



# IPO first-day returns, offer price revisions, volatility, and form S-1 language<sup>☆</sup>

Tim Loughran<sup>\*</sup>, Bill McDonald<sup>1</sup>

Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556-5646, USA



## ARTICLE INFO

### Article history:

Received 10 May 2012

Received in revised form

19 November 2012

Accepted 24 December 2012

Available online 27 February 2013

### JEL classification:

G14

G18

G24

### Keywords:

IPOs

First-day returns

Bookbuilding

S-1 filings

Textual analysis

## ABSTRACT

Form S-1 is the first SEC filing in the initial public offering (IPO) process. The tone of the S-1, in terms of its definitiveness in characterizing the firm's business strategy and operations, should affect investors' ability to value the IPO. We find that IPOs with high levels of uncertain text have higher first-day returns, absolute offer price revisions, and subsequent volatility. Our findings provide empirical evidence for the theoretical models of uncertainty, bookbuilding, and prospect theory.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

For U.S. firms, one of the first steps in going public is filing a Form S-1 on the Securities and Exchange Commission's (SEC) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. Typically, the S-1 offers investors their first detailed glimpse of a firm's business model and financial statements. For more anticipated initial

public offerings (IPOs), like Google, Facebook, or Groupon, the S-1 filing can generate extensive media coverage.

Imagine an investor trying to value an IPO on the basis of the S-1 filing. U.S. IPOs do not typically have a long history of tangible information (i.e., past positive earning streams, revenue, or even dividends) to facilitate forecasting cash flows. IPOs usually have negative trailing earnings, few assets-in-place, and limited experience/success selling their products. The S-1 filing, however, will have ample amounts of intangible information concerning future business plans or potential problems that might arise. The tone of the document, in terms of its definitiveness in characterizing the company's business strategy, its articulation of the firm's competitive position within an industry, and confidence in projecting financial outcomes, should impact investors' assessment of value. The more uncertain text contained in an S-1 filing, the more ambiguous are future cash flow projections, and thus the more challenging it is for an investor to value the IPO.

<sup>☆</sup> We thank Daniel Bradley, Stephen Ferris, James Fuehrmeyer, Gerald Hoberg, Jennifer Marietta-Westberg, Jay Ritter, Paul Schultz, William Schwert (the editor), Sophie Shive, Kent Womack, two anonymous referees, and seminar participants at University of South Florida for helpful comments.

<sup>\*</sup> Corresponding author. Tel.: +1 574 631 8432.

E-mail addresses: [Loughran.9@nd.edu](mailto:Loughran.9@nd.edu) (T. Loughran), [mcdonald.1@nd.edu](mailto:mcdonald.1@nd.edu) (B. McDonald).

<sup>1</sup> Tel.: +1 574 631 5137.

Specifically, we argue that S-1 filings with higher proportions of uncertain/weak modal words like *may*, *could*, *depend*, and *approximately*, or negative words like *loss*, *failure*, *termination*, and *adverse* should make it more difficult for investors to precisely assimilate the value-relevant information, thus generating more uncertainty about future outcomes. In a similar context, Epstein and Schneider (2008) note that when the information quality of an asset is difficult to judge, investors will treat signals as ambiguous and demand compensation for holding the asset.

Our paper links uncertain text in the Form S-1 language with first-day returns, offer price revisions, and subsequent volatility for a sample of 1,887 completed U.S. IPOs during 1997–2010. Supplementing this sample with data from 793 withdrawn IPOs, we also find that uncertain text only marginally impacts the probability of the offering being withdrawn. Managers and underwriters co-author the S-1. Does the tone of their writing provide any guidance about the uncertainty surrounding the IPO's valuation? Several key theories of IPO pricing (see Ritter, 1984; Rock, 1986) predict that uncertainty should matter in initial returns. Yet, the previous literature offers few ex ante and direct proxies for uncertainty. For example, firm age, sales, and IPO gross proceeds are frequently used as ex ante proxies of uncertainty but clearly could be measuring many other aspects of the offering. Alternatively, the standard deviation of ex post returns produces impressive empirical relations, but obviously is only available after the initial price is observed.<sup>2</sup>

We propose that the S-1's tone provides a direct proxy for ex ante uncertainty about an IPO's valuation. To measure document tone, we use the Loughran and McDonald (2011) six sentiment word lists (uncertain, weak modal, negative, positive, legal, and strong modal). The Loughran and McDonald word lists have been used in the literature to gauge tone in newspaper articles (Gurun and Butler, 2012; Dougal, Engelberg, Garcia, and Parsons, 2012), 10-Ks (Feldman, Govindaraj, Livnat, and Segal, 2010; Jegadeesh and Wu, 2011), time-stamped CNBC transcripts (Shabani, 2011), articles published on Seeking Alpha's Web site (Chen, De, Hu, and Hwang, 2011), and even managerial vocal cues in conference calls (Mayew and Venkatachalam, 2012). Their six word lists were created specifically for financial documents.

Our sample is intensive in firms with low sales and negative trailing earnings. Some IPOs, in spite of having zero trailing sales, had enormous market values right after going public (Corvis Corp.: \$28 billion, WebVan Group: \$8 billion, and EarthShell Corp: \$2.4 billion). These types of low sales, negative earnings firms are exactly where we would expect the subtlety of word choice to be most important to investors. It is not clear that our emphasis on uncertain, weak modal, and negative words to measure the IPO's information environment would apply equally as well to large, established firms with highly profitable operations.

Low trailing sales IPOs with high uncertain/negative tone in the S-1 filing are often backed by venture capitalists (VCs). The positive linkage we identify between uncertain/negative text and VC backing should not be surprising. We would expect IPOs backed by VCs to have more uncertain text, since venture capitalists are often viewed as capital providers of the last resort.

Beatty and Ritter (1986) demonstrate a positive link between ex ante uncertainty about an IPO's value and its expected initial return. According to the two authors, firms with higher ex ante uncertainty should experience higher first-day returns. Consistent with their hypothesis, we report that uncertain, weak modal, and negative word frequencies in the Form S-1 are significantly related to first-day returns after controlling for other valuation-relevant variables.

We find that the percentages of uncertain, weak modal, and negative words in the S-1 are much more powerful variables in explaining levels of underpricing than many commonly used IPO control variables, such as venture capital dummy, top-tier underwriter dummy, or trailing annual sales. For example, we find that a one-standard deviation increase in the proportion of weak modal or negative words is positively linked to an increase in first-day returns by an economically significant 4%. A one-standard deviation increase in the proportion of S-1 uncertain words is associated with a 3% increase in first-day returns.

Although the role of weak modal and uncertainty word lists in measuring ambiguity in the initial valuation process is obvious, negative words in the context of IPO pricing can be viewed from two perspectives. One could posit that a preponderance of negative words might have a negative effect on the expected offering price. However, to the extent that the tone of an S-1 filing has already been rationally impounded in the offer price, we would not expect negative words to have a mean effect on underpricing. Alternatively, as we argue, negative words can be seen as another source of uncertainty in projecting cash flows. Projecting presumably positive future cash flows for a firm reporting a long history of losses is clearly more difficult than making a positive projection for a firm with a stable history of positive cash flows. In this case, negative words should contribute to the uncertainty in the initial valuation and thus be linked to the level of underpricing.

