Prospectus Words, Parts of Speech, and Uncertainty in IPO Pricing

Raymond P. H. Fishe
David North
and
Aaron Smith*

This version: September 2015

ABSTRACT

We investigate how uncertainty is shaped from words in the risk factor section of IPO prospectuses using a new method to identify words that predict IPO first-day returns. We find that sentiment lists represent only a small fraction of the prospectus words associated with underpricing. When sentiment words do overlap our new words, there is a significantly greater impact on returns to such sentiment words. We investigate whether parts of speech help explain the development of investor sentiment. Nouns are found to be important determinants of first-day returns, but are under-represented on sentiment lists. Such nouns are often connected to an industry or technical function and are not really tonal. Verbs and adjectives appear as the most meaningful words on existing sentiment lists.

Keywords: Initial Public Offering (IPO), Investor Sentiment, False Discovery Rate, Parts of Speech, and Textual Analysis

JEL classification: G12, G14, G24

*Fishe: Department of Finance, Robins School of Business, University of Richmond, Richmond, VA 23173. Tel: (+1) 804-287-1269. Email: pfishe@richmond.edu. North: Department of Finance, Robins School of Business, University of Richmond, Richmond, VA 23173. Tel: (+1) 804-287-1269. Email: dnorth@richmond.edu. Tel: (+1) 804-287-1938. Smith: Department of Agricultural and Resource Economics, University of California at Davis, Davis, CA 95616. Email: adsmith@ucdavis.edu. Tel: (+1) 530-752-2138. We thank Richard Haynes, Steve Kane, Cassandra Marshall, Esen Onur, Michel Robe, and Steve Tallman for comments on this topic; Dylan Herman and Robert Little provided excellent research assistance. Corresponding author: Raymond P. H. Fishe.
Prospectus Words, Parts of Speech, and Uncertainty in IPO Pricing

I. INTRODUCTION

Investor perceptions play an important role in explaining asset returns, particularly for initial public offerings (IPOs) where there is no market history of news or announcement effects. Beatty and Ritter (1986) observe that investors require greater returns when an IPO is perceived to have more ex ante uncertainty. They find support for this idea using “use of proceeds” and “gross proceeds” data from the prospectus to measure ex ante uncertainty. Loughran and McDonald (2013) take a textual analysis approach to uncertainty by constructing a measure of prospectus sentiment based on word lists that reflect “negative”, “strong modal” and “uncertain” tones. They report that IPOs with greater use of these tonal words are more underpriced. Other researchers also find that prospectus sentiment affects first-day returns in IPOs (e.g., Ferris, Hao, and Liao, 2013; Feuerriegel, Schmitz, and Neumann, 2014).

The goal of this research is to examine how comprehensive sentiment list words are in capturing textual information correlated with IPO underpricing. We seek to define the set of words correlated with underpricing and to examine how much overlap they have with sentiment list words. In effect, we are investigating whether the connection between textual information in IPO prospectuses and initial returns is largely driven by sentiment words. Our hypothesis is that an investor’s view of an IPO and its risk is likely framed by many words—sentiment words (of course), but also words that characterize the industry, the firm’s product or service, or those that

---

connect to important risks faced by the firm. This research will investigate the validity of that hypothesis.

To find a broad set of correlated words we develop a new method to connect words to initial IPO returns. The method identifies two sets of words: one associated with high and another with low underpricing. The words associated with high underpricing are expected to be connected to companies whose IPOs produced high levels of uncertainty during the bookbuilding process.² We expect these words to have a meaningful overlap with sentiment words. Surprisingly, our analysis finds hundreds of new words that are not among those found on one of the Loughran and McDonald (2011) sentiment lists³, including words such as “bandwidth”, “networking”, and “engineering”. Such words suggest that investors may equate uncertainty to an emerging industry or technology, and thus additional context may be necessary to capture a typical investor’s view after reading an IPO prospectus.

The need for additional context is also emphasized by the cases in which our correlated words overlap with sentiment words. Overall, about 21% of our new words also appear on a LM tonal list. The majority of these overlaps occur with words on the “negative” sentiment list, but there are some overlaps with “positive”, “uncertainty”, and “litigious” lists. Importantly, we find overlaps with both words associated with high underpricing and words associated with low underpricing, and those words are on the same sentiment list. Thus, even sentiment list words may require additional context for investors to establish the true tonal implications of certain

² The “uncertainty” theory of IPO underpricing is also consistent with the Benveniste and Spindt (1989) and Sherman and Titman (2002) bookbuilding models. In those models, investors are compensated with underpricing for revealing valuable information to underwriters. The more uncertainty there is about a firm’s prospects, then the more information an underwriter may seek to help with initial pricing. Hence, there is likely to be greater underpricing to pay for this additional information.

words. In short, we find support for the hypothesis that high levels of uncertainty in IPOs are captured by many words only a fraction of which are found on popular sentiment lists.

Most of the LM words do not appear on our lists, so any predictive power of the LM words must come from a small subset. We illustrate this point by regressing first day returns on the fraction of all words in a document that appear on a particular list. We find that a one standard deviation change in the fraction of negative tone words in a document results in an 8.3% change in initial returns or about one-eighth the standard deviation of sample returns. However, the sentiment words that overlap our word lists have predicted impacts two to five times larger. Moreover, the non-overlapping words exhibit small estimated effects of the opposite sign. Thus, the FDR-derived words are an effective filter to isolate the sentiment words most correlated with subsequent returns.

Our method to find such correlated words is quite general and may be applied when there are fewer documents than words to test. To elaborate, we examine IPO prospectuses in the upper (and lower) 5% and 10% tails of the distribution of first-day returns. Using these tail documents, we select new words by testing whether each word appears more often than expected in the upper (or lower) tail of returns. For example, in our sample there are approximately 139 IPOs that comprise the upper 10% of first-day returns. If a given word appears in 200 prospectuses, then without other conditioning information we would expect that 20 of these prospectuses are in the upper tail. If we find significantly more than 20 prospectuses with this word in the upper 10% tail, then this word is flagged as possibly predictive of high underpricing.

---

4 Jegadeesh and Wu (2013) propose a method to score document tone using weights derived from regressing 10-K filing returns on word frequencies. The sign of these weights create positive and negative word tone groups. This method requires that the number of documents exceed the number of words, which is a limiting condition in many document samples including the IPO sample used here.
Our method requires us to conduct a large number of hypothesis tests, and our words lists are comprised of all words for which we reject the relevant null hypothesis. If we used standard normal critical values in these tests, then a large number of words may be flagged as correlated with tail returns purely by luck. This is known as a multiple testing problem. To address this problem and thereby limit these lucky words, we use the False Discovery Rate (FDR) method of Benjamini and Hochberg (1995). The FDR increases the critical value for significance so that the expected proportion of lucky words on our word lists is no more than 5%. This gives us confidence that our word lists are dominated by words that are truly associated with high (or low) first-day performance.

We implement our FDR-based method to include control variables—identified by other IPO theories—that are known to affect IPO first-day returns. With this approach, new words are identified after controlling for other factors, including industry effects using Fama-French 48-industry factors. Thus, these new words give support to the view that textual documents contain important pricing information beyond that found using other quantitative variables and beyond the specific risks of a given industry (cf., Hanley and Hoberg, 2010; Jegadeesh and Wu, 2013).

We validate our method using out-of-sample tests. We divide the sample in half and use the words found in the first half to predict first-day returns in the second half and vice versa. The sample division we use is consistent with the economies of scope findings in Gao, Ritter, and

---

5 Recent applications of the FDR method in finance include: Barras, Scaillet, and Wermers (2010) who sought to identify fund managers with significant alpha performance, Bajgrowicz and Scaillet (2012) who examine the success of technical trading rules applied to the DJIA index since 1897, Fishe and Smith (2012) who examined whether futures traders were informed of price changes over short horizons, and Harvey, Liu, and Zhu (2014) who examine multiple testing biases for asset pricing models.

6 Another method that may be used to identify words correlated with IPO returns is the least absolute shrinkage and selection operator, or LASSO, as described in Tibshirani (1996). This is a data mining routine that may systematically find the best set of correlated words. In contrast to our approach, LASSO takes no account of the Type I errors created by the systematic inclusion and elimination of words in the solution method.

7 See Ritter and Welch (2002) and Eckbo, Masulis, and Norli (2007) for discussions of the variables that affect underpricing in IPOs.
Zhu (2013). We show that there is significant additional explanatory power using the new words in both sets of predictive regressions.

We also explore whether the words we select and sentiment words are connected to sentence structure. Well written documents rely on nouns, verbs, adjectives, and adverbs as basic building blocks for each sentence. We examine whether these parts of speech act also as basic building blocks of investor sentiment. We compare the use of words on the sentiment lists to the use of words found with the FDR-based method. We find that many of the FDR-based words are nouns or equivalent to proper nouns and this part of speech is generally under-represented on sentiment lists, which focus more on adjectives, verbs, and adverbs. In effect, there is additional information to be found in the nouns used in the IPO prospectus.

