financial documents. Although negative words might be used less often in some media releases, there is no reason to believe that misclassification would be more likely.

IV. Results

A. Sample Description

Summary statistics for the sample of 41,842 10-Ks are reported in Table III. In the eight word list variables, the mean and median values are approximately equal. H4N words represent on average 2.365% of a 10-K compared to a median proportion of 2.334%. For the Harvard list with inflections, the mean is 3.841% and the median is 3.818%. As there are fewer words in the Fin-Neg list (1,202 words compared to 4,183 H4N-Inf words), it is not surprising that, on

average, a lower percentage of 10-K words are in the Fin-Neg word list (1.105%). The lowest average percentage of words on the lists (0.210%) is for litigious words.

Given that low-priced stocks are more affected by microstructure biases, our \$5 stock price requirement slightly reduces the filing period returns from those seen in other research. The mean and median 4-day filing period excess returns are slightly negative (-0.07% and -0.14%, respectively). The value-weighted CRSP index is used to convert the raw security returns to excess returns.

The mean market value is \$3.2 billion while the average book-to-market ratio is 0.58. Given that the \$5 price screen and Compustat data requirements remove many small firms from the analysis, the sample has mean and median institutional ownership of about 48%. Thus, our sample covers firms that would be of interest to institutional investors. More than half of our sample is listed on Nasdaq.

To provide a time-series pattern of the negative word lists, Figure 2 plots the median percent of words by calendar year. There are no dramatic or sudden changes during 1994-2007 in the percentage of negative words according to either the Harvard, Harvard with inflections, or Fin-Neg word lists. All three plots reveal a slight upward trend in the percentage of negative words over the sample period. There is nothing in Figure 2 that would suggest the proportion of negative words in a 10-K reflects other market-related time-series events.

Table IV reports correlations between the 4-day filing period returns and each word list. Not surprisingly, the highest correlation (0.868) is between the H4N list and the H4N list with inflections. The correlation between the H4N-Inf and the Fin-Neg list is a positive 0.362. The strongest correlation between filing period abnormal returns and the word lists is for Fin-Neg (-0.031) and weak modal words (-0.058).

B. Examining the Composition of Negative Tone

Which words have a stronger weight in determining the tone of a text? Does the frequency of a limited number of common negative words dominate the likelihood that a text will be classified as pessimistic? Are the Harvard words truly negative in a financial sense? Are some negative words specific to an industry?

For the entire 10-K universe on EDGAR of 113,154 firms, Table V reports the 30 most frequently occurring words from the H4N-Inf (Panel A) and Fin-Neg (Panel B) lists. The check mark in Panel A notes whether the word is also on the Fin-Neg list. The H4N-Inf word that appears most often and also appears on the Fin-Neg list is *against*.

The first column of numbers reports the fractional percentage for each word of the total negative word count. For example, the word *tax* accounts for 3.92% of the total count of all the H4N-Inf negative words. The next column lists the cumulative percentage, and the last column reports the percentage of all 10-Ks in which the listed word appears.

Panel A of Table V demonstrates the considerable misclassification of negative words in 10-K text according to the Harvard word list. The six words, *tax*, *costs*, *capital*, *loss*, *board*, and *expenses* account for more than one-fifth of the total count of “negative” words. Yet in the financial world, *costs* of a firm or the members of a *board* of directors or amount of *tax* paid are neutral in nature. Managers using this language are merely trying to describe their operations.

In some non-business situations, *foreign* or *vice* might be classified as negative words. In 10-K text, however, it does not imply negativity to characterize *foreign* operations or mention the *vice*-presidents of the firm. *Foreign* appears in more than 70% of 10-Ks and accounts for 1.09% of all negative words. *Vice* appears in about 94% of all 10-Ks and accounts for 1.36% of all counted negative words.

In other textual analysis research, the more negative words, the more pessimistic or negative the perceived tone of the text. When we eliminate the four words that also appear on the Fin-Neg list (*against, default, closing, and adverse*), Panel A reveals that more than 50% of the Harvard words classified as negative are not typically negative in the case of financial reporting.

Panel B of Table V reports the 30 most common words according to the Fin-Neg list; 24 of them also appear on the H4N-Inf list. Six words do not (*claims, litigation, discontinued, disclosed, unable, and construed*). Unlike the Harvard list, Fin-Neg includes words that are more likely to be negative in a financial sense.

As an example, in WorldCom's 2002 10-K shortly before filing for bankruptcy protection, the most frequent Fin-Neg words are *impairment, complaint, liquidation, litigation, adverse, and claims*. The six most frequent H4N-Inf words in that report are *costs, expense, tax, service, liabilities, and loss*. These Harvard negative words hardly capture the serious problems that WorldCom was experiencing at the time.

In untabulated results, we examine the difference between the H4N-Inf and Fin-Neg negative proportions for each firm in the sample across industries. Three Fama and French industries (Precious Metals, Insurance, and Coal) account for a much higher fraction of Harvard negative word counts. For Precious Metals and Coal, *mine* is one of the words with the highest count. *Mine* is an H4N negative word; but the word is used in 10-Ks merely to describe operations (e.g., Rochester *Mine*, gold *mine*, or coal *mine*).

In an extreme example, the word *mine* in the 1999 10-K of Coeur d'Alene Mines Corporation accounts for over 24% of all the H4N negative word counts. For the Insurance industry, the most used word on the H4N list is *capital*. To control for industry effects, all of our regressions will include Fama and French industry dummies.

Although using year-to-year differences in the proportional measures will in some cases lower the high levels attributable to the number of words misclassified by the Harvard lists, the magnitude of misclassification shown in Table V indicates that the measure would still remain very noisy.

C. 10-K Filing Period Returns and Negative Word Lists

One way to test word lists is to examine the market's reaction at the time of a 10-K filing. If tone matters, firms filing 10-Ks with a high proportion of negative words should, on average, experience negative excess returns around the filing date. Figure 3 reports the median filing period excess returns by quintiles according to the H4N, H4N-Inf, and Fin-Neg word lists. Median returns for both the H4N and H4N-Inf lists are basically flat across the quintiles. Thus, firms with a high proportion of Harvard negative words have essentially the same filing period announcement returns as firms with relatively few negative words on those lists.

The pattern produced by the Fin-Neg list is what we would expect. Firms including a lower percentage of pessimistic words have positive returns on the 4-days around the 10-K filing date compared to sharply negative median returns for the quintiles including a high percentage of negative words. The return pattern for Fin-Neg across the quintiles is monotonic.

We next examine the relation between Fin-Neg and filing period returns in a multivariate context using various control variables. Table VI reports regression results defining the dependent variable as the log of one plus the 4-day filing period excess returns. The independent variables are the proportions of words classified as negative using the two Harvard word lists and Fin-Neg. The control variables are size, book-to-market, institutional ownership, and a Nasdaq dummy. Included in each regression are Fama and French (1997) 48-industry dummies, calendar year dummies, and a constant.

We first separately consider each of the negative word lists with only industry and year dummies. Consistent with the visual evidence we see in Figure 3, there is no statistical relation between the 4-day filing period returns and either Harvard negative word list. In columns 1 and 2, there are insignificant coefficients on H4N and H4N-Inf. When Fin-Neg is included in a separate regression, its coefficient is negative and significant (t-statistic of -4.06). Thus, the higher the proportion of Fin-Neg words in a 10-K, the lower is the market's reaction on the filing date. It appears that the Fin-Neg word list does a much better job of linking 10-K tone with announcement returns than the two Harvard negative word lists.

To examine the robustness of this finding, we include the additional control variables in the regressions in columns 4 through 6. The H4N and H4N-Inf coefficients remain insignificant. The coefficient on Fin-Neg is smaller, but remains negative and statistically significant. The last column of Table VI shows that size and book-to-market are significant in explaining part of the excess 4-day filing period returns. Larger firms and value firms (i.e., high book-to-market) have higher abnormal filing period returns.

We should not overstate the regression results. Even for Fin-Neg, including size, book-to-market, institutional holdings, industry dummies, and calendar year dummies in the regression, the adjusted R-square is still low (only 1.71%). Only a small amount of the variation in filing period announcement returns is explained by the independent variables. Textual analysis is not the ultimate key to the returns cipher.

Table VII replicates the comparative results for the negative word lists from Table VI using normalized differences as the measure of negative tone. The results are essentially identical to those in Table VI. The t-statistics are closer to zero for the Harvard lists in this case,

even though we might expect the differencing to favor the lists with the highest levels of misclassification. Once again, the Fin-Neg word list is better at gauging financial text.

D. Additional Word Lists, Filing Date Returns, and Post-Filing Date Returns

We have some evidence that Fin-Neg is related to short-term announcement returns when the 10-K is filed. Is there a longer-run relation between the tone of the 10-K (as captured by negative words) and subsequent returns? Or are other word lists linked to short- or long-run stock returns?

Table VIII reports regression results for three different dependent variables: the log of one plus the excess 4-day filing period returns; one-year excess returns; and the mean-square-error of the market model for one year after the filing date. For all three regressions, the six word lists (negative, positive, uncertain, litigious, strong modal words, and weak modal words) are included as independent variables.

As in Table VII, we include a number of control variables, such as size and book-to-market. When all six word lists are included in the first regression, the negative, positive, litigious, and weak modal word coefficients are all significant.¹¹ Firms using fewer negative words, fewer positive words, more litigious words, and fewer weak modal words realize a more positive reaction from the market in the filing date event window. After controlling for other factors, the uncertain and strong modal word lists are not related to short-term filing period returns.

For the post-filing period one-year returns, a number of the word lists are linked to subsequent returns. In column 2, Fin-Neg, Fin-Unc, and MW-Weak all have negative

¹¹If H4N-Inf is added as an independent variable in column 1 of Table VIII, the other variable coefficients and t-statistics remain essentially identical. The coefficient on H4N-Inf is 0.018 with a t-statistic of 0.23.

coefficients while the litigious word list fraction has a positive sign. The negative relation between usage of uncertainty words and subsequent stock returns has been previously documented by Li (2006).¹² Of the four significant coefficients, the negative word list has the highest t-statistic (-9.98). Among the control variables, size, book-to-market, institutional ownership, and Nasdaq listing are all strongly related to subsequent returns.

The last column of Table VIII examines the relation between subsequent filing period return volatility and the six word lists. Five of the six word lists are linked to return volatility. In none of the three regressions in Table VIII is the strong modal word list significant at conventional levels. Negative, positive, uncertain, and weak modal word lists are all positively related to stock return volatility in the subsequent year.

The signs on the control variables seem reasonable. Smaller firms, growth firms, and Nasdaq companies with lower institutional ownership all have higher return volatilities in the year following the filing date. In column 3, the adjusted R-square is over 24%. Hence, short- and long-term returns are more difficult to explain, while word lists do a better job at explaining long-term volatility. The volatility/word tone linkage is consistent with the internet stock message board evidence presented in Antweiler and Frank (2004).