In addition, prior literature links potential negative outcomes with ex post uncertainty as measured by stock return volatility. For example, Clayton, Hartzell, and Rosenberg (2005) find long-lived increases in volatility following a change in the Chief Executive Officer (CEO). After forced CEO departures, the increases in subsequent volatility are even larger. Kothari, Li, and Short (2009) find that unfavorable news stories about a firm are linked with higher stock return volatility. In a principal component analysis, we find that the uncertain, weak modal, and negative word lists do appear to be measuring the same underlying attribute, a result consistent with the uncertainty interpretation of negative words.

Future researchers should consider using an aggregate uncertainty word list drawing words from the union of

<sup>2</sup> Jenkinson and Ljungqvist (2001) in their Table 3.2 provide a useful summary of variables used to measure uncertainty in the IPO process.

the uncertain, weak modal, and negative word lists.<sup>3</sup> This recommendation seems sensible if uncertainty comes in multiple forms (relating to strategic uncertainty, demand/sales uncertainty, and earnings predictability uncertainty). After removing word overlap between the three lists, we find that a one-standard deviation increase in the percent of aggregate uncertainty words is linked with a 5% increase in first-day returns.

When we examine post-IPO stock return volatility, we find that higher frequencies of uncertain, weak modal, and negative words are positively and significantly linked with higher volatility in a 60-day period following the offering. This finding suggests that the larger the proportion of uncertain/negative words in an S-1, the greater is the valuation uncertainty surrounding the firm, a result consistent with the evidence of Loughran and McDonald (2011). They find that uncertain, weak modal, and negative word proportions have positive and statistically significant coefficient values in regressions with stock return volatility in the one year following the 10-K filing as the dependent variable. Further, Tetlock (2007) reports that pessimism in a popular *Wall Street Journal* column weakly predicts subsequent increases in market volatility.

Our results also shed light on two other explanations of underpricing: prospect theory and bookbuilding. One prediction of prospect theory is that the tone of the S-1 could be related to upward price revisions. More uncertain/negative S-1 tone, perhaps reflecting a weak informational position of issuers relative to underwriters, allows for greater partial adjustment of offer prices. We report results consistent with this prediction.

Our empirical results also provide some indirect evidence on bookbuilding (Benveniste and Spindt, 1989; Benveniste and Wilhelm, 1990; Sherman and Titman, 2002). We report that IPOs with high levels of weak modal or negative words have higher absolute revisions in their offer prices. This could be interpreted as issuers using uncertain language in the S-1 to attract information production [fully consistent with Sherman and Titman (2002)]. On the other hand, we find that the tone of the final IPO prospectus (Form 424) is not a better predictor of subsequent volatility than text from the S-1. The tone of the S-1 is typically very similar to the tone of the final IPO prospectus.

We also find that changes in tone between the S-1 and 424 filings are unrelated to positive revisions in the offer price. In the context of bookbuilding, these results indicate that the information production process does not substantively impact the evolution of an IPO filings' aggregate tone. Note that this result does not preclude information production that could be reflected in very specific changes in financial statements or accounting estimates. That is, changing the tone of a single phrase, which is unlikely to be captured in our aggregate counts, or changing non-tonal statements of fact could have nontrivial implications for valuation. In fact, lack of tone revision might be predicted in the context of the

information production theory. Once the information providers have been attracted by the uncertainty contained in the S-1 text, retaining the negative text could be reasonable given the liability risks facing managers.

Our paper reports that more calendar time between the S-1 and 424 filings (the final IPO prospectus) is associated with significantly lower first-day returns. Companies with more problematic S-1 filings or facing adverse market conditions are often delayed in issuing stock. Some of the delay in going public could be due to the many days management needs to properly respond to SEC concerns about the IPO document or a sharp decline of IPO market conditions. When we examine the S-1 text of 793 withdrawn IPOs, we find that higher uncertain/weak modal frequencies marginally raise the probability of the IPO being withdrawn.

Surprisingly few academic papers have examined the role of S-1 filings in the pricing of initial public offerings. Papers usually instead focus on the information contained in the final IPO prospectus, which is actually filed on EDGAR on the day of or a few days following the IPO. In contrast to the Form 424, the Form S-1 is filed on EDGAR with ample time (median value of 88 calendar days prior to the IPO) for investors to analyze the document.

Two papers similar to our line of research are Hanley and Hoberg (2010, 2012). Hanley and Hoberg (2010) split the information contained in the S-1 into standard and informative components. Interestingly, they find that IPOs with more informative content in their S-1 have lower offer price revisions and first-day returns. In their later paper extending the litigation risk work of Lowry and Shu (2002) and Field, Lowry, and Shu (2005), Hanley and Hoberg (2012) find that issuers trade off first-day returns and disclosure in amended filings (S-1/A) as a hedge against the risk of future lawsuits. The two authors show that IPOs having higher risk of material omissions in their filings are more likely to hedge litigation exposure with higher levels of underpricing. A greater level of disclosure by issuers, as proxied by more meaningful revisions in their S-1/A filings during the bookbuilding period, lowers the probability of being sued by investors.

The key assumption of Hanley and Hoberg (2012), p. 243 is that an issuer having a large pre-IPO price adjustment but having only revised their filings slightly "is likely to have a material omission in the prospectus as the issuer did not disclose the information underlying the price change." Their focus is on whether or not the IPO firm changed the text in their S-1/A filings, versus our focus on the initial tone and its relation to underpricing, offer price revisions, and ex post volatility.

In summary, our paper analyzes the tone of information contained in the S-1 filing. In agreement with the predictions of prospect theory, we find that weak modal/negative words are related to upward revisions in the offer price. In the context of bookbuilding theories of information production, we report that the word frequencies in the S-1 (i.e., initial prospectus) are just as good at predicting post-IPO return volatility as the tone of the final IPO prospectus (Form 424). Further, changes in word frequencies between the Form 424 and S-1 are not significantly correlated with upward offer price revisions.

<sup>3</sup> This aggregated word list is provided at [http://www.nd.edu/~mcdonald/Word\\_Lists.html](http://www.nd.edu/~mcdonald/Word_Lists.html).

Thus, once information providers have been drawn to an offering, there might be no ex post incentive for managers to change the tone of the text.

Our central hypothesis is that the certainty and precision of a manager's S-1 writing when discussing future business strategies or current operations should impact IPO valuation. Our findings are consistent with the hypothesis that large amounts of uncertain text in an S-1 filing generally lead to more valuation uncertainty and, in turn, higher first-day returns, absolute offer price revisions, and subsequent return volatility.

## 2. Data and sample construction

Our initial sample includes 1,887 completed U.S. IPOs with an offer price of at least \$5 per share during the 1997–2010 time period. The SEC's date for mandatory electronic filing varies by form and firm size, with the number of S-1 filings appearing to stabilize in 1996. Thus, to allow for pre-IPO date S-1 filings, we begin our sample with firms going public in 1997.<sup>4</sup> Financial firms (i.e., banks and savings and loans), American Depository Receipts (ADRs), real estate investment trusts (REITs), and closed-end funds are excluded from the sample. Our sample is obtained from Thomson Financial Securities Data with additional items and corrections supplied by Professor Jay Ritter. The Center for Research in Security Prices (CRSP) is the source for stock prices and returns after the IPO.