Additionally, we decompose sentiment measures in first-day return regressions, parsing them into nouns, verbs, adverbs, and adjectives. Except for the litigious tonal words, these results show that nouns on sentiment lists are insignificant words. Our results also show that verbs and adjectives on the LM negative and combined negative/uncertain/weak modal lists are significant predictors of initial returns, so these lists capture many important action and descriptive words. Overall, we find evidence that parts of speech used in IPO prospectuses appear to significantly influence investor perceptions as measured by underpricing.

Our data include all IPOs issued between 1998 and 2005. After filtering for standard exclusions and data omissions, the sample contains 1,391 IPOs. We extract words from the Risk Factors section in each IPO prospectus as this section is known to be informative for initial

---

8 Other parts of speech are excluded here, such as pronouns, prepositions, conjunctions, and interjections. These word types are also important to sentence construction and meaning, but they are widely dispersed across documents and thus offer little discriminatory power for the FDR-based method.

9 The word filters we use eliminate proper names, but some still match words in the dictionary. The words “Oracle”, “Yahoo”, “Sun”, and “SAP” appear on our new word lists with all referring to companies in the IPO prospectuses, and not the underlying dictionary word.
pricing decisions (Hanley and Hoberg, 2010). Moreover, disclosure rules require firms to identify all material risks, so investors seeking to understand the uncertainty surrounding a firm’s future prospects will be expected to consider the risks described in this section. Counting each word no more than once within each risk factor section, the final document sample includes over 1.3 million words of which there are approximately 16,000 unique words. To preserve the power of our statistical tests, we examine the 4,508 unique words that appear in 30 or more prospectuses. The Appendix discusses the robustness of our findings to this restriction.

This paper proceeds as follows. In the next section, we describe the methods that we use to identify words significantly correlated with first-day returns. We discuss the FDR approach and explain how to condition our methods on other variables. Section 3 presents the IPO data used in this study, Section 4 offers our analytical results, and Section 5 provides our conclusions.

II. METHOD

To identify new words, we use a similar framework to Sebastiani (2002). We summarize each document by counting the number of times each word appears in that document. Thus, we define the variable \( w_j^d \) to equal the count of occurrences of word \( j \) in document \( d \).\(^{10}\) For a given document \( d \), the set \( \{w_1^d, w_2^d, w_3^d, \ldots, w_J^d\} \) characterizes the word content, where \( J \) is the length of the word vector.

Each document is associated with a binary success measure, \( x_d \in \{0,1\} \), that is the focus of study. In this study it is inferior or superior first-day IPO returns, but it could be accounting fraud, poor stock returns after an annual report filing, high trading volume following news about

\(^{10}\) We simplify the definition of documents to the frequency count of words as word counts are often used to examine category effects. However, documents may also be identified by other characteristics of words, such as the number of syllables or parts of speech (nouns, verbs, adjectives, etc.), but this generality is omitted here.
a company, etc. The empirical challenge is to identify the set of words $S$ that are more likely than average to appear in successful documents, i.e.,\(^{11}\)

$$S = \{j \in \mathcal{W}: Pr(w_j^d > 0|x_d = 1) > Pr(w_j^d > 0)\},$$

(1)

where $\mathcal{W}$ denotes the set of all words. To identify the full set $S$ with probability one, we would require an arbitrarily large number of sample documents. In a finite sample, we will only be able to identify the subset of words that we observe enough times that we can reject the null hypothesis that $Pr(w_j^d > 0|x_d = 1) = Pr(w_j^d > 0)$.

As with any statistical classification procedure, we face the possibility of making type I and type II errors. We want to identify a group of meaningful words, so we want to avoid diluting our set of selected words with words that are not meaningful, i.e., we wish to control the rate of erroneous membership in our lists. We use the false discovery rate method to select a set of words that is expected to contain at least 95% meaningful words (Benjamini and Hochberg, 1995). We describe how we implement the FDR method after introducing our success measures $x_d$ in the next section.

A. Measuring Word Success

We use the terms “superior” and “inferior” to signify returns that lie in the right and left tails of the empirical distribution, respectively. Specifically, our binary success measure is $x_d = 1(\ r_d < q_p \ )$ for the lower tail and $x_d = 1(\ r_d > q_p \ )$ for the upper tail, where the function, $1(\bullet)$, acts as a binary operator based on the given expression, $r_d$ is the return on the stock associated with document $d$, and $q_p$ is the threshold that defines a superior (or inferior) return. We define the threshold to be the $p^{th}$ quantile of the return distribution.

---

\(^{11}\) Equivalently, from Bayes rule, we could define $S$ as $S = \{j \in \mathcal{W}: Pr(x_d = 1|w_j^d > 0) > Pr(x_d = 1)\}$. 
The sample frequency of word $j$ in the lower tail is given by:

$$F_{j,\text{low}} = \frac{\sum_{d=1}^{D} 1(r_d < q_p, w_d^j > 0)}{\sum_{d=1}^{D} 1(w_d^j > 0)}, \quad (2)$$

and the frequency in the upper tail is similarly expressed as:

$$F_{j,\text{up}} = \frac{\sum_{d=1}^{D} 1(r_d > q_p, w_d^j > 0)}{\sum_{d=1}^{D} 1(w_d^j > 0)}. \quad (3)$$

For example, if a particular word appears in a total of 50 documents, then by random selection we would expect five of these documents to be in the lower 10th percentile of the distribution of the performance variable. To the extent that $F_{j,\text{low}} > 0.10$, we have evidence that this particular word is associated with more documents in the lower extremes of the performance variable than expected from random selection. Similarly, for the upper tail with a percentile cutoff at 90%, a finding that $F_{j,\text{up}} > 0.10$ is evidence that the word is associated with documents in the upper extremes of the performance variable.

Given the characterization of the performance variable percentiles, both $F_{j,\text{low}}$ and $F_{j,\text{up}}$ are proportions governed by parameters of the hypergeometric distribution. With a sufficient size for the total sample relative to $\sum_{d=1}^{D} 1(w_d^j > 0)$, the normal approximation may be applied to test the lower tail null, $H_0: F_{j,\text{low}} = \frac{n}{100}$, or the upper tail null, $H_0: F_{j,\text{up}} = \frac{100-n}{100}$, for each word $j$.

This approach generalizes to a regression framework when theory argues that other variables are important determinants of the performance variable, as is the case with IPO first-day returns.

Suppose that the performance variable is modeled as:

---

12 The tests for each word are from a subsample without replacement from the entire sample, which makes the test statistic follow a hypergeometric distribution. Given the sample sizes here, the binomial distribution is also a close approximation.

13 The literature on word content analysis suggests that certain words have more impact than other words based on frequency of use (Manning and Schütze, 2003; Jurafsky and Martin, 2009). Using median use across IPOs, we have estimated a word count variation of our method with high use words identified as those that exceed the median count benchmark. This extension identifies fewer significant words but does not meaningfully affect the basic results presented here.
where $\alpha$ and $\beta$ (a vector) are model parameters, $X_d$ is a matrix of explanatory variables, and $\epsilon_d$ is an i.i.d. random variable with the usual properties for ordinary least squares analysis. Note that using the empirical distribution of $r_d$ for percentiles gives the same success rates as using $\hat{\epsilon}_d = r_d - \bar{r}$ for percentiles. Thus, to allow other determinants to affect the performance variable, we first estimate eq. (4) by ordinary least squares and then evaluate the percentile function using the vector of regression residuals. Now the extremes of the distribution are conditional on having adjusted the performance variable for other known control variables. If a sizeable number of words are found to be significant after the regression adjustments, then this is evidence that word content information is not fully captured by the variables in $X_k$.

B. The False Discovery Rate

The methods described above generate test statistics for each word, which give rise to a multiple testing problem because there are thousands of unique words in most document samples (Miller, 1981).\(^{14}\) Although there are numerous approaches to this problem—such as control of the family-wise error rate or combining tests into single statistics—the framework developed by Benjamini and Hochberg (1995) and Storey (2002, 2004) to control the FDR has been convincingly applied to problems in finance (e.g., Barras, Scaillet, and Wermers, 2010; Bajgrowicz and Scaillet, 2012; Fishe and Smith, 2012; Harvey, Liu, and Zhu, 2014). The FDR equals the proportion of rejected hypotheses that are true. By controlling the FDR, we limit the number of words incorrectly classified as related to the extremes of the performance variable.