E. 10-b5 Filings and Material Weakness in Internal Controls

We next examine two different firm samples to see whether our six word lists capture language usage differences. Do companies accused of accounting improprieties or firms that self-report material weaknesses in internal controls use different language in their 10-Ks from other firms? Table IX reports the logit regression results. In the first two columns, the binary

¹²To gauge the risk sentiment of a 10-K, Li (2006) counts the frequency of six words (*risk*, *risks*, *risky*, *uncertain*, *uncertainty*, and *uncertainties*).

dependent variable is a dummy variable equal to one if there were a 10b-5 suit filed against the firm alleging accounting improprieties in the year after the 10-K filing date or if the 10-K were filed during the alleged violation period. Between January 1, 1994 and September 23, 2004, there were 507 firms in a potential universe of 31,750 firm-year observations.

The independent variables are the same six word lists and control variables. For all the logit regressions in Table IX, we correct the standard errors used to calculate the z-statistics for rare-event bias following Tomz, King, and Zeng (2003).

The first column in Table IX includes the control variables and the Fin-Neg list; the second column includes all six of the word lists. In the first regression, the coefficient on Fin-Neg is positive and statistically significant (z-statistic of 4.90). This implies that the more negative words used in the 10-K, the higher the chances a firm will be accused of accounting improprieties. Thus, firms undergoing Rule 10b-5 suits are different from other firms in terms of negative word counts, even after controlling for other variables.

The control variable results are as expected, given the evidence in Dechow, Ge, Larson, and Sloan (2008). Firms accused of accounting manipulation tend to be larger and more tilted toward growth (i.e., lower book-to-market ratios). In column 2, Fin-Neg continues to show a positive and significant coefficient when we include the other five word lists in the logit regression. Fin-Pos (negative coefficient) and Fin-Unc (positive coefficient) are also statistically significant.

In the last two columns of Table IX, the binary dependent variable is equal to one if within 18 months of the 10-K file date a firm disclosed a material weakness in a subsequent 10-K, 10-Q, or 8-K (N = 598). The material weakness sample is taken from Doyle, Ge, and McVay (2007). As in the 10b-5 sample, firms with a higher proportion of negative financial words are

more likely to report material weaknesses in their internal accounting controls. Here, however, firms with material weakness are smaller and tend to be tilted more toward value (i.e., higher book-to-market ratios).

For both the 10b-5 and material weakness regressions, it is not clear what we should expect about the word list coefficients. That is, we might expect a higher proportion of negative words for firms with underlying problems or else a lower proportion as managers were trying to minimize underlying problems. Our logit regressions suggest, nevertheless, that word lists can play a role in identifying firms experiencing unusual events.

F. 10-K Tone and Sarbanes-Oxley

The Sarbanes-Oxley Act was an important piece of government regulation signed into law on July 30, 2002. Section 302 of the Act requires managers to certify that their 10-K is free of untrue statements, material omissions, or misleading declarations. Did Sarbanes-Oxley affect the word selection of managers?

Figure 4 presents the pre- and post-SOX proportions for each word list. The legislation clearly had an effect on manager word selection. In untabulated tests, we see in all of the word lists (except litigious) statistically significant differences in word proportions across time periods.

Following the passage of SOX, 10-Ks include more negative, slightly fewer positive, and more uncertain words. Pre-SOX, 1.015% of all words were negative compared to 1.224% in the post-SOX period. The average 10-K in our sample includes 42,987 words. Thus, following passage of SOX, the average 10-K includes about 90 more negative words (526 versus 436).

The most interesting trend in Figure 4 is for modal words. Given that managers are now required to certify their 10-K text, it is enlightening that “weasel words” (or weak modal words)

are seen more while strong modal words are less common. Words like *must*, *shall*, and *always* are used in lower proportions, while *may*, *might*, *perhaps*, and *depends* increase in use following SOX. One clear negative ramification of Sarbanes-Oxley is that 10-Ks may have become less useful, with an increased use of weasel words. Managers appear to be weakening 10-K content to minimize their litigation exposure.

V. Conclusions

We find that many of the words classified as negative in the Harvard Psychological Dictionary (IV-4) are typically not negative in a financial sense. Words like *depreciation*, *liability*, *foreign*, and *mine* have little negative meaning in a 10-K. More important, we document for a large sample of US firms that more than half of the negative word count according to the Harvard list are words that are not negative sentiment words. That is, the majority of words tabulated as negative by the Harvard list are most likely misclassified in the case of financial reports.

We create instead a list of 1,202 words that we believe typically have negative financial meaning in financial reports, by examining all words that occur in at least 5% of the SEC's 10-K universe. In tests on the 10-K filing date, our negative word list is strongly related to announcement returns. Firms with a high proportion of negative words experience significantly lower stock returns than companies with a low proportion of negative words. For the same firms, we see no return pattern in the data using the Harvard list to classify negative words.

Other authors have examined the announcement effect when a high proportion of positive words appear in the financial text. Expanding on that premise, we create five other word classifications (positive, uncertain, litigious, and strong or weak modal words). In tests of market reaction around the 10-K filing date and one year after, we find evidence that some word lists are

related to stock returns and return volatility. Some of our word lists are also linked to company self-reported material weaknesses in firm accounting controls and to firms against which class action lawsuits have been filed.

Lastly, we present evidence that regulation has a significant impact on 10-K word choice. Following passage of the Sarbanes-Oxley Act, we see substantially fewer strong modal words (e.g., *highest*, *definitely*, and *must*) and many more weak modal words (e.g., *may*, *depends*, and *might*). One negative consequence of SOX is that managers appear less forthcoming in their presentation, thereby diluting a text's information content.

Our results do not suggest that textual analysis will resolve, to paraphrase Roll (1988), our profession's modest ability to explain stock returns. Our results and others', however, suggest that textual analysis can contribute to our predictive ability along with more traditional financial variables.

Most important, we show that textual analysis for financial researchers should not rely on word classification schemes derived outside the domain of business usage. Applying non-business word lists to accounting and finance topics can lead to a high misclassification rate. All textual analysis ultimately falls or stands by the categorization procedures.

References

- Antweiler, Werner, and Murray Z. Frank, 2004, Is all that talk just noise? The information content of Internet stock message boards, *Journal of Finance* 59, 1259-1293.
- Berelson, Bernard R., 1952, *Content Analysis in Communication Research*, The Free Press, Glencoe, IL.
- Das, Sanjiv, and Mike Chen, 2001, Yahoo! for Amazon: Opinion extraction from small talk on the web, Working paper, Santa Clara University.
- Davis, Angela, Jeremy Piger, and Lisa Sedor, 2008, Beyond the numbers: Managers' use of optimistic and pessimistic tone in earnings press releases, Working paper, University of Oregon.
- Dechow, Patricia, Weili Ge, Chad Larson, and Richard Sloan, 2008, Predicting material accounting manipulations, Working paper, University of California-Berkeley.
- Demers, Elizabeth, and Clara Vega, 2008, Soft information in earnings announcements: News or noise?, Working paper, INSEAD.
- Dovring, Karin, 1954, Quantitative Semantics in the 19th Century Sweden, *Public Opinion Quarterly* 18, 389-394.
- Doyle, Jeffrey, Weili Ge, and Sarah McVay, 2007, Accruals quality and internal control over financial reporting, *The Accounting Review* 82, 1141-1170.
- Easton, Peter D., and Mark E. Zmijewski, 1993, SEC Form 10K/10Q reports and annual reports to shareholders; Reporting lags and squared market model prediction errors, *Journal of Accounting Research* 31, 113-129.
- Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, Working paper, Northwestern University.
- Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.
- Fama, Eugene, and Kenneth French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153-193.
- Fama, Eugene, and Kenneth French, 2001, Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?, *Journal of Financial Economics* 60, 3-43.

- Feldman, Ronen, Suresh Govindaraj, Joshua Livnat, and Benjamin Segal, 2008, The incremental information content of tone change in management discussion and analysis, Working paper, INSEAD.
- Griffin, Paul, 2003, Got information? Investor response to Form 10-K and Form 10-Q EDGAR filings, *Review of Accounting Studies* 8, 433-460.
- Hanley, Kathleen Weiss, and Gerard Hoberg, 2008, Strategic disclosure and the pricing of initial public offerings, Working paper, University of Maryland.
- Henry, Elaine, 2006, Are investors influenced by how earnings press releases are written?, Working paper, University of Miami.
- Jordan, R.R., 1999, *Academic Writing Course*, Longman, London.
- Kothari, S.P., Xu Li, and James Short, 2008, The effect of disclosures by management, analysts, and financial press on cost of capital, return volatility, and analyst forecasts: A study using content analysis, Working paper, MIT.
- Li, Feng, 2006, Do stock market investors understand the risk sentiment of corporate annual reports?, Working paper, University of Michigan.
- Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221-247.
- Li, Feng, 2009, The determinants and information content of the forward-looking statements in corporate filings - a Naïve Bayesian machine learning approach, Working paper, University of Michigan.
- Manning, Christopher D. and Hinrich Schütze, 2003, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA.
- Pinker, Steven, 2007, *The Stuff of Thought: Language as a Window into Human Nature*, Penguin Group, New York, NY.
- Public Company Accounting Oversight Board (PCAOB), 2004, An audit of internal control over financial reporting performed in conjunction with an audit of financial statements. Auditing Standard No. 2, Washington, D.C..
- Roll, Richard, 1988, R^2 , *Journal of Finance* 43, 541-566.
- Sadique, Shibley, Francis In, and Madhu Veeraraghavan, 2008, The impact of spin and tone on stock returns and volatility: Evidence from firm-issued earnings announcements and the related press coverage, Working paper, Monash University.

- Stone, Philip J., Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie, 1966, *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press, Cambridge, MA.
- Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139-1168.
- Tetlock, Paul C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437-1467.
- Tomz, Michael, Gary King, and Langche Zeng, 2003, ReLogit: Rare events logistic regression, *Journal of Statistical Software* 8, 1-27.
- van Rijsbergen, C.J., S.E. Robertson, and M.F. Porter, 1980, *New Models in Probabilistic Information Retrieval*, University of Cambridge, Cambridge, England.