Form S-1 is the initial document on the EDGAR system for registering IPO stock offerings with the SEC. After the initial S-1 filing, the company typically will submit additional amended filings (S-1/A filings) to the EDGAR system. Form 424 is the final IPO prospectus, filed at or within days after the IPO. About 24% of our sample actually filed the Form 424 on EDGAR after the date when the IPO started trading on a major exchange according to CRSP. Unlike Hanley and Hoberg (2012) and Ferris, Hao, and Liao (forthcoming), the attention of this paper will primarily be on the S-1 filings.<sup>5</sup>

### 2.1. Parsing the S-1 and 424 documents

All Form 424 variants (e.g., 424A, 424B1, 424B8), Form S-1, and Form S-1/A filings are downloaded from the SEC's EDGAR Web site for the period 1994–2010. Filings for each firm in the IPO sample are then identified based on the SEC's Central Index Key (CIK). For each firm, we consider the first S-1 filing and the last filing occurring during days 0–5 relative to the IPO date. The last filing can be an S-1/A or a 424 variant. For each individual document, we then follow the parsing procedures described in

Loughran and McDonald (2011) and remove all ASCII-encoded segments (e.g., pdf's, jpg's), HTML, and XBRL.<sup>6</sup> The remaining text is then parsed into words, creating the various counts and proportions we use in the subsequent analysis.

Fig. 1 reports the median number of words contained in the S-1 and 424 filings as well as the number of IPOs by calendar year. The median number of words in both filings significantly increased during our sample period. The typical 1997 IPO contained less than 40,000 words in either the Form S-1 or 424 while the documents for the typical 2010 IPO had about 80,000 words. Li (2008) reports that 10-Ks also experienced a sharp rise in the average number of words since the 1990s. The median Form 424 has about 8% more words than the median Form S-1.

As reported in other papers (see, Gao, Ritter, and Zhu, 2011; Doidge, Karolyi, and Stulz, 2012), there has been a major drought in the number of U.S. firms going public over the last ten years. The early years of our sample had a large number of firms going public while the post-Internet bubble period had substantially fewer new offerings. The sample size reached its nadir in calendar year 2008 with only 18 IPOs.

### 2.2. Word lists

Unlike Hanley and Hoberg (2010, 2012) and Hoberg and Phillips (2010), whose focus is on gauging document likeness using the cosine similarity method, we examine text for simple frequencies of various word lists. To gain a measure of document tone, we use the Loughran and McDonald (2011) word lists to classify words into uncertain, weak modal, negative, positive, legal, and strong modal categories.<sup>7</sup> The two authors developed various word lists on the basis of their likely usage in financial documents. Thus, any of the 291 words in the uncertain word list, can reflect uncertainty regarding the IPO's future performance when appearing in financial documents.

Examples of uncertain words include *assume*, *approximately*, *risk*, and *believe*. The other lists are comprised of 27 weak modal words (e.g., *may*, *could*, *might*, and *possible*), 2,349 negative words (e.g., *loss*, *failure*, *decline*, *bankruptcy*, and *difficult*), 354 positive words (e.g., *beneficial*, *successful*, and *strong*), 871 legal words (e.g., *contract*, *lawsuits*, and *plaintiff*), and 19 strong modal words (e.g., *always*, *must*, and *will*). It is important to note that various words are sometimes on more than one list. For example, the 27 weak modal words are a subset of the uncertain word list. Also, words like *volatility*, *risky*, *unexpected*, and *unpredictable* are on both the uncertain and negative word lists. Although the Loughran and McDonald word lists are

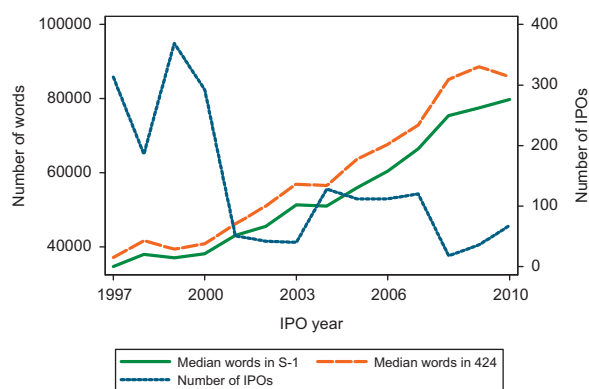
<sup>4</sup> The number of S-1 filings in 1994, 1995, and 1996 were 44, 88, and 928, respectively. The average number of filings per year for 1997–2010 was 953.

<sup>5</sup> Ferris, Hao, and Liao (forthcoming) examine Form 424 filings during 1999–2005 and find mixed results linking first-day returns and negative word frequencies. In their paper, the three authors did not examine S-1 filings, the impact of weak modal or uncertain word lists, offer price revisions, withdrawn IPOs, nor subsequent volatility.

<sup>6</sup> All of the S-1 filings in our sample contain a section entitled "Part II, Information not required in the prospectus." Since this section is specifically not included in the final IPO prospectus, varies randomly in length, and usually is dominated by legal contracts, we, like Hanley and Hoberg (2012), exclude this section from all analysis.

<sup>7</sup> The 2012 version of the word lists is available at [http://www.nd.edu/mcdonald/Word\\_Lists.html](http://www.nd.edu/mcdonald/Word_Lists.html).





**Fig. 1.** This graph plots the median number of words contained in the S-1 and 424 filings and the number of IPOs for calendar years 1997–2010. Form S-1 is the initial filing on EDGAR for registering the IPO offering with the Securities and Exchange Commission (SEC). Form 424 is the final IPO prospectus. The sample includes 1,887 US IPOs during 1997–2010 with an offer price of at least \$5 per share.

based on a sample of 10-Ks, they should apply equally well to S-1 and 424 filings.

For completeness, we will initially use all six word lists taken from Loughran and McDonald (2011), although we do not expect that all of these lists are likely to be relevant in the current context. Thus, higher proportions of positive S-1 text should not have the same effect for investors as when text is intensive in negative or uncertain word counts. Veronesi (1999) and Epstein and Schneider (2008) are examples of models that tell us that the response to positive and negative new information is not necessarily symmetric.

Additionally, prior work in textual analysis suggests that positive tone is difficult to accurately measure, since positive words are easily negated in ways difficult to programmatically identify. Both Tetlock (2007) and Loughran and McDonald (2011) note that the results for positive words are mixed because many times, negative phrases are wrapped in positive words. For example, as noted by managers of VA Linux Systems in their S-1: “Our business strategy may not be successful and we may not successfully manage these risks.”

### 2.3. IPO control variables

Our initial set of tests examines first-day returns for the IPO sample. First-day returns are defined as the percentage change from the offer price to the closing price. In our regressions, we use seven different control variables taken from the prior IPO literature on the basis of their ability to explain first-day returns.

The IPO specific control variables are: (1) *Up revision*—percentage upward revision from the mid-point of the filing range if the offer price is greater than the mid-point, else zero; (2) *VC dummy*—a dummy variable set to one if the IPO is backed by venture capital, else zero; (3) *Top-tier dummy*—a dummy variable set to one if the IPO’s lead underwriter has a value of eight or more using Carter and Manaster (1990) rankings as updated by Carter, Dark, and Singh (1998) and Loughran and Ritter (2004), else zero;

(4) *Positive EPS dummy*—a dummy variable set to one if the IPO has positive earnings per share (EPS) in the 12 months prior to going public, else zero; (5) *Prior Nasdaq 15-day returns*—the buy-and-hold returns of the CRSP Nasdaq value-weighted index over the 15-trading days prior to the IPO date; (6) *Share overhang*—the number of shares retained divided by the number of shares in the initial offering; and (7) *Sales*—trailing firm annual sales in millions of dollars. A number of prior papers use similar variables to explain first-day returns (see Bradley and Jordan, 2002; Hanley and Hoberg, 2012; Loughran and Ritter, 2004). See the Appendix for more detailed variable definitions.