To explain the FDR method, we use the example of a test for words connected to superior returns, i.e., the statistic in eq. (3). We posit three types of words, non-positive words with a less

\(^{14}\) Reducing words to root references, such as “declare” for declared, declaring, or declaration will help reduce multiple testing problems, but as our sample of IPOs shows there will still be thousands of test statistics.
than average chance of appearing in superior-return documents, null words with no connection to superior returns, and positive words with a greater than average chance of appearing in superior-return documents. The true population proportion of null words is denoted in this literature by $\pi_0$.

For each word $j$ found in a subset of the sample documents, we calculate a test statistic, $z_j$. This statistic has a known (possibly asymptotic) distribution under the null and is centered away from zero under the two alternative hypotheses. Consider a critical value, $c$, for a one-sided test of the hypothesis that the word is non-positive or null against the positive alternative.

Suppose that we randomly pick the $j^{th}$ word from among those for whom the null hypothesis is rejected. The FDR is the probability that this word is null, defined by:

$$FDR(c) = Pr\left( j \in \text{null} \mid z_j > c \right)$$

$$= \frac{Pr(z_j > c \mid j \in \text{null}) Pr(j \in \text{null})}{Pr(z_j > c)}$$

$$= \frac{Pr(z_j > c \mid j \in \text{null}) \pi_0}{Pr(z_j > c)}$$

where this derivation assumes that $z_j$ never exceeds $c$ for non-positive words, i.e., words for which $z_j > c$ are either null or positive.

Using a five percent error rate for the FDR control size, we choose the minimum critical value $c$ such that $FDR(c) \leq 0.05$. With this control, we reject the null hypothesis for each word with a $z$-statistic that exceeds $c$. Formally, our chosen set of words is

$$\tilde{S} = \max\{j \in W : Pr(j \in S \mid z_j > c) \leq 0.05\} \mu(W),$$

where $\mu$ is the counting measure.

Our method selects the largest set of meaningful words that is expected to contain no more than 5 percent null cases. This set will generally not contain all words in $S$. We will omit those words in $S$ that do not produce a large $z$ statistic, either because the word is weakly related to
extreme returns or because the word appears in relatively few documents and therefore doesn’t give us enough power to detect it. This omission is the price we pay for limiting the false discovery rate and is analogous to the tradeoff between size and power in a standard testing problem.

The key to implementing the FDR approach is to efficiently estimate the three terms in Eq. (5) using a sample of document words. We follow the methods described in Storey (2002) and Fishe and Smith (2012) to estimate these parameters.

The FDR approach has known benefits versus other multiple-testing methods. The critical value \( c \) is independent of the number of words tested because it controls the proportion of null cases for which the null hypothesis is rejected. In contrast, controlling the family-wise error rate requires the critical value to increase with the number of tests. Because our objective is to identify words meaningfully connected to extremes of a performance variable, we do not want a method that depends on the number of other words considered. The FDR approach also adjusts the critical value based on the probability a word is null (\( \pi_0 \)) in the population. As \( \pi_0 \) increases, the likelihood increases that a given tested word is null, and thus the FDR approach chooses a larger critical value.

Like most tests, however, the FDR method is not immune to violations of basic model assumptions. One assumption that may affect our results is that these test statistics are independent. Some words will surely satisfy this assumption, but others may be connected via sentence structure, “boiler plate” effects in prospectuses, or other latent dependencies. Fortunately, Benjamini and Yekutieli (2001) and Kim and van de Wiel (2008) show that the FDR method is quite robust when addressing samples with test dependencies. Even so, we offer
a note of caution that our 5% control rate may understate the true FDR if there are many such dependencies.

III. Data

Our sample of IPO prospectuses covers the years 1998–2005, during which over 2,000 firms went public. We identify these IPOs from the Securities Data Corporation (SDC) database. From those data we exclude: American Depository Receipts, unit offerings, limited partnerships, Real Estate Investment Trusts, closed-end funds, and those IPOs that are missing key variables, such as prior assets, returns or age information. We also remove firms with an offer price less than $5 per share and firms that are not available in the Center for Research on Security Prices (CRSP) database. After exclusions, the final sample contains 1,391 IPOs.

The final prospectus filing (SEC Form 424b) provides the document sample for our tests. Each prospectus is divided into several sections based on SEC rules: summary, risk, proceeds, dilution, selected financial data, management’s discussion and analysis, etc. We only examine the risk factors section in each prospectus. Hanley and Hoberg (2010, Table 5) find that this section offers significant “informative” content for price revisions during the filing period, although the management discussion section may also be informative. Importantly, Beatty and Welch (1996) suggest that the number of risk factors in the prospectus is related to ex ante uncertainty about the IPO. Addou and Dicle (2007) also show evidence that selected risk factors are important determinants of IPO underpricing during the internet bubble period in 1999-2000.

---

15 Jay Ritter has documented numerous errors in the SDC database (see his website and the document [http://bear.warrington.ufl.edu/ritter/SDC_corrections.pdf](http://bear.warrington.ufl.edu/ritter/SDC_corrections.pdf)). We correct errors in the initial price range, IPO price, days in registration, and ticker symbol variables found in the SDC database. We also use the Bloomberg system and company filings to correct several entries for offer dates and initial price ranges.

16 Loughran and McDonald (2013) find little difference in the text between the initial prospectus (Form S-1) and the final prospectus, so it unlikely to matter which document is used for this analysis.

17 We determine this from the significance of the informative content variable in the Hanley and Hoberg (2010) regressions measuring absolute change in offer price during the filing period.
The risk factors section aims to discuss all of the material risks individuals should consider before investing and is typically the first or second largest section in the prospectus based on the number of words. These facts suggest that it contains substantial information. Thus, if there is an association between word content, ex ante uncertainty, and initial pricing, it is at a minimum expected to be connected to the risk factor section.

We also use the word lists developed by Loughran and McDonald (2011, LM), who developed their lists because they found that the classification scheme in the Harvard IV-4 Psychosociological Dictionaries was poorly suited for financial applications. The principal problem was that they improperly assigned some words to a positive or negative tone when such words are actually neutral in a financial context. Examples are *taxes* and *cost*, which are classified as negative tone in the Harvard list, but carry no obvious tonal connotation in an IPO prospectus, 10-K report, or most other regulatory filings.

To create word vectors for each document, we keep both root words and their inflections, as inflections may add emphasis that is lost otherwise. However, we do filter a sizable number of other words from the risk factors section. We remove generic words, articles of speech, conjunctions, personal pronouns, prepositions, abbreviations, and single character words (199,274), stand-alone numbers and words with embedded numbers or special characters (85,679), and non-dictionary words (69,623) using the updated 12of12inf Master dictionary provided by Bill McDonald. After filtering and counting each word once within each prospectus, the final document sample includes 1,369,849 words of which there are 16,352 unique words.

---

18 Bill McDonald (http://www3.nd.edu/~mcdonald/Word_Lists.html) provides a comprehensive dictionary based on the “12of12inf” collection of common words. He has updated the base dictionary to include common words found in Form 10-K filings. The web site also includes tonal word lists developed by Loughran and McDonald (2010) and lists of “stop words,” such as names, dates, geographic references, and 278 generic words.
The first-day IPO return is the performance variable in our analysis. To compute this return, we use SDC data and SEC filings to identify offer prices, and CRSP data to identify closing prices on the first day of trading. We also include several control variables in the base case model. Both Ritter and Welch (2002) and Eckbo, Masulis, and Norli (2007) provide summaries of the IPO pricing literature and the variables that are found to affect initial returns. Table 1 show the base case control variables and provides sample statistics for these variables and initial returns, as well as a summary of document words.

As shown in Table 1, initial returns average 40.3% with a standard deviation across IPOs of 66.4%. Consistent with previous IPO samples, the 10th percentile, median, and 90th percentile cutoffs show substantial variation in first-day returns. This variation is partly due to the 1999-2000 bubble period when returns rose substantially. Loughran and Ritter (2004) and Aggarwal, Bhagat, and Rangan (2009) emphasize how sample periods affect IPO pricing. To control for these effects, we introduce both yearly and Fama-French (1993) 48-industry controls in our regressions.

The summary information for the control variables in Table 1 is fairly similar to other IPO samples. We use pre-IPO assets (in log form) to control for size differences, but also estimated our results using data on IPO proceeds (Ibbotson, Sindelar, and Ritter, 1988). The asset measure varied more than proceeds but there is little difference in how these measures affect the regression estimates. We also control for firm age as larger IPOs are likely to be older and less risky than smaller, newer companies.

Reputation of the lead underwriter is measured using market share, defined as the dollar amount underwritten relative to the total IPO capital during the sample period (Megginson and Weiss, 1991). This measure is positively correlated the reputation measure suggested by Carter
and Manaster (1990). The principal idea is that more reputable underwriters may reduce underpricing by certifying the quality of the issue to investors. Thus, reputation is expected to have a negative effect on underpricing, although this effect has been found to reverse itself in IPO samples beginning in the mid-1990s. Similarly, the number of days spent in registration is also an indicator of issue quality. Significant registration delays may imply that the SEC has questions about the filing and requires additional information. Thus, the longer the issuer spends in registration we might expect greater initial underpricing because of the added uncertainty. However, the effect may go in reverse for firms that receive negative information during bookbuilding and revise their offer price downward. These IPOs may have both lower underpricing and a longer time in registration.