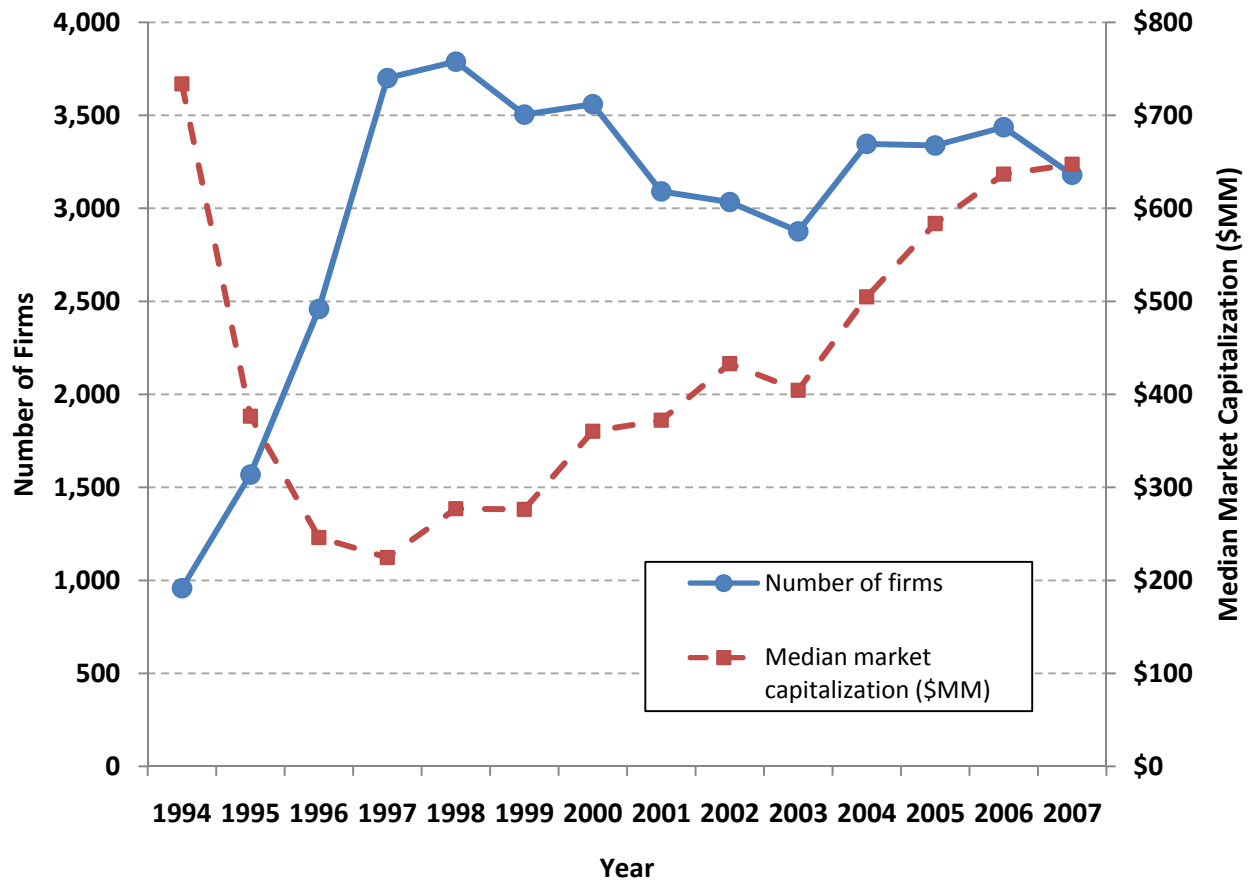


Figure 1. Number of firms and median market capitalization in millions of dollars by year for 10-K sample of 41,842 firm-year observations.

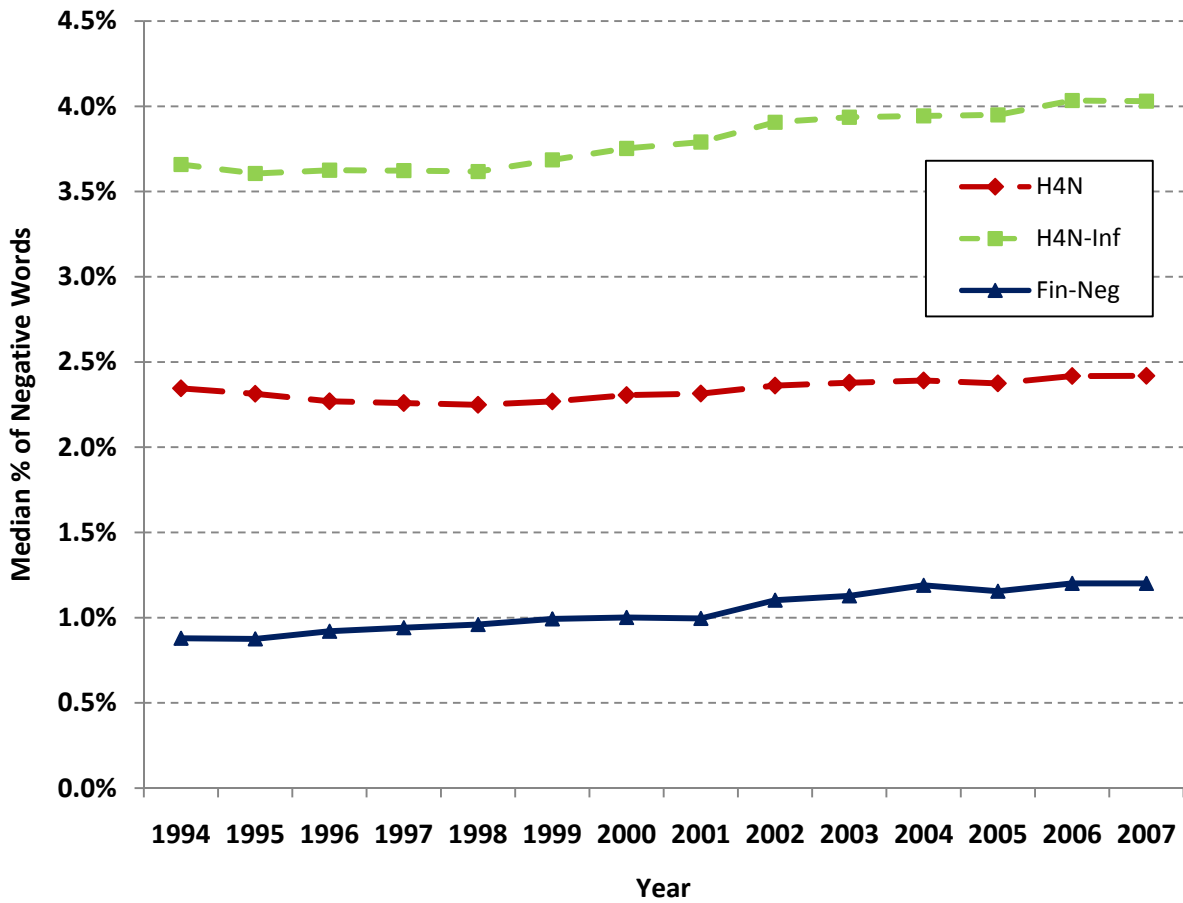


Figure 2. Median percent of negative words by calendar year for the Harvard-IV-4 Psychological Dictionary TagNeg word list (H4N), the H4N list with inflections (H4N-Inf), and the Financial-Negative (Fin-Neg) word list. Each point reflects the median value for the sample in a given year for the percent of negative words relative to total words occurring in a firm’s 10-K based on the specified word list. The sample contains 41,842 firm-year observations.

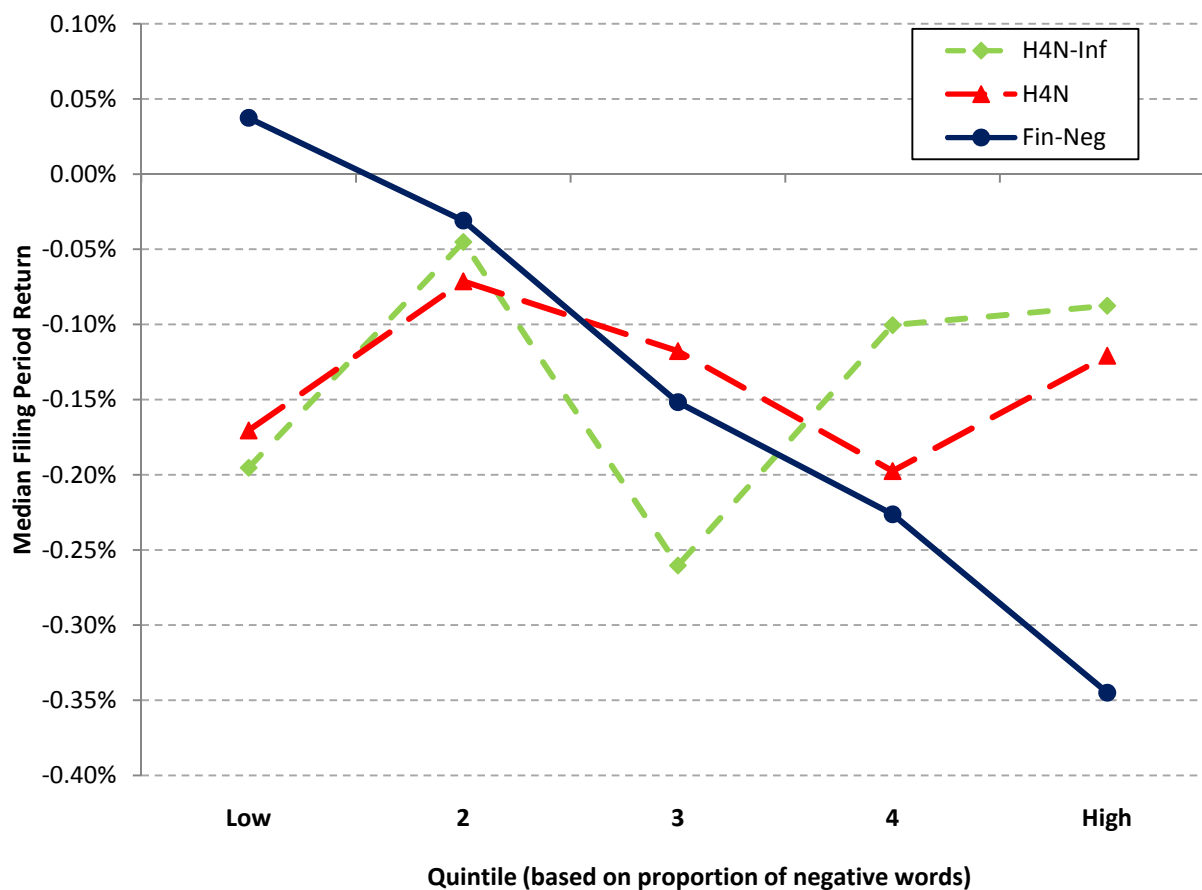


Figure 3. Median filing period excess return by quintile for the Harvard-IV-4 Psychological Dictionary TagNeg word list (H4N), the H4N list with inflections (H4N-Inf), and the Financial-Negative (Fin-Neg) word lists. For each of the three word lists, the sample of 41,842 10-Ks is divided into five portfolios based on the proportion of negative words. The filing period return is the log of one plus the cumulative excess returns for the 10-K file date through the subsequent three days, where excess return is a firm's common stock return minus the CRSP value-weighted market index return.