### 2.4. Data summary statistics

Summary statistics for the IPO sample are reported in Table 1. Our high mean (34.8%) and median (13.3%) first-day returns are a product of the large sample volume of IPOs during the Internet bubble. The 661 IPOs during 1999–2000 averaged 71.3% in first-day returns compared to 15.1% for the 1,226 IPOs during 1997–1998 and 2001–2010. As is typical with IPO samples, the firms are relatively small, with average trailing annual sales of only \$304 million. The average upward revision in the offer price from the mid-point of the filing price range is 11.7%. More than half of the IPOs are backed by venture capital while more than 75% of the sample is taken public by a top-tier underwriter.

Consistent with the evidence from Gao, Ritter, and Zhu (2011), a large fraction of the IPO sample has negative trailing EPS values. Only 37% of our sample has positive annual earnings in the year before going public. There is large variation in the number of days between the S-1 and 424 filings. While on average, 116.6 calendar days separate the two filings, the 5<sup>th</sup> percentile is 55 days compared to 280 days for IPOs at the 95<sup>th</sup> percentile.

Table 1 also reports sentiment word list percentages for both the S-1 and 424 filings. For each of the word lists, the mean and median values are quite similar. For example, the mean (1.31%) and median (1.32%) percentage of uncertain words in the S-1 are almost identical. The mean, median, and standard deviation for all the word lists from the S-1 are very close to the values contained in the 424. The S-1, for example, has mean, median, and standard deviation values for % *Uncertain* of 1.31%, 1.32%, and 0.19% compared to the respective values of 1.28%, 1.28%, and 0.18% in the 424.

Thus, although the final IPO prospectus contains several thousand more words, overall the typical tone for a 424 is very similar to the typical S-1 document tone. The word frequencies across the S-1 and 424 documents are highly correlated. For example, between the S-1 and 424 documents, the % *Uncertain* correlation is 0.936; the % *Weak modal* correlation is 0.965; while there is a 0.955 correlation for % *Negative*.

The correlations between key variables are reported in Table 2. All of the S-1 word list frequencies are highly linked with first-day returns except for strong modal words. Weak modal, negative, and positive word frequencies all have correlations greater than 0.15 with initial returns. Uncertain, weak modal, and negative word

**Table 1**

Summary statistics for IPO sample, 1997–2010.

The sample includes 1,887 U.S. IPOs with an offer price of at least \$5 per share. Financial firms, ADRs, REITs, and closed-end funds are excluded from the sample. Our sample is obtained from Thomson Financial Securities Data with additional items and corrections supplied by Professor Jay Ritter. Form S-1 is the initial EDGAR filing for registering the IPO offering with the Securities and Exchange Commission (SEC). Form 424 is the final IPO prospectus. The Loughran and McDonald (2011) word lists are used to classify words into uncertain, weak modal, negative, positive, legal, and strong modal categories. *Up revision* is the percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-point. See Appendix for other variable definitions.

Variable	Mean	Std. dev.	5th	Median	95th
<i>First-day returns</i>	34.8%	65.2%	−8.2%	13.3%	173.9%
<i>Sales (in millions)</i>	\$304.2	\$2,693	\$0.1	\$37.0	\$983.3
<i>Up revision</i>	11.7%	21.7%	0%	0%	50.0%
<i>VC dummy</i>	0.51	0.50	0	1	1
<i>Top-tier dummy</i>	0.78	0.42	0	1	1
<i>Positive EPS dummy</i>	0.37	0.48	0	0	1
<i>Prior Nasdaq 15-day return</i>	1.14%	5.90%	−8.86%	1.43%	10.40%
<i>Share overhang</i>	3.49	2.23	1.04	2.97	7.87
<i>Days between S-1 and 424</i>	116.6	98.2	55	88	280
<i>Post IPO return volatility</i>	3.21	3.67	0.34	1.79	10.12
<b>Form S-1 (initial IPO filing)</b>					
<i>Number of words</i>	47,827	19,272	28,214	42,027	83,605
% Uncertain	1.31	0.19	0.99	1.32	1.61
% Weak modal	0.64	0.17	0.35	0.64	0.92
% Negative	1.41	0.30	0.89	1.42	1.88
% Positive	0.97	0.17	0.71	0.97	1.26
% Legal	0.75	0.20	0.49	0.71	1.13
% Strong modal	0.53	0.11	0.36	0.52	0.72
<b>Form 424 (final IPO prospectus)</b>					
<i>Number of words</i>	52,784	22,594	30,012	45,890	95,140
% Uncertain	1.28	0.18	0.98	1.28	1.57
% Weak modal	0.62	0.16	0.35	0.62	0.89
% Negative	1.41	0.29	0.92	1.42	1.85
% Positive	0.94	0.16	0.69	0.93	1.21
% Legal	0.72	0.21	0.45	0.69	1.11
% Strong modal	0.53	0.11	0.36	0.52	0.72

**Table 2**

Correlations for the sample of 1,887 IPOs, 1997–2010.

The sample includes 1,887 US IPOs with an offer price of at least \$5 per share. Financial firms, ADRs, REITs, and closed-end funds are excluded from the sample. Our sample is obtained from Thomson Financial Securities Data with additional items and corrections supplied by Professor Jay Ritter. Form S-1 is the initial EDGAR filing for registering the IPO offering with the Securities and Exchange Commission (SEC). Form 424 is the final IPO prospectus. *First-day returns* is the percentage change from the offer price to the closing price. *Days between the S-1 and 424* is the number of calendar days between the S-1 and 424 filings. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. *Up revision* is the percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-point. *VC dummy* is set to one if the IPO is backed by venture capital, else zero. *Sales* is the trailing firm sales in millions of dollars. *Post-IPO return volatility* is the market model root-mean square error for each IPO over day +5 to day +64 relative to the IPO date. *Positive EPS dummy* is set to one if trailing EPS is positive, else zero. See Appendix for more detailed variable definitions.

	<i>First-day returns</i>	<i>Days between (S-1 &amp; 424)</i>	% Uncertain	% Weak modal	% Negative	% Positive	% Legal	% Strong modal	<i>Up revision</i>	<i>VC dummy</i>	<i>Sales</i>	<i>Post-IPO return volatility</i>
<i>First-day returns</i>	1.000											
<i>Log(days between S-1 &amp; 424)</i>	−0.159	1.000										
% Uncertain	0.106	0.089	1.000									
% Weak modal	0.183	0.081	0.835	1.000								
% Negative	0.151	0.103	0.600	0.674	1.000							
% Positive	0.171	−0.064	0.293	0.419	0.253	1.000						
% Legal	−0.184	0.063	0.019	−0.011	0.209	−0.222	1.000					
% Strong modal	0.030	−0.029	0.158	0.253	0.065	0.150	0.089	1.000				
<i>Up revision</i>	0.702	−0.135	0.070	0.144	0.127	0.156	−0.173	0.024	1.000			
<i>VC dummy</i>	0.243	−0.033	0.284	0.449	0.371	0.323	−0.173	0.123	0.226	1.000		
<i>Sales</i>	−0.037	0.023	−0.074	−0.102	0.006	−0.023	0.060	−0.066	−0.006	−0.099	1.000	
<i>Post-IPO return volatility</i>	0.367	−0.138	0.142	0.274	0.158	0.236	−0.219	0.095	0.319	0.283	−0.072	1.000
<i>Positive EPS dummy</i>	−0.162	0.054	−0.216	−0.300	−0.337	−0.186	0.100	−0.168	−0.148	−0.387	0.087	−0.298

frequencies are negatively correlated with the positive EPS dummy (correlations of  $-0.216$ ,  $-0.300$ , and  $-0.337$ , respectively). Weak modal, negative, and positive words have correlations with upward revisions in the offer price of more than  $0.125$ . As noted earlier, there is overlap between some of the word lists (i.e., all of the weak modal words are also defined as uncertain words). This explains the very high correlation between the weak modal and uncertainty lists ( $0.835$ ). Given the correlations and word overlap between the lists, we will run the regressions separately for each word list.