Hanley (1993) uses the ratio of offer price to midpoint of the filing range to measure how much firms compensate investors who provide information in the bookbuilding process. This variable is expected to be positively correlated with underpricing as more information, positive or negative, causes a divergence between the offer price and the range midpoint.\(^{19}\)

To control for market sentiment near the IPO offering date, we use the count of IPOs in the prior three months (Booth and Chua, 1996; and Eckbo, et al., 2007). Higher levels of prior market activity are expected to signify positive sentiment by investors. Effectively, the better the prior IPO market, the more capital is likely available to such offerings.

Our remaining control variables measure Nasdaq listings and venture capital backing (Ljungqvist and Wilhelm, 2003; Brav and Gompers, 1997).\(^{20}\) Approximately 79% of these IPOs

---

\(^{19}\) Edelen and Kadlec (2005) suggest that partial adjustment acts to increase the probability of a completed offering. This view suggests that partial adjustment may correlate with contemporaneous market returns.

\(^{20}\) We also considered pre-IPO ownership as a variable to explain management’s acceptance of underpricing. Ljungqvist and Wilhelm (2003) find that underpricing, particularly during the bubble period, is less when management and other related parties hold larger stakes in the company. This variable was insignificant in our base regressions.
are listed on the Nasdaq with 51% backed by venture capital firms. These two variables may act as controls for risk across sectors or the quality of companies, as venture capital firms bring their reputation to the market during an IPO.

We also show the word count information for our sample of IPO prospectuses in Table I. These risk factors sections have on average 3,258 words with a median of 3,128 words. We also provide summary data of the number of enumerated risks in these IPOs. On average there are 27.7 enumerated risks with a median of 27 risks. The similarity of these means and medians suggests symmetry in the distribution of word counts and enumerated risks across IPOs.

IV. ANALYSIS

We first estimate base case regressions as a reference to the explanatory power of the control variables used in our models. These regressions provide residuals for use by the FDR method. We then implement the FDR method to find the set of new words associated with tail returns. These new words are compared to the LM sentiment lists to understand how much they overlap. We then consider the impact of these overlapping sentiment words and test validity using out-of-sample tests. Lastly, we investigate whether parts of speech and subject matter act to connect sentiment and FDR words.

A. Base Case Regressions

The base case regressions are shown in Table 2 for the full sample and for the sample split between the first and second half of the period. The sample split occurs at the end of January 2000 and helps to isolate changes before and after the 1999-2000 technology bubble. This split point is important because Gao, Ritter, and Zhu (2013) find that IPO flow likely decreased after 2000 because of the increasing economies of scope offered in mergers. The split is also used to
validate to FDR method with out-of-sample results. The full sample regressions have 1,391 observations, the first-half has 683 observations, and the second half has 708 observations.

These regressions use Fama and French (1993) 48-industry factors and include separate yearly dummies to help isolate capital market trends.\textsuperscript{21} We cluster the data by month to correct standard errors for correlations between IPOs issued in a similar market environment. All data are winsorized at the 99\% level to reduce the impact of outliers, and standard errors were estimated using monthly clustering to control for correlations between IPO returns. Fama-French 48-industry factor indicators and yearly indicators are used as additional control variables. The adjusted R-squared in these base case models serves as a reference for the explanatory power of the new words found using the FDR method. Overall, the adjusted R-squared and coefficient estimates are consistent with previous studies using IPO data that cover our period (cf., Edelen and Kadlec, 2005).\textsuperscript{22}

B. Finding FDR Words and the Overlap with Sentiment Lists

We implement the FDR method using 5\textsuperscript{th} and 10\textsuperscript{th} percentiles for lower tail returns and 90\textsuperscript{th} and 95\textsuperscript{th} percentiles for upper tail returns. We apply these percentiles to the residual from the base case regression. We use only words that appear in at least 30 documents so as to attain sufficient statistical power in our tests. Out of the 16,352 unique words in the sample documents, 4,508 meet or exceed this filter. The FDR control rate is set at 5\% for all tests, so we expect no more than a 5\% error rate among the FDR significant words that we identify.

\textsuperscript{21} As a robustness check we also estimated first-day return regressions with controls for the 1999-2000 period and whether the IPO was a technology company. We excluded the yearly dummies and Fama-French 48-industry factors from these regressions. The full sample results were slightly worse in terms of R-squared.

\textsuperscript{22} The control variable coefficients support the previous literature, except for the size and venture capital measures. The venture capital backing effect has switched signs from the effect identified by Megginson and Weiss (1991) and the asset size measure is insignificant. A Nasdaq OMX listing tends to increase underpricing, the offer-to-midpoint of the filing range variable is significant and positive, the number of IPOs issued in the past three months has a negative effect on underpricing, and the age of the firm also has a negative effect.
Table 3 shows two panels of results. In Panel A, we report word counts of FDR significant words for each percentile test using the residual from the base case regression. We identify 66 (392) words significant in the lower 5th (10th) percentile tests or approximately 1.5% (8.7%) of the sample words, and 369 (675) words significant in the upper 95th (90th) percentile tests or approximately 8.2% (15.0%) of the sample words.

Recall that our primary objective is to compare these optimally chosen words to those on sentiment lists. These comparisons are also shown in Panel A. About 21% of both the lower and upper tail words are matched to a word on a LM tonal list. Of the overlaps, the vast majority (approximately 70%) are on the LM negative tone list with positive and uncertainty lists together accounting for between 21% (lower tail) and 27% (upper tail), respectively. The occurrence of same-tone words in both the upper and lower tail FDR word lists suggests context, not just tone, is important to investors to sort out the implications of these words. Without context such words send ambiguous signals and are likely to increase uncertainty and underpricing.

Panel B in Table 3 shows the top 20 words identified as FDR significant for each lower and upper tail test and indicates if any of these words overlap with a LM tonal list. To determine the top 20, we ranked by statistical significance. This panel shows a strong relationship between FDR significant words in same-tail tests: Eleven out of the top 20 words overlap in both the lower tail and upper tail tests. The total overlap for these results is that 98.5% of all 5th percentile words overlap the 10th percentile results and 90.8% of all the 95th percentile words overlap the 90th percentile results. This finding allows us to focus on the 10th and 90th percentile results in subsequent analyses.

A large number of the words associated with high underpricing have a technology theme. Words such as “versions”, “networking”, “redesign”, “user”, “engineering”, “hardware” and
“enhancements” bring to mind technology companies or products. This is understandable because these data include the internet bubble period so many companies going public had internet-related technologies or products.

The lower tail word list is less thematic. There are technology related words, such as “web”, “traffic”, and “viruses”. There are also two notable legal words: “libel” and “copyright”. These words certainly may bring up questions for investors, such as is the company being sued for libel, does it have copyright protection for it work, or does it need to obtain permission to use a copyright that maybe expensive? The lower tail list also includes some seemingly benign words: “privacy”, “content”, “commerce”, “assessed”, “adapt” and “medium”. The predictive power of these words is significant, yet their seeming neutrality again emphasizes that context is necessary to sort out the reasons why they are associated with such low returns.

C. Impacts of Tonal and FDR Words

To measure the predictive content of the document words, we follow sentiment literature and define a document-specific frequency variable:

$$f_d^t = \frac{\sum_{j=1}^{J} w_j^d \Psi_t(j)}{\sum_{j=1}^{J} w_j^d}, \quad (6)$$

where $J$ denotes the number of unique words in our documents and $\Psi_t(j)$ is the assignment function; it equals one if word $j$ appears on the sentiment list and zero otherwise.

Table 4 shows the predictive content of LM tonal words using the measure in (6) for each tonal list. We follow the same methods used to estimate the regressions in Table 2, where the first-day return regression produces an adjusted R-squared of 27.9% using only base case control variables. This value serves as a reference for the explanatory power of the LM tonal word lists.
Note that the LM strong and weak modal lists are omitted because they matched only eight and 19 words in our sample, respectively.

The regressions in Table 4 show that negative words have a significant positive effect on predicted first-day returns, which is consistent with existing literature. Uncertainty words also have a positive effect, but are only weakly significant. We also combine words on the negative, uncertain, and weak modal lists following the methods of Loughran and McDonald (2013). This variable also shows a significant positive effect on predicted returns, but the explanatory power is not meaningfully greater (t-statistic of 3.81 versus 3.79) than when negative words alone are included separately. These results suggest that the fraction of negative words is the principal reason for the correlation between underpricing and LM tonal measures. The usual interpretation is that the positive sign of this variable indicates that the tone conveyed by negative words may warn investors of potentially higher risks or greater uncertainty, which is subsequently associated with greater underpricing. Even though some LM variables are significant, the increase in adjusted R-squared from the base case regression is minimal—from 0.0 to 0.8 percentage points depending on the model—which is at most a 2.9% increase.