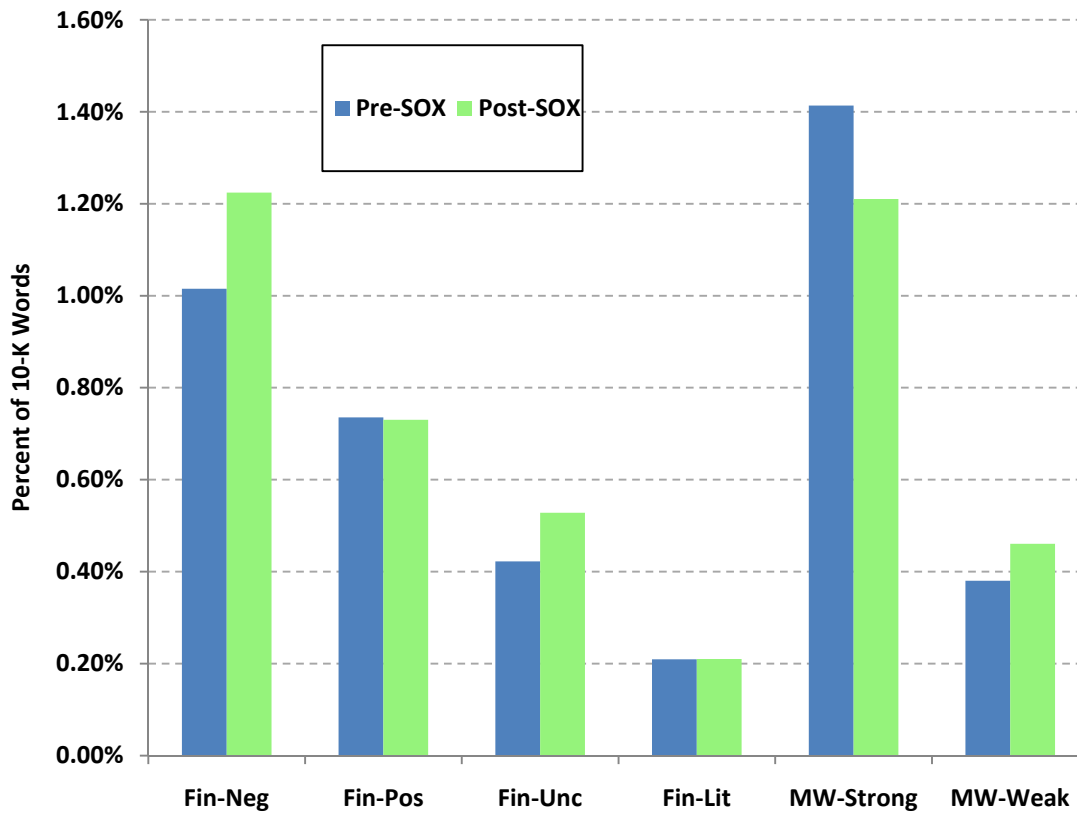


Figure 4. Percent of 10-K words by word list pre- and post-SOX. Reported percentages are for the 1994-2007 sample of 41,842 10-Ks with complete data. The Sarbanes-Oxley Act (SOX) was signed into law on July 30, 2002. Pre- and post-SOX averages for each word category except Fin-Lit are significantly different at the 0.05 level.

Table I
Summary of Finance and Accounting Literature using Word Classifications

For the word classification sources: H4N is the Harvard-IV-4 Psychological Dictionary; GI is Harvard's General Inquirer; LIWC is the Linguistic Inquiry and Word Count dictionary; CUVOALD is the Computer Usable Version of the Oxford Advanced Learner's Dictionary; and DICTION is a dictionary linguistic software. MD&A refers to the Management Discussion and Analysis section of the 10-K.

Authors	Source	Main Focus
Tetlock (2007)	GI / H4N	Links pessimism in the <i>Wall Street Journal</i> column "Abreast of the Market" with lower following day market returns.
Tetlock, Saar-Tsechansky, and Macskassy (2008)	H4N	Find that a high fraction of negative words in new stories forecasts lower firm earnings.
Demers and Vega (2008)	H4N	Find that soft information in quarterly earnings press releases is more informative than the earnings disclosure.
Engelberg (2008)	H4N	Examines investor information processing costs associated with earnings announcements and post-earnings announcement drift.
Sadique, In, and Veeraraghavan (2008)	H4N	Link the fraction of positive and negative words in earnings press releases and media news articles with stock returns.
Li (2008)	LIWC	Links proportion of self-referential and exclusive words in MD&A section of 10-K to earnings persistence.
Li (2009)	LIWC/GI	Uses Naïve Bayesian learning algorithm instead of dictionary to analyze forward looking statements in the MD&A section.
Kothari, Li, and Short (2008)	GI	Links tone of newspaper articles, corporate disclosures (including 10-Ks), and analyst reports to firm cost of capital and stock volatility.
Feldman, Govindaraj, Livnat, and Segal (2008)	GI	Link tone of MD&A section with stock returns around filing date.
Das and Chen (2001)	CUVOALD	Link stock market returns to sentiments on Yahoo's message board.
Henry (2006)	DICTION	Links tone of earnings press release with stock market reaction.
Davis, Piger, and Sedor (2008)	DICTION	Link language in earnings press releases with subsequent firm performance.

Table II
10-K Sample Creation

This table reports the impact of various data filters on initial 10-K sample size.

Source/Filter	Sample Size	Observations Removed
SEC EDGAR 10-K / 10-K405 1994-2007 complete sample	113,154	
CRSP PERMNO match	70,925	42,229
Reported on CRSP as an ordinary common equity firm	66,218	4,707
CRSP market capitalization data available	60,847	5,371
NYSE, AMEX or Nasdaq listing	60,781	66
Price on filing date day minus one \geq \$5	48,063	12,718
Complete returns for day 0-3 filing period	47,994	69
At least 60 days of returns in year before and after file date	47,710	284
Book-to-market Compustat data available and book value > 0	46,258	1,452
Institutional ownership data available on CDA/Spectrum	42,019	4,239
Number of words in 10-K $\geq 5,000$	41,967	52
Include only first filing in a given year	41,865	102
At least 180 days between a given firm's 10-K filings	41,842	23
Firm-Year Sample	41,842	
Number of unique firms	7,852	
Average number of years per firm	5.33	

Table III
Summary Statistics for the 1994-2007 10-K Sample

The first eight variables represent the proportion of occurrences for a given word list relative to the total number of words in a firm's 10-K. The word lists are available at <http://<author website>>. The filing period excess return is the cumulative excess returns from the 10-K file date through the subsequent three days, where excess return is the firms' common stock return minus the CRSP value-weighted market index return. Size is number of shares outstanding (in millions) times price of the stock as reported by CRSP on the day before the file date (or the first day within 22 days before the file date with complete data). Book-to-market, defined as in Fama and French (2001), is considered missing for negative values and then winsorized at 0.005 and 0.995. Institutional ownership is taken from CDA/Spectrum and is considered missing for negative values and winsorized to 100% for cases greater than 100%. One-year post-filing excess returns are cumulative excess returns excluding the five days surrounding the file date; there must be at least 60 days with complete returns data. The post mean-square-error is based on a market model with the CRSP value-weighted market return as the market index and using daily returns for one year, excluding the five days surrounding the file date. Market model estimates must also have at least 60 days with complete returns data. The Nasdaq dummy is set equal to one if the firm's stock is listed on Nasdaq on the day prior to the file date and zero otherwise. The statistics are based on the 41,842 10-Ks on EDGAR for the period 1994-2007 where complete data are available for each variable.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
<u>Word Lists</u>					
H4N (Harvard-IV-4)	2.365%	2.334%	0.354%	1.111%	6.461%
H4N-Inf (H4N w/ inflections)	3.841%	3.818%	0.553%	1.820%	9.154%
Fin-Neg (negative)	1.105%	1.066%	0.395%	0.160%	4.314%
Fin-Pos (positive)	0.733%	0.707%	0.180%	0.196%	2.947%
Fin-Unc (uncertain)	0.468%	0.457%	0.176%	0.059%	1.812%
Fin-Lit (litigious)	0.210%	0.196%	0.110%	0.016%	2.890%
MV-Strong (strong modal verbs)	1.330%	1.231%	0.426%	0.447%	3.441%
MV-Weak (weak modal verbs)	0.414%	0.392%	0.169%	0.000%	1.827%
<u>Other Variables</u>					
Event-period(0,3) excess return	-0.07%	-0.14%	6.42%	-74.15%	158.99%
Size	\$3,162	\$394	\$15,066	\$0.7	\$464,608
Book-to-market	0.5794	0.4970	0.4141	0.0180	5.7450
Institutional ownership	47.99%	47.98%	28.15%	0.00%	100.00%
One-year post-event abnormal return	2.87%	-4.57%	57.57%	-99.74%	1398.62%
Mean-square-error post-event	0.0011	0.0006	0.0017	0.0000	0.0689
Nasdaq dummy	0.550	1.0000	0.4975	0.0000	1.0000

Table IV
Correlations for Filing Period Excess Return and Word Lists

Correlations for the sample of 41,842 10-Ks over 1994-2007. The filing period excess return is the cumulative excess returns from the 10-K file date through the subsequent three days, where excess return is the firm's common stock return minus the CRSP value-weighted market index return. The other variables are the proportion of occurrences for a given word list relative to the total number of words in a firm's 10-K.

	Filing- period excess return	H4N	H4N-Inf	Fin-Neg	Fin-Pos	Fin-Unc	Fin-Lit	MW- Strong
H4N (Harvard-IV-4)	-0.004							
H4N-Inf (H4N w/ inflections)	0.003	0.868						
Fin-Neg (negative)	-0.031	0.354	0.362					
Fin-Pos (positive)	-0.029	0.044	0.012	-0.164				
Fin-Unc (uncertain)	-0.017	0.184	0.363	-0.073	0.177			
Fin-Lit (litigious)	0.023	0.080	0.085	0.281	-0.186	-0.224		
MW-Strong (strong modal words)	-0.005	-0.128	-0.364	0.303	-0.166	-0.678	0.218	
MW-Weak (weak modal words)	-0.058	0.079	0.012	0.365	0.055	0.061	0.071	0.139

Table V
Thirty Most Frequent Words Occurring in 10-Ks
from the H4N-Inf and Fin-Neg Word Lists

The H4N-Inf word list is based on the Harvard-IV-4 Psychological Dictionary TagNeg file. We extend the original word list to include appropriate inflections. The Fin-Neg word list includes negative words from a list of all words occurring in more than 5% of the full sample of 10-Ks filed over 1994-2007. There are 4,183 H4N-Inf words, based on 2,004 root words in the original H4N list. Fin-Neg consists of 1,202 words, including inflections. The word lists are available at: <http://<author website>>. The results in this table are based on the original sample of 113,154 10-Ks available on EDGAR over 1994-2007.