### 3. Empirical results

#### 3.1. First-day returns

In this section, we examine the link between first-day returns and the language used by managers and investment bankers in Form S-1. Table 3 reports the regression results with first-day returns, defined as the percentage change from the offer price to the closing price, as the dependent variable. All regressions in this paper also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. In all of the regression tables, the  $t$ -statistics are in parentheses and based on the standard errors clustered by year and industry.

The first two columns report regression results with the traditional IPO control variables. The only difference between the regressions in columns 1 and 2 is that *Up revision* is not in the first column. For the results in column 1, five of the six independent variables are statistically significant in explaining first-day returns. IPOs with venture capital-backing or top-tier underwriters have higher initial returns. The higher the returns on the value-weighted Nasdaq index over the 15-days prior to the IPO, the lower the trailing annual sales, and more shares retained by insiders relative to shares issued, the larger are first-day returns. The *Share overhang* variable should be thought of as a scarcity measure. The fewer shares offered to the public, the stronger should be investor demand for the asset. The only control variable not significant in the column 1 regression is the *Positive EPS dummy* variable. The  $R^2$  value for the first regression is  $32.8\%$ .

When *Up revision* is added as a control variable in column 2, the  $R^2$  value of the regression increases to  $56.0\%$  and three of the other variables (*VC dummy*, *Top-tier dummy*, and *Log(sales)*) are no longer significant. Consistent with several other papers (Hanley, 1993; Lowry and Schwert, 2002), a very substantial component of IPO first-day returns is the degree to which underwriters upwardly adjust the offer price relative to the mid-point of the initial IPO filing price

**Table 3**

Regressions with first-day returns as the dependent variable with Form S-1 sentiment proportions, 1997–2010.

The dependent variable, *First-day returns*, is the percentage change from the offer price to the closing price. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The  $t$ -statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Log(days between S-1 and 424)</i>			−4.93 (−2.18)	−4.92 (−2.19)	−4.95 (−2.13)	−5.15 (−2.13)	−5.18 (−2.15)	−5.05 (−2.09)
% Uncertain			17.18 (2.65)					
% Weak modal				23.23 (3.17)				
% Negative					14.27 (2.62)			
% Positive						2.40 (0.61)		
% Legal							−3.59 (−0.79)	
% Strong modal								−12.90 (−1.46)
<b>Control variables</b>								
<i>Up revision</i>		1.70 (20.54)	1.69 (20.35)	1.69 (19.98)	1.68 (19.99)	1.69 (20.43)	1.69 (20.54)	1.69 (20.72)
<i>VC dummy</i>	9.29 (1.94)	2.58 (0.85)	2.31 (0.78)	1.60 (0.60)	1.52 (0.55)	2.50 (0.84)	2.41 (0.83)	2.37 (0.78)
<i>Top-tier dummy</i>	9.01 (2.50)	2.20 (1.02)	2.18 (1.10)	1.99 (0.98)	1.89 (0.89)	1.80 (0.90)	1.84 (0.90)	2.07 (1.02)
<i>Positive EPS dummy</i>	2.39 (0.96)	2.93 (2.24)	2.92 (2.18)	2.64 (2.05)	4.38 (2.91)	2.94 (2.26)	2.96 (2.22)	3.04 (2.19)
<i>Prior Nasdaq 15-day returns</i>	2.18 (3.57)	0.95 (3.14)	0.94 (3.16)	0.95 (3.14)	0.97 (3.28)	0.97 (3.14)	0.97 (3.13)	0.98 (3.26)
<i>Share overhang</i>	8.39 (3.70)	4.19 (2.81)	4.09 (2.81)	4.01 (2.79)	4.05 (2.76)	4.15 (2.86)	4.16 (2.84)	4.21 (2.88)
<i>Log(sales)</i>	−2.50 (−3.07)	−1.46 (−1.66)	−0.83 (−0.91)	−0.54 (−0.56)	−1.15 (−1.38)	−1.38 (−1.59)	−1.40 (−1.58)	−1.77 (−1.84)
$R^2$ value	32.8%	56.0%	56.2%	56.2%	56.3%	56.1%	56.1%	56.1%

range. The farther the offer price is moved above the mid-point filing price, the higher are first-day returns.

Each of the last six regressions in Table 3 contains a different Loughran and McDonald (2011) word list as the main independent variable. These columns also include the seven IPO control variables as well as *Log(days between S-1 and 424)*. The coefficient on *Log(days between S-1 and 424)* is consistently negative and significant. A longer period between the filings indicates that more calendar time was needed to address concerns from the SEC and/or there are periods where high stock market volatility temporarily restricts the equity issuing ability of firms.

Half of the word lists have insignificant coefficients (% *Positive*, % *Legal*, and % *Strong modal*) while half have positive and statistically significant coefficient values in the presence of the control variables. In column 3, % *Uncertain* has a positive coefficient (*t*-statistic of 2.65). IPOs with a higher proportion of words like *uncertain*, *may*, *approximate*, and *pending* experience higher first-day returns.

What is the economic significance of the uncertain word proportion in the S-1? Multiplying the regression coefficient by the standard deviation of % *Uncertain* (17.18 times 0.19) implies an economically significant change in first-day returns of 3.3%. For IPOs, the amount of uncertain words in the initial filing is an important variable in explaining first-day returns.

In column 4, the coefficient on % *Weak modal* is 23.23 (*t*-statistic of 3.17). As with the proportion of uncertain words in the S-1, the frequency of weak modal words is an economically significant factor. A one-standard deviation increase in the percentage of weak modal words is linked with a 3.9% increase in first-day returns (23.23 coefficient value multiplied by standard deviation of 0.17).

Column 5 reports that % *Negative* is also related to first-day returns. IPOs with more negative tone in their S-1 experience higher first-day returns. A one-standard deviation increase in the percentage of negative words in the S-1 is associated with a 4.3% rise in first-day returns (14.27 coefficient value multiplied by the standard deviation of 0.30). More negative tone in the S-1 from underwriters and management, probably reflecting bleak operating performance and an uncertain future against larger, more successful competitors, is linked with more underpricing. Notice that S-1 uncertain, weak modal, and negative word frequencies have a more important role in explaining first-day returns than commonly used explanatory variables like *VC dummy*, *Top-tier dummy*, or trailing annual *Sales*.

In untabulated results, we also created an aggregate measure of uncertainty by combining the uncertain, weak modal, and negative word lists after removing word overlap between the three lists. On average, 2.70% of S-1 words are in this aggregate uncertainty word list. When this aggregate uncertainty measure is added as an independent variable to the Table 3 first-day return regressions, it has a coefficient value of 11.50 (*t*-statistic of 2.91). Thus, a one-standard deviation increase in the percentage of aggregate uncertainty words is associated with a 5% increase in first-day returns (11.50 multiplied by standard deviation of 0.43).