The FDR-derived words we have found may be considered a set of words important for framing investor perceptions of IPO risk and uncertainty. Approximately 21% of these FDR-derived words overlap words on the LM tonal lists. To test whether our view is supported we re-estimated the regressions in Table 4 using only the FDR-derived words that overlap words on each of the LM tonal lists, but included measures for both upper tail and lower tail overlapping words. Thus, we have FDR-interaction measures for each type of tone. We also re-estimated the these results in the form shown in Table 4 using only the LM words that did not overlap with a word on the FDR-derived lists. For these overlaps, we used the results from the lower (10%) and
upper (90%) tail tests. Our findings are summarized in Table 5 by showing how a one-standard deviation change in the measured variable impacts predicted first-day returns.

The first column of results in Table 5 shows the impact of the LM tonal measures using the regression coefficients estimated in Table 4, where there is no adjustment to the words included. Statistically significant impacts based on underlying p-values are marked with an asterisk. For example, a one standard deviation change (1.4%) in the LM negative tone measure implies a statistically significant impact of an 8.3% increase in first-day returns. Only the LM negative tone and the combined negative/uncertainty/weak modal tones are found to have significant impacts at the 95% level or above.

The estimated impacts are much larger for LM words that overlap either the FDR lower or upper tail word lists. The middle two columns of results in Table 5 show the impact of LM words that are also found on the upper and lower lists, respectively. Here the magnitudes of these impacts are demonstrably greater when compared to the impacts of LM sentiment words in the first column. It is also clear that selected subsets of LM words have significant and opposing effects, which may explain why the coefficient for the positive tone list in Table 4 is not significant. Even if we were to net these effects against each other because there are some overlapping words on the FDR lists, the net effect would be 1.5 to 3.0 (1.8 to 4.9) times greater for the negative (positive) words subset compared to the impact of the full-list LM sentiment words.

The final column of Table 5 shows the impact of the LM words that do not overlap with words on the FDR-derived lists. Two results are seen from these impacts. First, the magnitudes of these impacts are small compared to any of the other columns. Given that nearly 80% of the

---

23 Researchers have been puzzled about the insignificance of the positive tonal list. Jegadeesh and Wu (2013) discuss this puzzle and offer an alternative approach to classifying words as positive or negative tone. Their approach shows that a positive document tone affects market returns.
LM words are not FDR words, there are large numbers of tonal words that produce only weak sentiment signals. Second, the subset of negative tone words isolated here show a significant negative impact on first-day returns. Previous work with the negative tone word list and IPO returns find a positive correlation. This result shows that that correlation is driven by only a small fraction of the words on the LM negative tone list. In effect, the majority of those words provide a weak signal that lowers first-day returns in IPOs. This finding is more intuitively appealing as negative words should convey negative tone and thereby warn after-market investors about the IPO risks.

Taken together, the results in Table 5 suggest that FDR-derived words may be an effective filter to isolate the set of important sentiment words in a document, at least into the negative, positive, uncertain, and litigious groups studied here. Those FDR words that overlap these lists appear to be the best at capturing investor sentiment. The words that do not overlap FDR-derived words have small effect sizes and, in the case of negative sentiment words, have the opposite sign. This finding suggests that the negative tone list contains words with an ambiguous meaning, so that textual information from a document is not well defined by a sentiment list.

A word of caution is necessary. Although words on sentiment lists are exogenous to returns, the subsets that overlap our FDR-derived words are selected based on their correlation with returns. We therefore expect the overlapping words to be statistically significant in Table 5. However, these regressions are useful for quantifying the predictive content of the FDR words. We selected the words based on whether they have a statistically significant correlation with returns and the regression reveals the magnitude of those correlations.
D. Out-of Sample Validation

Table 6 includes the variables $f^T_d$ from equation (6), defined using FDR-based words, in the base case regressions and in the split sample. For the split sample, we estimate the FDR words from the first half of the sample and use those words to create $f^T_d$ measures for the second half of the sample. We then reverse this procedure to create frequency measures for the first half of the sample.

In the first half of the sample, the FDR method identifies 325 (77) words associated with the upper 90% (95%) tail and 89 (1) words associated with the lower 10% (5%) tail, respectively. In the second half of the sample, the FDR method identifies 84 (54) words associated with the upper 90% (95%) tail and 41 (4) words associated with the lower 10% (5%) tail, respectively. Because there are so few FDR words associated with the lower 5% tails, we exclude frequency measures for those words in our regressions.

Table 6 shows both the in-sample and out-of-sample impact of the FDR word frequency measures. All of the control variables in the base case regressions are included in these models. The models combine FDR significant words found using the 10th/90th percentile and the 5th/95th percentile tests. As expected, the in-sample upper and lower measures are highly significant for both sets of percentiles. The out-of-sample data also show highly significant results when the second half FDR words are used to predict initial returns in the first half of the data. The first half FDR words are also predictive in the second half of the data, but these results are not as strong being significant at the 10% level. In total, these findings show that the FDR-based methods are robust in out-of-sample testing.

At a higher level, the changes in adjusted R-squared in the full sample in Table 6 compared to the results in Table 4 are a measure of how well our method does in identifying important
words. Because the set of words is broader than that suggested by sentiment lists, it appears that IPO prospectuses contain substantial additional information on pricing beyond that captured by tonal measures. This information would include words that do not fit easily into tonal categories, such as “versions”, “engineering”, “copyright”, “hardware”, “networking” or “redesign”.

E. Parts of Speech

The fact that most FDR-derived words do not fit well into tonal categories leads us to ask whether there is a more fundamental approach to understanding investor sentiment and IPO uncertainty. We consider whether the parts of speech offer insight into sentiment effects as both nouns and verbs are essential to clear, understandable writing. Also, adjectives and adverbs add clarity and emphasis to a sentence. It is thus possible that these parts of speech may be the building blocks of an investor’s perception of an IPO. If so, they may help explain why the words on sentiment lists have lower explanatory power than those found using the FDR-based method. To test this conjecture, we use the online Oxford English dictionary to identify how FDR and LM sentiment words are used in our IPO sample. We then examine whether these different parts of speech affect initial IPO returns.

Table 7 shows the distribution of nouns, adjectives, verbs, and adverbs on the four LM tonal lists, the combined negative/uncertain/weak modal lists, and for the FDR-based words found using the lower and upper tails with the 10th and 90th percentile cutoffs. For the tonal lists, the negative, positive, and uncertain tone words found in our sample have significantly fewer nouns and more adjectives than the FDR lists. This also arises for the combination of negative, uncertain, and weak modal lists that Loughran and McDonald (2013) use as a textual measure for ex ante uncertainty. The top 20 list in Panel B of Table 3 supports this finding with over one-half of these words serving as nouns in the document sample. Combining verbs and adverbs as they
complement each other shows that these parts of speech are over-represented on the sentiment lists compared to the FDR-based word results. The litigious list is the exception with a distribution almost similar to the FDR upper tail list of words.

What do these findings mean? A sentence must have both a noun and a verb. An adjective adds clarity to a noun and an adverb adds emphasis to a verb. These are the most common parts of speech that offer meaning, so it sensible if the reader of an IPO prospectus forms a perception using these basic elements. In effect, sentiment or an investor’s view of uncertainty is principally built from the reading of nouns, verbs, adjectives, and adverbs. Table 7 shows that nouns are relatively more important than any other part of speech for both highly over-priced and highly underpriced IPOs. The implication is that sentiment lists would be more accurate measures of sentiment if they included relatively more nouns and possibly fewer verbs and adjectives. Specifically, from the 90th percentile list in Panel B of Table 3, nouns such as “deployment”, “versions”, “engineering”, “hardware”, “networking”, and “enhancements” serve to identify highly underpriced IPOs and hence IPOs that are likely to have greater ex ante uncertainty.24

To better understand the value of parts of speech in predicting initial returns, we decompose the LM lists into nouns, adjectives, verbs, and adverbs. We then estimate separate effects for each part of speech based on the relative frequency of such words in the risk factor section of our IPO sample. Note that some words—such as “engineering” or “redesign”—may be used in some sentences as one part of speech and in other sentences as another part of speech. We have credited such words as belonging to multiple parts of speech when we found these different usages in the sample documents.

Table 8 shows regressions from Table 4 re-estimated using parts-of-speech variables entered as separate regressors. The fraction of nouns, verbs, adjectives, and adverbs by tonal category are

24 Note that “engineering” also serves as an adjective in the sample documents.
measured as in equation (6) for each document. The table shows the estimated coefficient, p-value in parentheses beneath the coefficient, and the R-squared for each separate regression beneath the p-value. The base case control variables described in Table 1 are also included in these regressions.