Panel A: H4N-Inf				
Word is also in Fin-Neg	Word	% of Total H4N-Inf Word Count	Cumulative % of Total H4N-Inf Word Count	% of All 10-Ks Including Word
	TAX	3.92	3.92	91.71
	COSTS	3.70	7.62	89.24
	CAPITAL	3.49	11.11	91.65
	LOSS	3.33	14.44	90.77
	BOARD	3.07	17.50	95.13
	EXPENSES	2.99	20.50	90.23
	COST	2.72	23.21	88.46
	EXPENSE	2.49	25.70	89.06
	SERVICE	2.41	28.12	88.78
	LIABILITIES	2.36	30.47	89.17
	TAXES	2.18	32.65	90.99
	LIMITED	2.13	34.78	92.12
	LOSSES	1.64	36.42	89.12
	LIABILITY	1.53	37.95	88.56
	RISK	1.45	39.40	92.41
	MAKE	1.39	40.79	96.07
	VICE	1.36	42.15	93.69
✓	AGAINST	1.31	43.46	85.18
	ORDER	1.14	44.60	86.24
✓	DEFAULT	1.13	45.73	60.49
	FOREIGN	1.09	46.82	72.11
	DECREASE	1.01	47.83	88.05
	CHARGES	0.93	48.77	78.32
	RESTRICTED	0.92	49.69	67.85
✓	CLOSING	0.90	50.59	75.61
	EXCESS	0.89	51.47	86.84
	MATTERS	0.86	52.33	99.37
	DEPRECIATION	0.84	53.17	85.54
	CHARGE	0.78	53.95	79.66
✓	ADVERSE	0.78	54.73	82.72

Table V – continued

Panel B: Fin-Neg				
Word is Also in H4N-Inf	Word	% of Total Fin-Neg Word Count	Cumulative % of Total Fin-Neg Word Count	% of All 10- Ks Including Word
✓	AGAINST	4.08	4.08	85.18
✓	DEFAULT	3.51	7.58	60.49
	CLAIMS	3.42	11.00	76.14
✓	CLOSING	2.79	13.79	75.61
✓	ADVERSE	2.41	16.20	82.72
✓	IMPAIRMENT	2.25	18.45	59.89
✓	FAILURE	2.02	20.47	72.55
✓	LIMITATION	1.97	22.45	72.75
	LITIGATION	1.53	23.98	62.03
✓	ADVERSELY	1.50	25.47	80.11
✓	FORCE	1.32	26.79	75.11
	DISCONTINUED	1.29	28.09	67.05
✓	DAMAGES	1.28	29.37	45.20
✓	BREACH	1.27	30.64	59.40
✓	BANKRUPTCY	1.24	31.87	50.55
✓	LIQUIDATION	1.03	32.90	47.92
✓	UNPAID	1.02	33.92	57.91
✓	LIMITATIONS	0.94	34.86	57.11
✓	DECLINE	0.86	35.72	70.54
	DISCLOSED	0.83	36.55	67.16
✓	DAMAGE	0.82	37.37	68.17
	UNABLE	0.80	38.17	46.88
	CONSTRUED	0.70	38.87	76.12
✓	HAZARDOUS	0.70	39.57	66.63
✓	DEFICIT	0.68	40.24	45.30
✓	OMITTED	0.67	40.91	35.47
✓	VIOLATION	0.67	41.58	47.55
✓	CONTRARY	0.61	42.19	76.59
✓	DOUBTFUL	0.60	42.79	47.89
✓	IMPAIRED	0.57	43.36	38.80

Table VI
Comparison of Negative Word Lists using Filing Period Excess Return Regressions

The dependent variable in each regression is the log excess returns cumulated from the 10-K file date through the subsequent three days, where excess return is the firm's common stock return minus the CRSP value-weighted market index return. The first three variables represent the proportion of the given word list occurrences relative to the total number of words appearing in a firm's 10-K. The Fin-Neg word list is available at <http://<author website>>. The control variables are as previously defined. Fama-French (1997) industry dummies (based on 48 industries), year dummies, and a constant are also included in each regression. Robust t-statistics are reported in parentheses. The estimates are based on the sample of 41,842 10-Ks over 1994-2007.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<u>Word Lists</u>						
H4N (Harvard-IV-4)	-0.116 (-1.23)			-0.154 (-1.61)		
H4N-Inf (H4N with inflections)		-0.053 (-0.87)			-0.094 (-1.53)	
Fin-Neg (negative)			-0.405 (-4.06)			-0.379 (-3.81)
<u>Control Variables</u>						
Log(size)				0.001 (3.90)	0.001 (3.95)	0.001 (3.90)
Log(book-to-market)				0.006 (8.98)	0.006 (8.98)	0.006 (8.96)
Institutional ownership				0.000 (0.02)	-0.000 (-0.00)	0.000 (0.17)
Nasdaq dummy				-0.003 (-4.32)	-0.003 (-4.33)	-0.003 (-4.09)
Adjusted R-square	1.28%	1.28%	1.32%	1.67%	1.67%	1.71%

Table VII
Comparison of Negative Word Lists using Filing Period Excess Return Regressions
Normalized Differences

All variables are defined as in Table VI, except the negative word list variables are now normalized differences. The normalized difference is defined as: (current period negative word proportion - average negative word proportion for the same Fama-French 48 industry) / standard deviation of the prior year Fama-French 48 industry proportion. Industry dummies, year dummies, and a constant are also included in each regression. Robust t-statistics are reported in parentheses. The estimates are based on 32,478 10-Ks over 1994-2007. All coefficients are multiplied by one hundred.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<u>Word Lists</u>						
H4N (Harvard-IV-4)	-0.025 (-0.75)			-0.031 (-0.93)		
H4N-Inf (H4N with inflections)		-0.025 (-0.77)			-0.039 (-1.20)	
Fin-Neg (negative)			-0.116 (-3.18)			-0.118 (-3.24)
<u>Control Variables</u>						
Log(size)				0.125 (4.81)	0.126 (4.84)	0.124 (4.77)
Log(book-to-market)				0.493 (6.84)	0.494 (6.86)	0.494 (6.88)
Institutional ownership				0.094 (0.54)	0.089 (0.51)	0.130 (0.74)
Nasdaq dummy				-0.181 (-2.29)	-0.182 (-2.29)	-0.169 (-2.14)
Adjusted R-square	1.15%	1.15%	1.19%	1.46%	1.46%	1.49%

Table VIII
Additional Word Lists, Filing Period Returns, and Post-Filing Period Returns

The dependent variables in the three columns below are: (1) Filing Period Excess Return – the log of one plus the cumulative excess returns from the 10-K file date through the subsequent three days, (2) Post-Filing Excess Return – the log of one plus the cumulative excess return for the calendar year following the file date, and (3) MSE Post-Filing – the mean-square-error of a market model regression with the CRSP value-weighted market return as the market index and using daily returns for one calendar year. Excess returns in (1) and (2) are the firm’s common stock returns minus the CRSP value-weighted market index returns. For both post-filing measures, the first five days following the file date are excluded, and the period must have a minimum of 60 return observations. Fama-French (1997) industry dummies (based on 48 industries), year dummies, and a constant are also included in each regression.

Variable	Dependent Variable		
	Filing Period Excess Return (1)	Post-Filing Excess Return (2)	MSE Post-Filing (3)
<u>Word Lists</u>			
Fin-Neg (negative)	-0.464 (-4.03)	-9.149 (-9.98)	0.048 (18.81)
Fin-Pos (positive)	-0.406 (-2.08)	-0.698 (-0.45)	0.027 (6.13)
Fin-Unc (uncertain)	-0.398 (-1.48)	-6.490 (-3.13)	0.022 (3.43)
Fin-Lit (litigious)	1.128 (3.10)	18.501 (6.11)	-0.083 (-8.14)
MV-Strong (strong modal verbs)	-0.008 (-0.07)	-0.400 (-0.45)	-0.003 (-1.12)
MV-Weak (weak modal verbs)	-0.971 (-3.73)	-13.672 (-7.00)	0.059 (10.58)
<u>Control Variables</u>			
Log(size)	0.001 (3.95)	0.011 (5.23)	-0.000 (-23.92)
Log(book-to-market)	0.005 (8.62)	0.094 (17.94)	-0.000 (-17.26)
Institutional ownership	0.001 (0.39)	0.101 (7.82)	-0.001 (-14.99)
Nasdaq dummy	-0.002 (-3.26)	-0.022 (-3.67)	0.000 (21.03)
Adjusted R-square	1.79%	7.81%	24.76%

Table IX
Logit Regressions for Shareholder Class Action Suits for Accounting Issues
and Self-Reported Material Weakness in Internal Controls

The binary dependent variable for the logit regressions in columns (1) and (2) is equal to one if a firm had a 10b-5 class action lawsuit filed in the year after the 10-K filing date or if the 10-K file date falls within the purported violation period reported in the 10b-5 (N = 507). The 10b-5 sample includes only firms accused of accounting fraud. To parallel the dates available in the 10b-5 sample, all 10-Ks from 1/1/1994 through 9/23/2004 are included in the sample. In columns (3) and (4) the binary dependent variable is set equal to one if within 18 months of the 10-K file date a disclosure of material weakness is reported in a subsequent 10-K, 10-Q, or 8-K (N = 598). The material weakness sample is taken from Doyle, Ge, and McVay (2007). To parallel the dates of their dataset, the observation period for this second sample is all 10-Ks from 2/1/2001 through 10/30/2005. For all regressions, the standard errors used to calculate the z-statistics, in parentheses, are corrected for rare-event bias using the method of Tomz, King, and Zeng (2003). Dummy variables for the Fama-French five industry classifications, year dummies, and a constant are also included in the logit regressions.

Variable	Logit Dependent Variable			
	Fraud		Material Weakness	
	(1)	(2)	(3)	(4)
<u>Word Lists</u>				
Fin-Neg (negative)	52.142 (4.90)	57.166 (4.82)	54.072 (4.72)	63.520 (4.84)
Fin-Pos (positive)		-114.813 (-4.07)		-50.787 (-1.79)
Fin-Unc (uncertain)		94.291 (2.79)		42.042 (1.18)
Fin-Lit (litigious)		-90.298 (-1.63)		-171.136 (-2.89)
MV-Strong (strong modal verbs)		0.741 (0.05)		16.213 (1.03)
MV-Weak (weak modal verbs)		-25.253 (-0.91)		-6.511 (-0.22)
<u>Control Variables</u>				
Log(size)	0.167 (5.14)	0.173 (5.41)	-0.147 (-4.19)	-0.137 (-3.89)
Log(book-to-market)	-0.175 (-2.87)	-0.199 (-3.26)	0.104 (1.57)	0.097 (1.43)
Institutional ownership	0.917 (4.30)	0.862 (4.08)	0.872 (4.58)	0.772 (4.02)
Nasdaq dummy	-0.007 (-0.06)	0.000 (0.00)	0.190 (1.80)	0.184 (1.74)
Sample size	31,570	31,570	12,878	12,878

Reviewer's Appendix
Word Lists – these will be included by reference to the authors' website