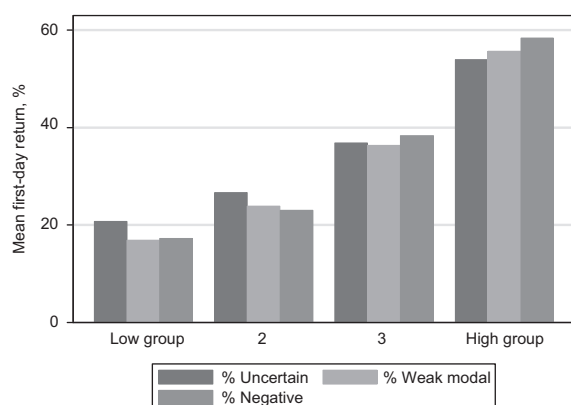
Our finding of a relation between negative word frequencies and first-day returns is in contrast to the evidence of Ferris, Hao, and Liao (forthcoming). They find (for samples of Tech and non-Tech IPOs) insignificant coefficients on the aggregate percentage of negative words in Form 424 after controlling for offer price revision as we do in Table 3. The difference between the two papers is not driven by our use of word frequencies from the S-1 instead of from the Form 424. In our Table 3 regressions, if word frequencies are obtained from the Form 424, uncertain, weak modal, and negative word frequencies all have positive and significant coefficient values in the presence of the control variables. For the three word lists, the coefficient values and *t*-statistics when using the Form 424 are 18.64 (*t*-statistic of 4.08), 23.07 (*t*-statistic of 5.85), and 11.51 (*t*-statistic of 2.83), respectively.

Focusing on the sample period (1999–2005) of Ferris, Hao, and Liao (forthcoming), they report mean and median Form 424 negative word percentages of 1.02% and 1.18% compared to our respective Form 424 values of 1.48% and 1.47%. Thus, our mean negative word frequencies from the 424, for a similar IPO sample, are about 50% higher. The difference is due to the tabulation of total word counts. They have an average word count more than 100,000 words higher than our Form 424 average during an identical time period. Ferris et al. count all tokens not containing digits as words. Our methodology, following Loughran and McDonald (2011), only counts tokens as words when they appear in a master dictionary. The master dictionary does not include proper nouns, single letter words, and industry-specific terminology (like chemical names and technical terms).

As a robustness check to the Table 3 analysis, in untabulated results, we added a number of additional control variables to the regressions in columns 3, 4, and 5. The additional control variables are *Log(number of words in S-1)*, *Log(pre-issue book value of assets)*, *Log(1 + number of years since the IPO's founding as of the offering date)*, *Nasdaq dummy* (set to one if the IPO listed initially on Nasdaq, else zero), and *Tech dummy*. The IPO founding dates are from the Field and Karpoff (2002) and Loughran and Ritter (2004) data sets. *Tech dummy* is set to one if the IPO is in the technology business [according to Loughran and Ritter (2004)], else zero. Adding these additional variables had only minor effects on the coefficient values and significance levels for uncertain, weak modal, and negative word frequencies. For example, the coefficient on % *Uncertain* increases slightly to 17.53 (from 17.18) while the coefficient on % *Negative* also rises slightly to 14.68 (from 14.27). In none of the regressions are the additional control variables significant in the presence of *Up revision*.

How much power does a simple sort of the word lists have in explaining differences in first-day returns without the presence of the IPO control variables? To address this question, Fig. 2 plots IPO first-day returns sorted by the frequencies of uncertain, weak modal, and negative words. We only want to place firms into groups using information that would have been known at the time of the IPO. Thus, each IPO during 1998–2010 is placed into one of four groups on the basis of the prior year's





**Fig. 2.** This chart plots the mean IPO first-day returns sorted, in each case, by uncertain, weak modal, and negative word quartiles during 1998–2010. First-day returns are defined as the percentage change from the offer price to the closing price. Each IPO during 1998–2010 is placed into a quartile on the basis of the prior year's percentage of uncertain, weak modal, or negative words in the S-1 filing.

distribution of the percentage of uncertain, weak modal, or negative words.

Fig. 2 reports a monotonic rise in initial returns from the smallest to largest quartile. For example, IPOs in the low weak modal quartile have average first-day returns of 16.9% compared to 55.7% for the high weak modal groups, a difference of 38.8%, between the extreme quartiles. Annually sorting IPOs by either uncertain, weak modal, or negative word frequencies clearly produces large first-day return differences between the quartiles.

### 3.2. Ten most frequent words in S-1 and 424

To assist the reader in understanding which words have a high weight in our analysis, we present in Table 4 the ten most frequently occurring words by the three word categories and by document type. As would be expected from prior research in this area, a few words generate a disproportionate share of the word count frequencies. Among the 291 uncertain words, ten words account for about 72% of all the uncertain words appearing in the S-1. Among weak modal words, the skewness is even higher with two words, *may* and *could*, accounting for about 87% of the total weak modal word count. The top ten negative words account for only about 30% of the cumulative S-1 negative word count. The table reports a strong similarity between the words used in the S-1 and the 424. For example, *may* accounts for 29.78% of the S-1 uncertain words while it represents 29.29% of the Form 424 uncertain words.<sup>8</sup>

Table 4 also shows that unusual words are not driving the word counts. All of the reported words appear to reflect the premise of their respective lists. It should not be surprising that words like *loss*, *adversely*, *adverse*, *against*, and *losses* are the most frequent negative words appearing in either the S-1 or 424 documents. Also note

that *may* and *could* are the top two words in both the uncertain and weak modal lists.

### 3.3. Link between word list levels and revisions in offer prices

An obvious extension of our analysis would be to examine the link between word lists and revisions in the offer price from the mid-point of the filing range. Hanley and Hoberg (2012) hypothesize that a portion of the partial adjustment phenomenon is linked with issuer's and banker's desire to minimize their litigation exposure. Bankers might also not fully revise the offer price in the face of receiving positive feedback from investors. Benveniste and Spindt (1989) and Sherman and Titman (2002) view underpricing as compensation to investors for revealing information about the IPO valuation to the underwriters during the bookbuilding process. Hanley and Hoberg (2010) argue that IPOs with more standard S-1 content would be more likely to solicit information from investors during bookbuilding.

In Table 5, the dependent variable is *Absolute(revision)*, defined as the absolute value of the percentage revision in the offer price from the mid-point of the filing range. We use the absolute value of the offer price revision since we are interested in the magnitude of the revision, not its direction. Our expectation would be that more uncertain IPOs, as proxied by higher frequencies of uncertain, weak modal, or negative words, would have higher absolute revisions in their offer prices. The higher the uncertainty surrounding the IPO's valuation, the more likely new information (positive or negative) revealed during the bookbuilding process will have a significant impact in offer price revisions. In all the regressions, the six different IPO control variables, an intercept, Fama and French 48-industry dummies, and calendar year dummies are included.

Across the six columns, five of the IPO control variables are consistently significant. The regressions imply IPOs with VC-backing, negative trailing EPS, higher returns on the Nasdaq index in the prior 15-days before the offering, more retained shares, and smaller trailing sales have larger absolute revisions.

Of the six word lists, three (*% Weak modal*, *% Negative*, and *% Positive*) are significantly linked with absolute revisions in the offer price. The positive and significant coefficients on *% Weak modal* and *% Negative* are consistent with bookbuilding theories. More uncertainty about the IPO valuation (proxied by frequencies of weak modal/negative words) should produce significantly larger absolute price revisions. As large institutional investors provide underwriters with their assessments of the IPO's valuation, this information should be reflected in the offer price. If the information from investors is quite positive, underwriters would increase the offer price relative to the mid-point of the filing range. Negative information from investors would have the opposite effect. This is further proof that weak modal and negative word frequencies are related to uncertainty.