The results in Table 8 support the findings in Table 7, particularly for nouns. The nouns included on LM tonal lists are not statistically significant, except possibly for the litigious list, which has a negative coefficient. Compared to the representation of nouns in FDR-based words, these findings suggest that the tone or sentiment captured by meaningful nouns is substantially missing from the LM negative and uncertain tonal lists. In contrast, the LM lists appear to best capture verbs and adjectives that are meaningful sentiment words. Adverbs only appear significant on the positive tone list. We also estimated a combined verb/adverb model for the LM word lists. These results were consistent in significance to the verb-alone results, so the adverb part of speech may be a less meaningful building block for investor sentiment.

Over all of the LM-based results in Table 8, the highest R-squared is 28.7%, while the full sample results in Table 6 show the lowest R-squared for FDR-based models is 32.2%. The difference in explanatory power suggests that tonal lists may be improved with additional words from all parts of speech or from the development of separate lists, such as a list of nouns that describe technical things or functions, such as “bandwidth”, “networking”, “redesign”, or “hardware”.

\[\text{25} We\ also\ estimated\ regressions\ that\ included\ all\ parts\ of\ speech\ measures\ for\ both\ LM\ lists\ and\ FDR-based\ words.\ These\ results\ confirm\ the\ claims\ made\ for\ the\ separate\ regressions.\]
6. CONCLUSIONS

Previous research has improved our understanding of how investor sentiment, measured by document tone, impacts asset returns and thereby offer a new avenue for testing pricing theories. This research analyzed the carefully crafted tonal word lists of Loughran and McDonald (2011) to determine whether those words broadly captured the collection of words associated with IPO underpricing. Our analysis developed a new method to identify words associated with high (or low) underpricing in IPOs. We show that there is hundreds of non-sentiment words associated with underpricing. These new words offer significant explanatory power in out-of-sample tests and they may help to filter out the really important tonal words on the LM lists.

We also investigated how parts of speech—nouns, verbs, adjectives, and adverbs—affect first-day IPO returns with the hypothesis being that parts of speech are significant determinants of sentence clarity and meaning, so that they may also be basic building blocks of investor perceptions of uncertainty. Decomposing the FDR-based words into parts of speech shows that nouns are over-represented relative to their use in sentiment word lists. Regression results for LM tonal words suggest that more nouns—either tonal based, industry connected, or general “things”—would be helpful to characterizing investor perceptions of IPO risks. Overall, our new method shows that hundreds of words, particularly nouns, not on sentiment lists are important to IPO pricing.
REFERENCES


Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, working paper, Kellogg School of Management, Northwestern University.


Feuerriegel, Stefan, Schmitz, Jonas Tobias, and Dirk Neumann, 2014, What matters most? How tone in initial public offering filings and pre-IPO news influences stock market returns, working paper, University of Freiburg, Freiburg, Germany.


____, 2013, IPO first-day returns, offer price revisions, volatility, and form S-1 language, forthcoming in the *Journal of Financial Economics*.


_____ , 1949, Human behaviour and the principle of least effort, Addison-Wesley, Cambridge, MA.
### Table 1

**First-day Returns, IPO Control Variables, and Document Words**

This table provides summary data on IPO returns, control variables, and document words. The first-day return is calculated from the offer price to the close of trading on the first day. The base case control variables are the age of the firm as of the filing date, number of days in registration, number of completed IPOs within three months of the offer date, the ratio of offer to the midpoint of the initial filing range, the market share of the lead underwriter during the sample period, pre-IPO assets, and dummies for IPOs listed on the Nasdaq market and whether the offering was backed by venture capital firms. Document words are for word counts in the risk factor section of the prospectus. The number of risk factors is based on counts of each enumerated risk. The information on means, standard deviations, percentile cutoffs, and medians are summarized for our sample of 1,391 IPOs issued between 1998 and 2005. These data are collected from the Securities Data Corporation database and filings with the Securities and Exchange Commission.

<table>
<thead>
<tr>
<th>Sample Characteristics</th>
<th>Average</th>
<th>Std. Dev.</th>
<th>10th Percentile</th>
<th>Median</th>
<th>90th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-day Return</td>
<td>40.34%</td>
<td>66.39%</td>
<td>-3.57%</td>
<td>15.41%</td>
<td>124.37%</td>
</tr>
<tr>
<td><strong>Base Case Control Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Firm at IPO filing (years)</td>
<td>14.1</td>
<td>20.1</td>
<td>2.0</td>
<td>7.0</td>
<td>34.0</td>
</tr>
<tr>
<td>Number of Days in Registration</td>
<td>107.9</td>
<td>70.6</td>
<td>59.0</td>
<td>83.0</td>
<td>59.0</td>
</tr>
<tr>
<td>Number of IPOs in prior 3 months</td>
<td>78.2</td>
<td>42.3</td>
<td>20.0</td>
<td>84.0</td>
<td>131.0</td>
</tr>
<tr>
<td>Offer to midpoint of initial filing range</td>
<td>1.06</td>
<td>0.28</td>
<td>0.73</td>
<td>1.02</td>
<td>1.40</td>
</tr>
<tr>
<td>Market Share of Lead Underwriter</td>
<td>4.06%</td>
<td>2.68%</td>
<td>0.01%</td>
<td>4.38%</td>
<td>7.10%</td>
</tr>
<tr>
<td>Pre-IPO Assets ($ millions)</td>
<td>1,443.5</td>
<td>14,373.0</td>
<td>10.0</td>
<td>55.3</td>
<td>708.0</td>
</tr>
<tr>
<td>NASDAQ listings</td>
<td>78.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venture capital backing indicator</td>
<td>51.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document Words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Factor Section</td>
<td>3,258</td>
<td>1,271</td>
<td>1,847</td>
<td>3,128</td>
<td>4,776</td>
</tr>
<tr>
<td>Number of Risk Factors</td>
<td>27.7</td>
<td>8.1</td>
<td>18.0</td>
<td>27.0</td>
<td>38.0</td>
</tr>
</tbody>
</table>
The effects of commonly used IPO control variables are shown using the full sample and split samples to compare changes in these effects before and after the 1999-2000 technology bubble. The sample split occurs at the end of January 2000. The control variables used are described in Table 1. All data are winsorized at the 99% level to reduce the impact of outliers with standard errors estimated using monthly clustering to control for correlations between IPO returns. Fama-French (1993) 48-industry factor indicators and yearly indicators are used as additional control variables. The full sample regressions have 1,391 observations, the first-half has 683 observations with the split at end of January, 2000, and the second half has 708 observations. Levels of significance are shown by p-values in parentheses beneath each coefficient.

### Table 2
**IPO First-Day Returns**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Full Sample</th>
<th>First Half of Sample</th>
<th>Second Half of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-IPO Assets ($ millions)</td>
<td>0.002</td>
<td>-0.014</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.877)</td>
<td>(0.507)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>NASDAQ listings</td>
<td>0.045</td>
<td>0.111</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.074)</td>
<td>(0.326)</td>
</tr>
<tr>
<td>Venture Capital backing</td>
<td>0.167</td>
<td>0.158</td>
<td>0.159</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Market Share of Lead Underwriter</td>
<td>4.184</td>
<td>6.506</td>
<td>1.140</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Number of Days in Registration</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.003)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>Offer to midpoint of initial filing range</td>
<td>0.218</td>
<td>0.081</td>
<td>0.372</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.069)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Number of IPOs in prior 3 months</td>
<td>-0.002</td>
<td>-0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.004)</td>
<td>(0.805)</td>
</tr>
<tr>
<td>Age of Firm at IPO filing (years)</td>
<td>-0.056</td>
<td>-0.104</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.84)</td>
</tr>
<tr>
<td>Constant term</td>
<td>-0.030</td>
<td>-0.158</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>(0.816)</td>
<td>(0.544)</td>
<td>(0.796)</td>
</tr>
<tr>
<td>Yearly Indicators</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>FF48 Industry Indicators</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

*Adjusted R²* 27.9% 31.1% 27.8%
Table 3  
Words Associated with First-day IPO Tail Returns

In a sample of 1,391 IPOs, there are 4,508 words that are repeated in the risk factor section of 30 or more IPO prospectuses. This sample of words is analyzed using FDR methods with a 5% control rate to identify which words are statistically associated with the lower and upper tails of regression-adjusted returns using the residual from the Fama-French 48-industry adjusted regression shown in Table 2. Panel A shows the total count of FDR significant words for two lower and upper tail percentiles. The 5th (10th) percentile identifies words associated with IPOs whose sample performance places them in the lower 5th (10th) percentile of first-day returns. The 90th (95th) percentile finds words associated with IPOs whose sample performance places them in the upper 90th (95th) percentile of first-day returns. These results are compared to tonal word lists created by Loughran and McDonald (2011, LM) to count the overlapping words. Panel B shows the top 20 words identified as FDR significant for each lower and upper tail test. In this panel, the abbreviation "INS" refers to the Immigration and Naturalization Service. Words that overlap those on a tonal list are marked (n) negative list, (p) positive list, (u) uncertain list, or (l) litigious list.