Fin-Neg (negative)	BACKWARD	COLLISION	COUNTERFEIT	DELAYING
ABANDON	BAD	COLLISIONS	COUNTERMEASURE	DELAYS
ABANDONED	BALK	COLLUSION	CRIME	DELETE
ABANDONING	BAN	COMBATING	CRIMES	DELETED
ABANDONMENT	BANKRUPT	COMPLAIN	CRIMINAL	DELETERIOUS
ABANDONMENTS	BANKRUPTCIES	COMPLAINED	CRISES	DELETING
ABANDONS	BANKRUPTCY	COMPLAINING	CRISIS	DELETION
ABNORMAL	BANNED	COMPLAINT	CRITICIZE	DELETIONS
ABNORMALITIES	BANNING	COMPLAINTS	CRITICIZED	DELIBERATE
ABNORMALLY	BANS	COMPLICATE	CRITICIZES	DELIBERATELY
ABOLISH	BARRED	COMPLICATED	CRITICS	DELINQUENCIES
ABOLISHED	BARRIER	COMPLICATES	CULPABILITY	DELINQUENCY
ABROGATE	BARRIERS	COMPLICATING	CULPABLE	DELINQUENT
ABRUPT	BOYCOTT	COMPLICATIONS	CURTAIL	DEMISE
ABSENCE	BOYCOTTS	COMPROMISE	CURTAILED	DEMISED
ABSENCES	BREACH	COMPROMISED	CURTAILING	DEMISES
ABSENT	BREACHED	COMPROMISES	CURTAILMENT	DEMOLISH
ABSENTEEISM	BREACHES	COMPROMISING	CURTAILMENTS	DEMOLISHED
ABUSE	BREACHING	CONCEDING	CUT	DEMOLITION
ABUSES	BREAK	CONCERN	CUTBACK	DEMOTION
ABUSIVE	BREAKDOWN	CONCERNED	CUTBACKS	DENIAL
ACCIDENT	BREAKDOWNS	CONCERNS	CUTS	DENIALS
ACCIDENTAL	BREAKING	CONCILIATION	CUTTING	DENIED
ACCIDENTALLY	BREAKS	CONDEMN	DAMAGE	DENIES
ACCIDENTS	BRIBE	CONDEMNATION	DAMAGED	DENY
ACCUSATION	BRIBERY	CONDEMNED	DAMAGES	DENYING
ACCUSED	BRIBES	CONDEMNING	DAMAGING	DEplete
ADULTERATED	BROKEN	CONFESS	DAMPEN	DEPLETED
ADULTERATION	BURDEN	CONFESSED	DAMPENED	DEPLETING
ADVERSARIAL	BURDENED	CONFESSES	DANGER	DEPLETION
ADVERSARY	BURDENING	CONFESSING	DANGEROUS	DEPRECIATION
ADVERSE	BURDENS	CONFESSION	DANGERS	DEPRESS
ADVERSELY	BURDENSOME	CONFISCATED	DEBARMENT	DEPRESSED
ADVERSITY	BURNED	CONFISCATION	DEBARRED	DEPRESSING
AFTERMATH	CALAMITY	CONFLICT	DECEASED	DEPRIVATION
AGAINST	CANCEL	CONFLICTING	DECEIT	DEPRIVE
AGGRAVATED	CANCELED	CONFLICTS	DECEIVE	DEPRIVED
ALIENATE	CANCELING	CONFRONT	DECEPTIVE	DEPRIVES
ALIENATION	CANCELLATION	CONFRONTED	DECLINE	DEPRIVING
ALLEGATION	CANCELLATIONS	CONFRONTING	DECLINED	DEROGATE
ALLEGATIONS	CANCELLED	CONFUSED	DECLINES	DEROGATION
ALLEGE	CANCELLING	CONFUSING	DECLINING	DEROGATORY
ALLEGED	CANCELS	CONFUSION	DEFACE	DESTROY
ALLEGEDLY	CARELESSNESS	CONSPIRED	DEFAMATION	DESTROYED
ALLEGES	CASUALTIES	CONSPIRING	DEFAMATORY	DESTROYS
ALLEGING	CASUALTY	CONSTRUE	DEFAULT	DESTRUCTION
AMBIGUITY	CATASTROPHE	CONSTRUED	DEFAULTED	DESTRUCTIVE
AMBIGUOUS	CATASTROPHES	CONSTRUING	DEFAULTING	DETENTION
ANNOY	CATASTROPHIC	CONTEMPT	DEFAULTS	DETER
ANNOYANCE	CAUTION	CONTEND	DEFEAT	DETERIORATE
ANNUL	CAUTIONARY	CONTENDED	DEFEATED	DETERIORATED
ANNULLED	CAUTIONED	CONTENDING	DEFECT	DETERIORATING
ANNULMENT	CAUTIONS	CONTENDS	DEFECTIVE	DETERIORATION
ANOMALOUS	CAUTIOUS	CONTENTION	DEFECTS	DETERRENT
ANTITRUST	CEASE	CONTRADICT	DEFEND	DETERRING
APPROPRIATED	CEASED	CONTRADICTED	DEFENDANT	DETRACT
ARBITRARY	CEASES	CONTRADICTION	DEFENDANTS	DETRACTING
ARGUE	CEASING	CONTRARY	DEFENDED	DETRIMENT
ARGUED	CENSURE	CONTROVERSIAL	DEFENDING	DETRIMENTAL
ARGUING	CENSURES	CONTROVERSIES	DEFENDS	DEVIATE
ARGUMENT	CIRCUMVENT	CONTROVERSY	DEFENSIVE	DEVIATED
ARGUMENTS	CLAIMING	CONVICED	DEFER	DEVIATES
ARREST	CLAIMS	CONVICTION	DEFICIENCIES	DEVIATION
ARRESTED	CLOSED	CORRECTED	DEFICIENCY	DEVIATIONS
ARRESTS	CLOSEOUT	CORRECTING	DEFICIENT	DEVOLVE
ARTIFICIALLY	CLOSING	CORRECTION	DEFICIT	DIFFICULT
ASSAULT	CLOSINGS	CORRECTIONS	DEFICITS	DIFFICULTIES
ATTENUATION	CLOSURE	CORRECTS	DEFRAUD	DIFFICULTY
ATTRITION	CLOSURES	CORRUPT	DEGRADATION	DIMINISH
AVOID	COERCE	CORRUPTION	DEGRADE	DIMINISHED
AVOIDANCE	COERCION	COSTLY	DEGRADED	DIMINISHES
AVOIDED	COERCIVE	COUNTERCLAIM	DEGRADING	DIMINISHING
AVOIDING	COLLAPSE	COUNTERCLAIMED	DELAY	DIMINUTION
AVOIDS	COLLAPSED	COUNTERCLAIMS	DELAYED	DISADVANTAGE

DISADVANTAGED	DISRUPTIONS	ESCALATIONS	HAZARD	INCONSISTENT
DISADVANTAGEOUS	DISRUPTIVE	EVICT	HAZARDOUS	INCONVENIENCE
DISADVANTAGES	DISSATISFACTION	EVICTED	HAZARDS	INCONVENIENT
DISAGREE	DISSATISFIED	EVICTION	HINDER	INCORRECT
DISAGREED	DISSENT	EXACERBATE	HINDERED	INCORRECTLY
DISAGREEMENT	DISSENTERS	EXACERBATED	HINDERING	INDECENT
DISAGREEMENTS	DISSENTING	EXCEPTIONS	HINDRANCE	INDEFEASIBLE
DISAGREES	DISSOLUTION	EXCESSIVE	HOSTILE	INDEFEASIBLY
DISALLOW	DISTORT	EXCESSIVELY	HOSTILITIES	INDICTED
DISALLOWED	DISTORTED	EXCULPATION	HOSTILITY	INDICTMENT
DISALLOWING	DISTORTION	EXCULPATORY	HURT	INEFFECTIVE
DISALLOWS	DISTRACT	EXONERATE	IDLE	INEFFECTIVENESS
DISAPPOINTING	DISTRACTED	EXONERATION	IDLED	INEFFICIENCIES
DISAPPROVAL	DISTRACTION	EXPOSE	IDLING	INEFFICIENT
DISAPPROVE	DISTRACTIONS	EXPOSED	ILL	INELIGIBILITY
DISAPPROVED	DISTRESS	EXPOSES	ILLEGAL	INELIGIBLE
DISAPPROVES	DISTRESSED	EXPOSING	ILLEGALITY	INEXPERIENCED
DISAPPROVING	DISTURB	EXPROPRIATION	ILLEGALLY	INFRINGE
DISASTER	DISTURBANCE	EXPULSION	IMBALANCE	INFRINGED
DISASTERS	DISTURBANCES	FAIL	IMBALANCES	INFRINGEMENT
DISCIPLINARY	DISTURBED	FAILED	IMMATURE	INFRINGEMENTS
DISCLAIM	DISTURBING	FALLING	IMMORAL	INFRINGES
DISCLAIMED	DISTURBS	FAILS	IMPAIR	INFRINGING
DISCLAIMER	DIVERSION	FAILURE	IMPAIRED	INIMICAL
DISCLAIMERS	DIVORCE	FAILURES	IMPAIRING	INJUNCTION
DISCLAIMS	DIVULGE	FALSE	IMPAIRMENT	INJUNCTIONS
DISCLOSE	DIVULGED	FALSELY	IMPAIRMENTS	INJURE
DISCLOSED	DOUBTFUL	FALSIFICATION	IMPAIRS	INJURED
DISCLOSES	DOWNGRADE	FAULT	IMPEDE	INJURIES
DISCLOSING	DOWNGRADED	FAULTS	IMPEDED	INJURIOUS
DISCONTINUANCE	DOWNGRADES	FAULTY	IMPEDIMENT	INJURY
DISCONTINUATION	DOWNGRADING	FEAR	IMPEDIMENTS	INQUIRY
DISCONTINUE	DOWNSIZE	FEARS	IMPEDING	INSENSITIVE
DISCONTINUED	DOWNSIZED	FELONY	IMPENDING	INSTABILITY
DISCONTINUES	DOWNSIZING	FICTITIOUS	IMPERATIVE	INSUBORDINATION
DISCONTINUING	DOWNTIME	FINE	IMPERFECTIONS	INSUFFICIENCY
DISCOURAGE	DOWNTURN	FINED	IMPERMISSIBLE	INSUFFICIENT
DISCOURAGED	DOWNTURNS	FINES	IMPLICATED	INSURRECTION
DISCOURAGING	DOWNWARD	FIRE	IMPOSE	INSURRECTIONS
DISCREPANCIES	DOWNWARDS	FIRING	IMPOSES	INTERFERE
DISCREPANCY	DRAG	FLAWS	IMPOSING	INTERFERED
DISGORGEMENT	DRAIN	FORBIDDEN	IMPOSSIBILITY	INTERFERENCE
DISHONEST	DRASTIC	FORCE	IMPOSSIBLE	INTERFERES
DISHONESTY	DRASTICALLY	FORCED	IMPRACICAL	INTERFERING
DISHONOR	DRAWBACKS	FORCES	IMPRISONMENT	INURE
DISHONORED	DROUGHT	FORCIBLE	IMPROPER	INURES
DISINTERESTED	DURESS	FORCING	IMPROPERLY	INURING
DISLOYALTY	DYSFUNCTION	FORECLOSE	IMPROPRIETIES	INVALID
DISMISS	EASING	FORECLOSED	IMPROPRIETY	INVALIDATE
DISMISSAL	EGREGIOUS	FORECLOSES	INABILITY	INVALIDATED
DISMISSALS	EMBEZZLEMENT	FORECLOSING	INACCESSIBLE	INVALIDATING
DISMISSED	ENCROACH	FORECLOSURE	INACCURACIES	INVALIDATION
DISMISSING	ENCROACHMENT	FORECLOSURES	INACCURACY	INVALIDITY
DISPARAGE	ENCROACHMENTS	FOREGO	INACCURATE	INVESTIGATION
DISPARAGEMENT	ENCUMBER	FOREGONE	INACTION	INVESTIGATIONS
DISPARAGING	ENCUMBERED	FORFEIT	INADEQUACIES	INVOLUNTARILY
DISPARATE	ENCUMBERING	FORFEITED	INADEQUACY	INVOLUNTARY
DISPARITY	ENCUMBERS	FORFEITING	INADEQUATE	IRREGULAR
DISPLACE	ENCUMBRANCE	FORFEITS	INADEQUATELY	IRREGULARITIES
DISPLACED	ENCUMBRANCES	FORFEITURE	INADVERTENT	IRREGULARITY
DISPLACEMENT	ENDANGER	FRAUD	INADVERTENTLY	IRREPARABLE
DISPOSSESS	ENDANGERED	FRAUDULENT	INADVISABLE	IRREPARABLY
DISPOSSESSED	ENJOIN	FRAUDULENTLY	INAPPLICABLE	JUDGMENTAL
DISPROPORTIONATE	ENJOINED	FRIVOLOUS	INAPPROPRIATE	JUSTIFIABLE
DISPROPORTIONATELY	ENJOINING	FRUSTRATE	INAPPROPRIATELY	JUSTIFICATION
DISPUTE	ENJOINS	FRUSTRATED	INCAPABLE	JUSTIFIED
DISPUTED	ENURE	FRUSTRATION	INCAPACITATED	JUSTIFIES
DISPUTES	ERODE	GRIEVANCE	INCAPACITY	JUSTIFY
DISPUTING	ERODED	GRIEVANCES	INCIDENCE	JUSTIFYING
DISQUALIFICATION	EROSION	GROSSLY	INCIDENT	KICKBACK
DISQUALIFIED	ERRATIC	GUILTY	INCIDENTS	KICKBACKS
DISQUALIFY	ERRONEOUS	HAMPER	INCOMPATIBILITY	KNOWINGLY
DISQUALIFYING	ERRONEOUSLY	HAMPERED	INCOMPATIBLE	LACK
DISREGARD	ERROR	HARASSMENT	INCOMPETENCE	LAGGED
DISREGARDED	ERRORS	HARM	INCOMPETENCY	LAGGING
DISREGARDING	ESCALATE	HARMED	INCOMPETENT	LAGS
DISRUPT	ESCALATED	HARMFUL	INCOMPLETE	LATE
DISRUPTED	ESCALATES	HARMING	INCONCLUSIVE	LAUNDERING
DISRUPTING	ESCALATING	HARMS	INCONSISTENCIES	LAYOFF
DISRUPTION	ESCALATION	HARSH	INCONSISTENCY	LAYOFFS