It is important to note that the positive coefficient on *% Uncertain* is not statistically significant. Thus, the

<sup>8</sup> Note that the capitalized "May" is excluded from all counts.

**Table 4**

Top ten most frequent words by category and document type.

The Loughran and McDonald (2011) word lists are used to classify S-1 and 424 words for US IPOs during 1997–2010.

Panel A: Uncertain words			
Word	Form S-1: % Of uncertain count	Form S-1: Cumulative % of uncertain counts	Form 424: % Of uncertain count
May	29.78%	29.78%	29.29%
Could	12.29%	42.07%	12.06%
Approximately	10.01%	52.08%	11.43%
Believe	5.54%	57.62%	5.60%
Risk	4.25%	61.87%	4.28%
Risks	2.66%	64.53%	2.55%
Believes	2.42%	66.95%	2.43%
Assumed	2.12%	69.07%	1.56%
Intangible	1.44%	70.51%	1.43%
Assumptions	1.34%	71.85%	1.31%
Panel B: Weak modal words			
Word	Form S-1: % Of weak modal count	Form S-1: Cumulative % of weak modal counts	Form 424: % Of weak modal count
May	61.55%	61.55%	61.59%
Could	25.39%	86.94%	25.35%
Depend	2.49%	89.43%	2.37%
Possible	2.35%	91.78%	2.36%
Might	2.27%	94.05%	2.27%
Depends	1.37%	95.42%	1.29%
Appearing	1.09%	96.51%	1.17%
Depending	1.09%	97.60%	1.07%
Uncertain	0.77%	98.37%	0.78%
Nearly	0.39%	98.76%	0.38%
Panel C: Negative words			
Word	Form S-1: % Of negative count	Form S-1: Cumulative % of negative counts	Form 424: % Of negative count
Loss	10.37%	10.37%	10.56%
Adversely	2.64%	13.01%	2.46%
Adverse	2.56%	15.57%	2.44%
Against	2.44%	18.01%	2.42%
Losses	2.37%	20.38%	2.38%
Claims	2.22%	22.60%	2.33%
Closing	2.12%	24.72%	2.25%
Termination	2.12%	26.84%	2.22%
Deficit	2.03%	28.88%	2.05%
Failure	1.61%	30.48%	1.79%

frequencies of uncertain words do not help explain absolute price revisions in the presence of the control variables. Also of note is that the coefficient on % *Positive* is significant ( $t$ -statistic of 2.17). As mentioned before, we are hesitant to interpret positive words because they can be easily negated in ways that are difficult to identify computationally. However, a possible explanation for the significant coefficient on % *Positive* is that the creditability of managers/underwriters might be diminished by having an overly positive S-1. This explanation is consistent with the evidence from Larcker and Zakolyukina (2012) that deceptive CEOs use significantly more positive emotion words (e.g., *awesome*, *fabulous*, *groovy*, and *marvelous*) in conference calls.

To reduce the opacity of our results from the reader's perspective and give helpful intuition, it makes sense to provide some examples. Since our analysis is about aggregate loadings, providing specific sentence examples might be misleading. Instead, we will note that firms like Sycamore Networks, FairMarket Inc., and Registor.com

Inc. all had high levels of uncertain/weak modal/negative S-1 text coupled with large first-day returns and absolute revisions in the offer price. IPOs with low levels of uncertain text, initial returns, and absolute revisions include Fairchild Semiconductor, AECOM Technology Corp., and American Axle & Manufacturing. Fairchild Semiconductor is particularly emblematic. In spite of being a high-tech firm going public in the midst of the Internet bubble, its initial filing mirrored its maturity as a firm with a long history in technology.

### 3.4. Implications for theories of IPO pricing

Using S-1 tone, we can also empirically examine other theories of IPO pricing. Prospect theory (see Loughran and Ritter, 2002) assumes that managers will focus on the change in their wealth at the time of the IPO rather than on the more relevant benchmark, i.e., the level of their wealth. Excessive underpricing occurs when managers agree to the underwriter's request to partly increase the offer price in the

**Table 5**

Regressions with absolute value of revision as the dependent variable with Form S-1 sentiment proportions, 1997–2010.

The dependent variable, *Absolute(revision)*, is the absolute value of the percentage revision in the offer price from the mid-point of the filing range. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
<i>Log(days between S-1 and 424)</i>	–0.28 (–0.19)	–0.22 (–0.15)	–0.16 (–0.10)	–0.21 (–0.14)	–0.29 (–0.19)	–0.31 (–0.21)
% Uncertain	1.59 (0.67)					
% Weak modal		6.66 (2.28)				
% Negative			7.65 (3.02)			
% Positive				3.92 (2.17)		
% Legal					–2.63 (–1.50)	
% Strong modal						0.46 (0.14)
<b>Control variables</b>						
<i>VC dummy</i>	2.89 (2.94)	2.63 (2.60)	2.31 (2.69)	2.73 (2.84)	2.77 (2.91)	2.93 (3.04)
<i>Top-tier dummy</i>	1.15 (1.04)	1.16 (1.05)	1.12 (1.05)	1.03 (0.93)	1.11 (1.03)	1.12 (1.00)
<i>Positive EPS dummy</i>	–2.65 (–3.55)	–2.74 (–3.53)	–1.89 (–2.72)	–2.69 (–3.71)	–2.65 (–3.56)	–2.65 (–3.55)
<i>Prior Nasdaq 15-day returns</i>	0.48 (2.65)	0.47 (2.64)	0.48 (2.82)	0.48 (2.65)	0.48 (2.67)	0.48 (2.65)
<i>Share overhang</i>	2.37 (4.83)	2.33 (4.74)	2.31 (4.92)	2.35 (4.79)	2.38 (4.87)	2.38 (4.73)
<i>Log(sales)</i>	–1.11 (–3.09)	–0.92 (–2.34)	–1.02 (–3.12)	–1.13 (–3.14)	–1.16 (–3.54)	–1.15 (–3.42)
<i>R<sup>2</sup> value</i>	19.5%	19.6%	20.1%	19.5%	19.5%	19.5%

presence of excess investor demand only when insiders simultaneously receive the good news of unexpected increases in their personal wealth. Loughran and Ritter (2002) argue that leaving money on the table in the IPO should be viewed as indirect compensation for underwriters.

In the context of prospect theory, one might suppose that the lead underwriter could more easily take advantage of issuers who write an initial prospectus with high levels of uncertain language because they are in a weak informational position. Managers cannot easily conceal their weak position from the underwriter because they face litigation risk and SEC scrutiny if the prospectus contains misstatements. In such an environment, S-1 filings with high uncertain/negative language would have the underwriter optimally (1) lowball the initial offer price; (2) only partly increase the final offer price in the face of high investor demand; and (3) indirectly benefit by allocating underpriced shares that will materialize under prospect theory. Thus, prospect theory suggests that an unusually high amount of uncertain language in the S-1 generates, on average, predictable upward offer price revisions.