<table>
<thead>
<tr>
<th>Residual in FF 48-Industry Regression</th>
<th>Lower Tail</th>
<th>Upper Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5th p-tile</td>
<td>10th p-tile</td>
</tr>
<tr>
<td>FDR Significant Count</td>
<td>66</td>
<td>392</td>
</tr>
<tr>
<td>Words on LM List:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>10</td>
<td>57</td>
</tr>
<tr>
<td>Positive</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Litigious</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Words not on LM Lists</td>
<td>51</td>
<td>312</td>
</tr>
</tbody>
</table>

Panel A: Count and Tonal Distribution of FDR significant words

#1 libel (l) libel (l) deployed deployed
#2 viruses viruses slowly (n) deployment
#3 outages (n) web deployment engineering
#4 web slowly (n) widespread older
#5 defamation (n) break (n) versions ordering
#6 unsettled (u) INS engineering copyright
#7 privacy outages (n) copyright widespread
#8 shortfall (n) commerce hardware unproven (u)
#9 content adapt enhancements (p) redesign
#10 connected unsettled (u) networking versions
#11 break (n) stored internationally ordering
#12 accessed medium evolving emerge
#13 INS traffic reluctant (n) problem (n)
#14 copyrights defamation (n) commerce interoperate
#15 user online copy scalability
#16 communication unproven (u) redesign enhancements (p)
#17 commerce transmission user warranty
#18 internet privacy undetected (n) undetected (n)
#19 prevalent print unproven (u) slowly (n)
#20 gauge user solutions deploy
Table 4
Predictive Power of LM Tonal Words for First-Day Returns

The effects of each LM tonal list on first-day returns are shown using the full sample and including base case control variables. For each IPO, the LM variable is the fraction of tonal words in the risk factors section divided by the total number of words in that section. In addition, the the effect of combining the negative, uncertainty, and weak modal lists into one variable is shown. The LM strong and weak modal lists are omitted because they matched only eight and 19 words in our sample, respectively. The base case control variables used are described in Table 1. All data are winsorized at the 99% level to reduce the impact of outliers with standard errors estimated using monthly clustering to control for correlations between IPO returns. Fama-French (1993) 48-industry factor indicators and yearly indicators are used as additional control variables. Each regression has 1,391 observations. Levels of significance are shown by p-values in parentheses beneath each coefficient.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Model Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>LM Tonal Word List:</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>5.85</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>Positive</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
</tr>
<tr>
<td>Uncertainty</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Litigious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative/Uncertainty/Weak Modal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Base Case Control Variables:</td>
<td>yes</td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
</tr>
<tr>
<td>Yearly indicators</td>
<td></td>
</tr>
<tr>
<td>FF48 Industry indicators</td>
<td>yes</td>
</tr>
</tbody>
</table>

Adjusted $R^2$ (Base Case = 27.9%) 28.6% 27.9% 28.0% 27.9% 28.7%
Table 5
Impact of LM Tonal Words on Predicted First-Day Returns

The expected effects of a one standard deviation change on first-day returns are shown for the LM tonal variables using the coefficient estimates in Table 5, the interaction between LM tonal list words and FDR words, and the impact of LM tonal words that do not intersect FDR-derived words. The latter impacts are from regressions with separate measures for the overlap with lower and upper tail FDR words. The "interaction" regressions select only the LM tonal words that are included on FDR-derived word lists found using the lower (10%) and upper (90%) tail tests. The last column shows the impact of LM tonal list words for regressions of the form in Table 4 in which none of the LM words overlap the lower (10%) and upper (90%) FDR-derived word lists. All impacts are measured in percent. An asterisk indicates that the coefficient in the underlying regression model is significant at the 95% level or above.

<table>
<thead>
<tr>
<th>Tonal List Variable</th>
<th>Impact of the LM Words</th>
<th>Impact of LM Words Overlapping FDR Upper (90%) Tail Words</th>
<th>Impact of LM Words Overlapping FDR Lower (10%) Tail Words</th>
<th>Impact of LM Words not on either FDR list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>8.3%*</td>
<td>42.8%*</td>
<td>-18.0%*</td>
<td>-3.9%*</td>
</tr>
<tr>
<td>Positive</td>
<td>2.3%</td>
<td>18.4%*</td>
<td>-7.1%*</td>
<td>-1.5%</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>3.5%</td>
<td>17.0%*</td>
<td>-5.0%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Litigious</td>
<td>-0.6%</td>
<td>11.8%*</td>
<td>-2.7%</td>
<td>-2.7%</td>
</tr>
<tr>
<td>Negative/Uncertainty/Weak Modal</td>
<td>8.8%*</td>
<td>45.8%*</td>
<td>-25.1%</td>
<td>-2.9%*</td>
</tr>
</tbody>
</table>
Table 6
Split-Sample Validation of FDR Words

The full sample base case regression in Table 2 is re-estimated using the percentage of FDR significant words in the risk factor section of each IPO to create new control variables. Results are shown by pairing the upper and lower FDR significant words in the 10th and 90th percentile tests and the 5th and 95th percentile tests. Split sample results are estimated after identifying the FDR significant words using only the split samples. The FDR significant words in the 1st half of the sample are then used in regressions on the second half of the sample and vice versa. The control variables described in Table 1 are also included in these regressions. Regressions are estimated using robust methods for the FDR significant words found using the conditioning or residual method described in the text. All data are winsorized at the 99% level to reduce the impact of outliers. Sample sizes are shown at the bottom of the table. Levels of significance are shown by p-values in parentheses beneath each coefficient. An "n.a." implies that there were too few observations on the FDR words for this tail to include this measure in the regressions.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>10th and 90th Percentile Tests</th>
<th>5th and 95th Percentile Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FDR Words in 2nd half on 1st half data</td>
<td>FDR Words in 1st half on 2nd half data</td>
</tr>
<tr>
<td></td>
<td>Full Sample</td>
<td>on 1st half data</td>
</tr>
<tr>
<td>FDR % Upper tail words</td>
<td>9.07</td>
<td>9.79</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>FDR % Lower tail words</td>
<td>-6.22</td>
<td>-20.84</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Variables from base case</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Yearly indicators</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>FF48 Industry indicators</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>36.7%</td>
<td>34.6%</td>
</tr>
<tr>
<td>Sample Size</td>
<td>1391</td>
<td>683</td>
</tr>
</tbody>
</table>
Each word found using the FDR-based method is classified as a part of speech from the online Oxford English dictionary. Words whose usage changes with sentence context are assigned to multiple parts of speech if that usage is found in the sample documents. Words on the Loughran and McDonald (2011, LM) tonal lists are similarly classified if they appear in the sample. The LM strong and weak modal lists are omitted because they matched only eight and 19 words in our sample, respectively. However, the combination of negative, uncertain and weak modal is shown for comparison. The count of upper and lower tail FDR and LM words found in the sample is shown along with the fraction of nouns, adjectives, verbs, and adverbs. The p-values for the chi-squared goodness-of-fit test comparing the parts of speech distribution of sentiment words to the corresponding distribution for FDR significant words is shown in the final two columns.

<table>
<thead>
<tr>
<th></th>
<th>Sample Word Count</th>
<th>Percentages</th>
<th>p-values</th>
<th>Goodness-of-fit test v. Upper</th>
<th>Goodness-of-fit test v. Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR Words in Upper (90%) Tail</td>
<td>675</td>
<td>45.78</td>
<td>18.88</td>
<td>30.62</td>
<td>4.72</td>
</tr>
<tr>
<td>FDR Words in Lower (10%) Tail</td>
<td>392</td>
<td>43.69</td>
<td>18.06</td>
<td>35.15</td>
<td>3.11</td>
</tr>
<tr>
<td><strong>LM Tonal List</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>1533</td>
<td>34.49</td>
<td>21.03</td>
<td>40.19</td>
<td>4.29</td>
</tr>
<tr>
<td>Positive</td>
<td>248</td>
<td>28.71</td>
<td>30.28</td>
<td>33.44</td>
<td>7.57</td>
</tr>
<tr>
<td>Uncertain</td>
<td>220</td>
<td>30.26</td>
<td>23.99</td>
<td>33.58</td>
<td>12.18</td>
</tr>
<tr>
<td>Litigious</td>
<td>374</td>
<td>48.67</td>
<td>15.27</td>
<td>29.42</td>
<td>6.64</td>
</tr>
<tr>
<td>Negative/Uncertain/Weak Modal</td>
<td>1771</td>
<td>33.47</td>
<td>21.41</td>
<td>39.32</td>
<td>5.80</td>
</tr>
</tbody>
</table>
The regressions in Table 4 are re-estimated using parts-of-speech variables entered as separate regressors in individual models. The fraction of nouns, verbs, adjectives, and adverbs by tonal category are measured for each document to compute these regressors. The table shows the estimated coefficient, p-value in parentheses beneath the coefficient, and the R-squared for each separate regression beneath the p-value. All of the base case control variables described in Table 1 are included in these regressions. The base case regression R-squared is 27.9%. These data are winsorized at the 99% level to reduce the impact of outliers. All regressions have 1,391 observations.