LIMITATION	OBJECTED	PROTESTED	SLOWER	UNDERPAYMENTS
LIMITATIONS	OBJECTING	PROTESTING	SLOWEST	UNDERSTATED
LIMITING	OBJECTION	PROTESTS	SLOWING	UNDERSTATEMENT
LIMITS	OBJECTIONABLE	PROTRACTED	SLOWLY	UNDESIRABLE
LINGERING	OBJECTIONS	PURPORT	SLUGGISH	UNDETECTED
LIQUIDATE	OBSCENE	PURPORTED	SPILL	UNDETERMINED
LIQUIDATED	OBSCENITY	PURPORTEDLY	SPILLAGE	UNDISCLOSED
LIQUIDATES	OBSCOLESCENCE	PURPORTING	SPILLED	UNDISCOVERED
LIQUIDATING	OBsolete	PURPORTS	SPILLING	UNDUE
LIQUIDATION	OBSTACLE	QUESTION	SPILLS	UNDULY
LIQUIDATIONS	OBSTACLES	QUESTIONABLE	STOLEN	UNECONOMIC
LIQUIDATOR	OFFENDING	QUESTIONED	STOPPAGE	UNECONOMICAL
LITIGANT	OMISSION	QUESTIONING	STOPPAGES	UNETHICAL
LITIGANTS	OMISSIONS	QUESTIONS	STOPPED	UNEXPECTED
LITIGATE	OMIT	QUIT	STOPPING	UNEXPECTEDLY
LITIGATED	OMITS	RACKETEER	STOPS	UNFAIR
LITIGATING	OMITTED	RACKETEERING	STRESS	UNFAIRLY
LITIGATION	OPPORTUNISTIC	RECALL	STRESSED	UNFAVORABLE
LOCKOUT	OPPORTUNISTICALLY	RECALLED	STRESSES	UNFAVORABLY
LOCKOUTS	OPPOSE	RECALLS	STRESSING	UNFIT
LOSE	OPOSED	RECESSION	STRINGENT	UNFORESEEABLE
LOSES	OPOSING	RECESSIONARY	SUBPOENA	UNFORESEEN
LOSING	OPPOSITION	RECESSIONS	SUBPOENAS	UNFORTUNATELY
MALFEASANCE	OVERAGE	RECKLESS	SUBSTANDARD	UNFULFILLED
MALFUNCTION	OVERBURDEN	RECKLESSLY	SUFFERED	UNINSURED
MALFUNCTIONING	OVERCAPACITY	RECKLESSNESS	SUFFERING	UNINTENDED
MALFUNCTIONS	OVERCHARGED	RECOMPUTED	SUFFERS	UNINTENTIONAL
MALICIOUS	OVERCHARGES	REDRESS	SURRENDER	UNINTENTIONALLY
MALPRACTICE	OVERCOME	REFUSAL	SURRENDERED	UNJUST
MANIPULATE	OVERDUE	REFUSALS	SURRENDERING	UNJUSTLY
MANIPULATED	OVERPAID	REFUSE	SURRENDERS	UNLAWFUL
MANIPULATING	OVERPRODUCTION	REFUSED	SUSPECT	UNLAWFULLY
MANIPULATION	OVERRUN	REFUSES	SUSPECTED	UNMARKETABLE
MARKDOWN	OVERRUNS	REFUSING	SUSPECTS	UNOCCUPIED
MARKDOWNS	OVERSTATED	REJECT	SUSPEND	UNPAID
MISAPPLICATION	OVERSTATEMENT	REJECTED	SUSPENDED	UNPERFORMED
MISAPPROPRIATE	OVERSUPPLY	REJECTING	SUSPENDING	UNPLANNED
MISAPPROPRIATED	OVERTLY	REJECTION	SUSPENDS	UNPREDICTABILITY
MISAPPROPRIATING	PENALIZED	REJECTIONS	SUSPENSION	UNPREDICTABLE
MISAPPROPRIATION	PENALTIES	REJECTS	SUSPENSIONS	UNPRODUCTIVE
MISCALCULATION	PENALTY	RELUCTANCE	SUSPICIOUS	UNPROFITABLE
MISCALCULATIONS	PERJURY	RELUCTANT	TERRORISM	UNQUALIFIED
MISCHIEF	PICKETING	REPORTEDLY	TERRORIST	UNREASONABLE
MISCONDUCT	PLAINTIFF	REPOSSESSED	TESTIFY	UNREASONABLY
MISDEMEANOR	PLAINTIFFS	REPOSSESSION	THREAT	UNRECOVERABLE
MISFEASANCE	PLEA	REPOSSESSIONS	THREATEN	UNRELIABLE
MISLEAD	PLEAD	REPUDIATE	THREATENED	UNREPORTED
MISLEADING	PLEADED	REPUDIATED	THREATENING	UNREST
MISMANAGEMENT	PLEADING	REPUDIATION	THREATENS	UNSAFE
MISREPRESENTATION	PLEADINGS	RESIGN	THREATS	UNSATISFACTORY
MISREPRESENTATIONS	PLEAS	RESIGNATION	TOLERATE	UNSATISFIED
MISREPRESENTED	PLED	RESIGNATIONS	TOLERATED	UNSAVORY
MISREPRESENTING	POOR	RESIGNED	TORNADO	UNSOLD
MISS	POORLY	RESIGNING	TORNADOES	UNSOUND
MISSED	POSES	RESIGNS	TORTUOUS	UNSTABLE
MISSTATED	POSING	REVOKE	TRAGIC	UNSUCCESSFUL
MISSTATEMENT	POSTPONE	REVOKED	TRAUMATIC	UNSUCCESSFULLY
MISSTATEMENTS	POSTPONED	REVOKES	TROUBLE	UNSUITABILITY
MISTAKE	POSTPONEMENT	REVOKING	TROUBLED	UNSUITABLE
MISTAKEN	POSTPONEMENTS	RISKIER	TROUBLES	UNSURE
MISTAKES	POSTPONING	RISKY	UNABLE	UNSUSPECTED
MISUNDERSTANDING	PRECIPITATED	SABOTAGE	UNACCEPTABLE	UNTIMELY
MISUSE	PREJUDICE	SECRECY	UNANTICIPATED	UNUSABLE
MISUSED	PREJUDICED	SECRET	UNAPPROVED	UNUSUAL
MORATORIUM	PREJUDICES	SEIZE	UNATTRACTIVE	UNUSUALLY
MORATORiums	PREMATURE	SEIZED	UNAUTHORIZED	UNWANTED
NEGATIVE	PREMATURELY	SEIZURE	UNAVAILABILITY	UNWARRANTED
NEGATIVELY	PRESSING	SERIOUS	UNAVAILABLE	UNWILLING
NEGLIGENCE	PRETRIAL	SERIOUSLY	UNAVOIDABLE	UNWILLINGNESS
NEGLIGENT	PROBLEM	SEVER	UNAWARE	URGENCY
NEGLIGENTLY	PROBLEMATIC	SEVERE	UNCOMPLETED	URGENT
NONCOMPETITIVE	PROBLEMS	SEVERED	UNCONTROLLABLE	USURIOUS
NONCOMPLIANCE	PRONE	SEVERELY	UNCONTROLLED	USURY
NONCONFORMING	PROSECUTE	SEVERING	UNCORRECTED	VANDALISM
NONCONFORMITY	PROSECUTED	SEVERITY	UNCOVER	VERDICT
NONDISCLOSURE	PROSECUTES	SLANDER	UNCOVERED	VERDICTS
NONPAYMENT	PROSECUTING	SLOW	UNDERESTIMATE	VIOLATE
NONPERFORMANCE	PROSECUTION	SLOWDOWN	UNDERMINE	VIOLATED
NONPERFORMING	PROSECUTIONS	SLOWDOWNS	UNDERPAID	VIOLATES
NONPRODUCTIVE	PROTEST	SLOWED	UNDERPAYMENT	VIOLATING

VIOLATION	BENEFICIALLY	ENHANCING	PRESTIGIOUS	ASSUMES
VIOLATIONS	BENEFIT	ENJOY	PROACTIVE	ASSUMING
VIOLATORS	BENEFITED	ENJOYABLE	PROFICIENCY	ASSUMPTION
VIOLENCE	BENEFITING	ENJOYED	PROFITABILITY	ASSUMPTIONS
VIOLENT	BENEFITTED	ENJOYING	PROFITABLE	BELIEVE
VOLATILE	BENEFITTING	ENJOYMENT	PROFITABLY	BELIEVES
VOLATILITY	BETTER	ENJOYS	PROGRESS	CONDITIONAL
VULNERABILITIES	BOLSTERED	ENTHUSIASM	PROGRESSED	CONDITIONALLY
VULNERABILITY	BOOM	ENTHUSIASTIC	PROGRESSES	CONFUSINGLY
VULNERABLE	BOOST	EXCELLENCE	PROGRESSING	CONFUSION
WARN	BOOSTED	EXCELLENT	REBOUND	CONTINGENCIES
WARNING	BREAKTHROUGH	EXCEPTIONAL	REBOUNDED	CONTINGENCY
WARNINGS	BREAKTHROUGHS	EXCEPTIONALLY	REGAIN	CONTINGENT
WEAK	CHARITABLE	EXCITED	REGAINED	CONTINGENTLY
WEAKEN	CHARITIES	EXCITEMENT	REWARD	CROSSROADS
WEAKENED	CHARITY	EXCITING	REWARDED	DEPEND
WEAKENING	COLLABORATE	EXCLUSIVE	REWARDING	DEPENDENT
WEAKENS	COLLABORATED	EXCLUSIVELY	REWARDS	DEPENDENCE
WEAKER	COLLABORATING	EXCLUSIVITY	SATISFACTION	DEPENDENCIES
WEAKEST	COLLABORATION	EXEMPLARY	SATISFACTORILY	DEPENDENCY
WEAKNESS	COLLABORATIONS	FAVORABLE	SATISFACTORY	DEPENDING
WEAKNESSES	COLLABORATIVE	FAVORABLY	SATISFIED	DEPENDS
WILLFULLY	COLLABORATOR	FRIENDLY	SATISFIES	DEVIATE
WORSE	COLLABORATORS	GAIN	SATISFY	DEVIATION
WORSEN	COMPLIMENT	GAINING	SATISFYING	DEVIATIONS
WORSENER	COMPLIMENTARY	GREAT	SMOOTH	DIFFERED
WORSENING	CONCLUSIVE	GREATER	SMOOTHLY	DIFFERING
WORST	CONCLUSIVELY	GREATEST	STABILITY	DOUBTFUL
WORTHLESS	CONDUCTIVE	GREATLY	STABILIZATION	DOUBTS
WRONG	CONSTRUCTIVE	HAPPY	STABILIZE	EXPECT
WRONGDOING	CONSTRUCTIVELY	HIGHEST	STABILIZED	FLUCTUATE
WRONGFUL	COURTEOUS	HONOR	STABILIZING	FLUCTUATED
WRONGFULLY	CREATIVE	HONORABLE	STABLE	FLUCTUATES
	CREATIVITY	HONORED	STRENGTH	FLUCTUATING
Fin-Pos (positive)	DELIGHTED	HONORING	STRENGTHEN	FLUCTUATION
ABLE	DEPENDABILITY	HONORS	STRENGTHENED	FLUCTUATIONS
ABUNDANCE	DEPENDABLE	IDEAL	STRENGTHENING	IMPRECISE
ABUNDANT	DESIRABLE	IMPRESSED	STRENGTHENS	IMPRECISION
ACCOMPLISH	DESIRED	IMPRESSIVE	STRENGTHS	IMPROBABLE
ACCOMPLISHED	DESPITE	IMPROVE	SUCCEED	INDEFINITE
ACCOMPLISHING	DESTINED	IMPROVED	SUCCEDED	INDEFINITELY
ACCOMPLISHMENT	DILIGENT	IMPROVEMENT	SUCCESSING	INDETERMINABLE
ACCOMPLISHMENTS	DILIGENTLY	IMPROVEMENTS	SUCCEEDS	INDETERMINATE
ACHIEVE	DISCOVER	IMPROVES	SUCCESS	INTANGIBLE
ACHIEVED	DISCOVERED	IMPROVING	SUCCESSSES	INTANGIBLES
ACHIEVEMENT	DISCOVERIES	INFORMATIVE	SUCCESSFUL	PENDING
ACHIEVEMENTS	DISCOVERING	INNOVATE	SUCCESSFULLY	POSSIBILITIES
ACHIEVES	DISCOVERS	INNOVATION	SURPASS	POSSIBILITY
ACHIEVING	DISCOVERY	INNOVATIONS	SURPASSED	POSSIBLE
ADEQUATE	DISTINCTION	INNOVATIVE	SURPASSING	POSSIBLY
ADEQUATELY	DISTINCTIONS	INNOVATOR	TRANSPARENCY	PREDICTABILITY
ADVANCED	DISTINCTIVE	INTEGRITY	TREMENDOUS	PREDICTING
ADVANCEMENT	DONATION	INVENT	UNPARALLELED	PREDICTION
ADVANCEMENTS	DONATIONS	INVENTED	UNPRECEDENTED	PREDICTIONS
ADVANCES	DRAMATIC	INVENTION	UPTURN	PREDICTIVE
ADVANCING	DRAMATICALLY	INVENTIONS	VERSATILE	PREDICTOR
ADVANTAGE	DREAM	INVENTOR	VERSATILITY	PREDICTS
ADVANTAGED	EASIER	INVENTORS	VIBRANT	PRESUMABLY
ADVANTAGEOUS	EASILY	LEADERSHIP	WORTHY	PRESUME
ADVANTAGEOUSLY	EASY	LEADING		PRESUMED
ADVANTAGES	EFFECTIVE	LUCRATIVE	Fin-Unc (uncertain)	PRESUMPTION
ALLIANCE	EFFICIENT	MERITORIOUS	ABEYANCE	PRESUMPTIONS
ALLIANCES	EFFICIENTLY	OPPORTUNITIES	AMBIGUITIES	PROBABILITIES
APPRECIATE	EMPOWER	OPPORTUNITY	AMBIGUITY	PROBABILITY
APPRECIATED	EMPOWERED	OUTPERFORM	AMBIGUOUS	PROBABLE
APPRECIATES	EMPOWERING	OUTPERFORMED	ANOMALIES	PROBABLY
APPRECIATION	EMPOWERS	OUTSTANDING	APPEARED	RANDOM
ASSURE	ENABLE	PERFECT	APPEARING	RANDOMLY
ASSURED	ENABLED	PERFECTLY	APPEARS	REEXAMINATION
ASSURES	ENABLES	PLEASANT	APPROXIMATE	REEXAMINE
ASSURING	ENABLING	PLEASED	APPROXIMATED	REINTERPRETED
ATTAIN	ENCOURAGED	PLEASURE	APPROXIMATELY	ROUGHLY
ATTAINED	ENCOURAGEMENT	POPULAR	APPROXIMATES	RUMORS
ATTAINING	ENCOURAGES	POPULARITY	APPROXIMATING	SOMEWHAT
ATTAINMENT	ENCOURAGING	POSITIVE	APPROXIMATION	SPECULATE
ATTAINS	ENHANCE	POSITIVELY	APPROXIMATIONS	SPECULATION
ATTRACTIVE	ENHANCED	PREEMINENT	ARBITRARILY	SPECULATIVE
BEAUTIFUL	ENHANCEMENT	PREMIER	ARBITRARY	SUDDEN
BEAUTY	ENHANCEMENTS	PREMIERE	ASSUME	SUDDENLY
BENEFICIAL	ENHANCES	PRESTIGE	ASSUMED	UNCERTAIN

UNCERTAINTIES	JURISDICTIONAL
UNCERTAINTY	JURY
UNCLEAR	JUSTICE
UNFAMILIAR	LAWYER
UNIDENTIFIABLE	LAWYERS
UNIDENTIFIED	LEGAL
UNKNOWN	LEGALITY
UNPROVED	LEGALIZED
UNPROVEN	LEGALLY
UNRESOLVED	LEGATEE
UNSEASONABLY	LEGATEES
UNSETTLED	LEGISLATED
UNSPECIFIED	LEGISLATION
VARIABILITY	LEGISLATIVE
VARIABLE	LEGISLATORS
VARIABLES	LEGISLATURE
VARIANCE	LEGISLATURES
VARIANCES	PLAINTIFF
VARIATION	PLAINTIFFS
VARIATIONS	PREJUDICIAL
VARIED	PROBATE
VARIES	QUITCLAIM
VARY	REBUTTAL
VARYING	REGULATION
	REGULATIONS
Fin-Lit (litigious)	REGULATOR
ADJUDGED	REGULATORS
ADJUDGING	REGULATORY
ADJUDICATE	REHEARING
ADJUDICATED	REMAND
ADJUDICATING	REMANDED
ADJUDICATION	REMANDING
ADMISSIBILITY	RULING
ADMISSIBLE	RULINGS
ADMISSION	SURETIES
ADMIT	SURETY
ADMITS	TESTIMONY
AFFIDAVIT	TORT
AFFIDAVITS	TORTS
AGGRIEVED	WRIT
ALLEGATION	WRITS
ALLEGATIONS	
ALLEGE	Modal Words – Strong
ALLEGED	ALWAYS
ALLEGEDLY	BEST
ALLEGES	CLEARLY
ALLEGING	DEFINITELY
APPELLATE	HIGHEST
ARBITRATE	IS
ARBITRATED	LOWEST
ARBITRATION	MUST
ARBITRATOR	NEVER
ARBITRATORS	SHALL
ATTEST	UNDOUBTEDLY
ATTESTATION	WILL
ATTESTED	
ATTESTING	Modal Words – Weak
ATTORNEY	CONCEIVABLE
ATTORNEYS	COULD
BREACH	DEPENDING
BREACHED	DEPENDS
BREACHES	MAY
BREACHING	MAYBE
CLAIMANT	MIGHT
CLAIMANTS	OCCASIONALLY
COLLUSION	PERHAPS
COMPENSATORY	POSSIBLE
DEFENDANT	POSSIBLY
DEFENDANTS	SELDOMLY
DEPOSE	SOMETIMES
DEPOSITION	SUGGEST
DEPOSITIONS	UNCERTAINLY
DESIST	
INTERLOCUTORY	
INTERROGATORIES	
JUDICIAL	
JUDICIALLY	
JUDICIARY	
JURIES	
JURISDICTION	