### Table 8
Effects of Parts of Speech on First-day Returns

The table shows the estimated coefficient, p-value in parentheses beneath the coefficient, and the R-squared for each separate regression beneath the p-value. All of the base case control variables described in Table 1 are included in these regressions. The base case regression R-squared is 27.9%. These data are winsorized at the 99% level to reduce the impact of outliers. All regressions have 1,391 observations.

<table>
<thead>
<tr>
<th>Part of Speech Variable: Fraction found in IPO document</th>
<th>Negative</th>
<th>Positive</th>
<th>Uncertain</th>
<th>Litigious</th>
<th>Negative/ Uncertain/ Weak Modal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>-3.65</td>
<td>-7.85</td>
<td>11.35</td>
<td>-10.27</td>
<td>-2.15</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.105)</td>
<td>(0.340)</td>
<td>(0.051)</td>
<td>(0.384)</td>
</tr>
<tr>
<td></td>
<td>27.9%</td>
<td>28.0%</td>
<td>27.9%</td>
<td>27.9%</td>
<td>27.9%</td>
</tr>
<tr>
<td>Verbs</td>
<td>12.69</td>
<td>-9.91</td>
<td>22.57</td>
<td>4.65</td>
<td>12.16</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.248)</td>
<td>(0.008)</td>
<td>(0.615)</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td>28.3%</td>
<td>27.9%</td>
<td>28.4%</td>
<td>27.9%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Adjectives</td>
<td>27.49</td>
<td>0.66</td>
<td>0.69</td>
<td>-12.42</td>
<td>23.16</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.916)</td>
<td>(0.673)</td>
<td>(0.669)</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td>28.7%</td>
<td>27.9%</td>
<td>27.9%</td>
<td>27.9%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Adverbs</td>
<td>9.18</td>
<td>62.95</td>
<td>-10.57</td>
<td>40.42</td>
<td>6.44</td>
</tr>
<tr>
<td></td>
<td>(0.277)</td>
<td>(0.000)</td>
<td>(0.592)</td>
<td>(0.315)</td>
<td>(0.379)</td>
</tr>
<tr>
<td></td>
<td>27.9%</td>
<td>28.7%</td>
<td>27.9%</td>
<td>27.9%</td>
<td>27.9%</td>
</tr>
</tbody>
</table>

**Base Case Control Variables:**

- **Independent Variables:** Yes, Yes, Yes, Yes, Yes
- **Yearly indicators:** Yes, Yes, Yes, Yes, Yes
- **FF48 Industry indicators:** Yes, Yes, Yes, Yes, Yes
This appendix addresses two potential issues that could affect the robustness of our results: censoring bias and outlier effects.

A. Censoring Bias

Zipf’s law argues that the frequency of use of a word is inversely proportional to its rank in a document (Zipf, 1935, 1949). If Zipf’s law applies across rather than within documents, then we would confront a censoring problem. For example, if we have 1,000 documents and if the most frequently used word is found in all documents, Zipf’s law implies that the second most frequent word is found in one-half of the documents, the third most frequent in one-third of the documents, etc. At the 28th most frequent word we would expect it to repeat in only about 36 documents, so we would have relatively few words to choose among for these tests if we used a document repeat cutoff of \( L = 30 \) or more to ensure statistical power. Fortunately, regression estimates reject Zipf’s law across our IPO document sample.\(^{26}\)

In our sample there are 16,352 unique words. Approximately, 28% of these meet a repeat cutoff threshold of 30 or more documents.\(^{27}\) The remaining 72% is 11,844 words, so we are concerned about the effects of a censored sample on our conclusions. We cannot include these words without sacrificing statistical power, but we can examine how such censoring is likely to

\(^{26}\) Ordinary least squares estimates of the regression model, \( \ln(\text{Word Rank}) = a + b \ln(\text{Count of Documents}) + \text{error term} \), produces \( \hat{a} = 10.9 \) and \( \hat{b} = -0.625 \) with an adjusted R-squared of 0.086. The estimate of \( b \) is significantly different from -1.0, which is a rejection of Zipf’s law. Also, the adjusted R-squared suggests that this is weak fit for the Power law to be robust in our sample data.

\(^{27}\) Heller and Gur (2012) show that the FDR procedure yields more conservative results using discrete data; this allows us some flexibility in selecting a repeat cutoff value for \( L \).
bias the success ratios in eqs. (2) and (3), and whether censoring is likely to affect the total count of words included in the inferior or superior performance lists.

A censoring bias would imply that the success ratios were different in the censored and uncensored samples. We test this implication. Specifically, we are censoring on a variable \((K_j^+)\), which as defined above represents the number of repeats across the IPO sample for word \(j\). Thus, we test whether success ratios for data in which \(K_j^+ < L\) are different from success ratios for data in which \(K_j^+ \geq L\). This test provides information on the impact of words by frequency-of-use groups.

In our analysis, the cutoff is \(L = 30\) documents. The omitted group contains 11,844 words compared to 4,508 in the included group. We compute success ratios for each word in the included and omitted groups for each of the four percentiles used in our tests. We then test whether the average success ratio is different between the included and omitted groups. The results show that for the 90th and 95th percentile tests, the average success ratio is higher for the included group of words, while it is not statistically different for the 5th and 10th percentile tests.\(^{28}\) These results suggest that the censoring bias is irrelevant in the lower tail tests, but is the opposite of what previous literature suggests for the upper tail tests. In effect, censoring on words that repeat in 30 or more IPO documents creates a sample that is more likely to contain high impact, FDR significant words at least for the upper tail tests.

An alternative test is to simulate whether less frequently repeated words would add to the total count of words on our lists if they were pooled with the more frequently used words. To this end, we developed such a test using a bootstrap simulation approach. We present the

\(^{28}\) The average success rate in the omitted group for the 5th (10th) percentile test is 0.049 (0.111) and 0.051 (0.104) for the included group with a p-value for the difference equal to 0.42 (0.08). For the upper tail, the average success rate in the omitted group for the 90th (95th) percentile test is 0.064 (0.031) and 0.095 (0.047) for the included group with a p-value for the difference equal to 0.000 (0.000).
analytics of this method and the simulation results in a Supplemental Appendix.\textsuperscript{29} We find no evidence to support a claim that censoring on repeats affects the total count of FDR significant words, or that lower use words have higher impact as measured by success rates.

\textit{B. Sensitivity to Return Outliers}

Our FDR method depends on IPO documents from the tail areas of the distribution of first-day returns. As such, our results may be partly driven by a connection to tail area returns. To examine this issue, we re-estimated both the results in Table 5 and Table 6 using only IPOs from the middle 80\% of first-day returns. There were 1,113 observations in these regressions. For Table 5, we found the significance levels to decrease with the LM uncertainty tone measure now insignificant even at the 10\% level. For Table 6, the impacts using overlapping FDR and LM tonal words were largely similar for the upper tail measures except for the litigious tone which was no longer significant; the other upper tail interactions remained significant at the 95\% level or above. The LM/FDR lower tail impact measures all became insignificant.

While we are not surprised that removing 20\% of the tail data changes results, particularly levels of significance, we are somewhat surprised that it only meaningfully affects the lower tail overlapping measures in Table 6. The issue this presents is whether the FDR lower tail words are less effective or whether the LM words that overlap these FDR lower tail words are the problem. To address this issue, we included a new variable that measured the ratio of FDR words that are non-LM tonal words to total words in an IPO. This ratio is computed for FDR upper tail and lower tail words separately. We then re-estimated the group regressions using the middle 80\% dataset, which included LM/FDR overlapping measures and these new FDR/non-LM variables. In each model, both the upper and lower tail FDR/non-LM coefficients were highly significant.

\textsuperscript{29} The Supplemental Appendix provides additional tables to examine the robustness of the results in Tables 5 and 6; it is available from the authors via email.
with the upper tail sign being positive and the lower tail sign being negative. In these models, the
LM/FDR upper tail overlapping terms were significant as before but none of the LM/FDR lower
tail terms were significant. Thus, the subset of LM tonal words that overlap the lower tail FDR
words lose their predictive power when tail area data are removed, but the remaining FDR words
still have predictive value for IPO returns.