In Table 6 we test the prediction that uncertain/negative text helps explain upward revisions in offer prices. The dependent variable, *Up revision*, is the percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-

point, else zero. Of the six word lists, three lists (weak modal, negative, and positive) have positive and significant coefficient values while the coefficient on legal words is significant and negative. The positive coefficient values on negative (significant at the 6% level) and weak modal are consistent with lowballing of initial price estimates by underwriters and with outcomes suggested by prospect theory.<sup>9</sup> Note that, referring to the unique tonal words in Table 4 and consistent with the findings in Table 6, weak modal words (e.g., *depend*, *possible*) are more likely to reveal a lack of clear vision by management, versus uncertain words (e.g., *approximately*, *believe*, and *risk*) which are more associated with the fundamental risks of the business.

Similarly, in this context, the importance of legal terminology in making the assertions of management less ambiguous is consistent with the regression results, where upward revisions are reduced by the proportion of legalese. In the six regressions, the coefficient on *Top-tier dummy* is positive and significant. This evidence is also in agreement

<sup>9</sup> Since the focal point of Loughran and Ritter (2002) is money left on the table in the IPO, where both the offer price and first closing price are more than what was expected, we use upward revision as the variable of interest. However, if the dependent variable is the entire level of the price revision, none of the six word lists have significant coefficient values.

**Table 6**

Regressions with *up revision* as the dependent variable with Form S-1 sentiment proportions, 1997–2010.

The dependent variable, *Up revision*, is the value of the percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-point,  $((\text{offer price} - \text{mid-point}) / \text{mid-point}) \times 100$  if offer price > mid-point, else zero. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
<i>Log(days between S-1 and 424)</i>	–3.11 (–2.41)	–3.08 (–2.36)	–3.01 (–2.40)	–3.09 (–2.34)	–3.11 (–2.44)	–3.10 (–2.39)
% Uncertain	0.89 (0.40)					
% Weak modal		3.47 (2.16)				
% Negative			5.91 (1.94)			
% Positive				1.84 (2.10)		
% Legal					–2.72 (–2.49)	
% Strong modal						–1.94 (–0.41)
<b>Control variables</b>						
VC dummy	3.91 (4.11)	3.78 (3.81)	3.46 (4.03)	3.84 (3.95)	3.78 (4.38)	3.89 (4.13)
Top-tier dummy	3.78 (3.35)	3.78 (3.47)	3.76 (3.68)	3.72 (3.45)	3.75 (3.49)	3.80 (3.60)
Positive EPS dummy	–0.29 (–0.31)	–0.34 (–0.36)	0.30 (0.31)	–0.31 (–0.33)	–0.29 (–0.30)	–0.28 (–0.29)
Prior Nasdaq 15-day returns	0.73 (3.65)	0.73 (3.63)	0.73 (3.81)	0.73 (3.65)	0.73 (3.66)	0.73 (3.68)
Share overhang	2.43 (5.68)	2.41 (5.55)	2.38 (5.82)	2.42 (5.56)	2.43 (5.62)	2.44 (5.65)
Log(sales)	–0.54 (–2.36)	–0.44 (–2.15)	–0.46 (–2.40)	–0.55 (–2.44)	–0.56 (–2.19)	–0.62 (–2.41)
R <sup>2</sup> value	28.2%	28.2%	28.5%	28.2%	28.3%	28.2%

with the Lowry and Schwert (2004) finding that high quality underwriters (the bankers probably more likely to take advantage of weak issuer informational positions) are more likely to lowball the preliminary price range.

The results of Table 6, however, are also consistent with the evidence of Sherman and Titman (2002). It could be that the underwriter intentionally lowballs IPOs with high weak modal/negative text in order to create an incentive for investors to engage in information production. We could expect the IPOs with substantial uncertain/negative language to have, on average, low preliminary offer prices, large upward price revisions, and high first-day returns due to the need of bankers to compensate investors for their information production. An added benefit for IPOs in having institutional investors spending more resources to correctly price the offering is to reduce the litigation risk faced by managers (see Hanley and Hoberg, 2012).

### 3.5. Link between word list changes and revisions in offer prices

The bookbuilding theory of IPO pricing focuses on information production in the time interval between the S-1 and 424 filing. In Table 7 we link upward revisions in offer price with word list percentage changes between the

S-1 and 424 filings. For example, the *Change in % uncertain* variable is defined as the percent uncertain in the 424 minus the percent uncertain in the S-1. To the extent bookbuilding produces changes in tone from the S-1 to the 424 filings, we would expect these changes to relate to the price revision from the initial to final filing.

None of the three word lists most relevant in the prior results are significant in these regressions. In the six individual word list regressions, only the *Change in % strong modal* word list is marginally significant. The lack of significant coefficients on changes in tone does not appear to be a result of the dependent variable. If we instead select the absolute revision in the offer price (i.e., *Absolute(revision)*) as the dependent variable with the same independent variables as in Table 7, we find, in untabulated results, that none of the tone changes in the individual word lists are significant even at the 10% level.

Our measures are broad averages across an entire document, and thus not powerful enough to identify very specific changes in a few key phrases that could be an artifact of bookbuilding. Also, the lack of significant tone differences between the S-1 and 424 might be consistent with the information production theory. Once information providers have been attracted by the uncertain/negative text, there might be no reason to revise the tone

**Table 7**

Regressions with *up revision* as the dependent variable with changes in sentiment proportions, 1997–2010.

The dependent variable, *Up revision*, is the value of the percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-point,  $((\text{offer price} - \text{mid-point}) / \text{mid-point}) \times 100$  if offer price > mid-point, else zero. The Loughran and McDonald (2011) word lists are used to classify 424 and S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. *Change in % uncertain*, for example, is defined as the % *Uncertain* in the Form 424 minus the % *Uncertain* in the S-1. Thus, if the 424 has a lower percentage of uncertain words than was contained in the S-1, the *Change in % uncertain* variable will have a negative value. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
<i>Log(days between S-1 and 424)</i>	–3.13 (–2.41)	–3.21 (–2.39)	–3.10 (–2.38)	–3.10 (–2.42)	–3.08 (–2.47)	–3.00 (–2.34)
<i>Change in % uncertain</i>	–5.50 (–1.21)					
<i>Change in % weak modal</i>		–7.88 (–1.28)				
<i>Change in % negative</i>			–2.07 (–0.36)			
<i>Change in % positive</i>				5.20 (0.63)		
<i>Change in % legal</i>					9.02 (0.82)	
<i>Change in % strong modal</i>						19.16 (1.92)
<b>Control variables</b>						
<i>VC dummy</i>	3.95 (4.09)	3.97 (4.07)	3.91 (4.16)	3.91 (4.12)	3.93 (4.04)	3.93 (4.01)
<i>Top-tier dummy</i>	3.78 (3.58)	3.78 (3.58)	3.75 (3.51)	3.76 (3.50)	3.77 (3.51)	3.86 (3.57)
<i>Positive EPS dummy</i>	–0.30 (–0.31)	–0.29 (–0.30)	–0.31 (–0.32)	–0.26 (–0.27)	–0.24 (–0.25)	–0.13 (–0.14)
<i>Prior Nasdaq 15-day returns</i>	0.73 (3.65)	0.73 (3.66)	0.73 (3.67)	0.73 (3.67)	0.73 (3.64)	0.74 (3.73)
<i>Share overhang</i>	2.43 (5.61)	2.43 (5.65)	2.43 (5.63)	2.44 (5.59)	2.40 (5.55)	2.43 (5.61)
<i>Log(sales)</i>	–0.54 (–2.44)	–0.53 (–2.53)	–0.56 (–2.45)	–0.58 (–2.39)	–0.60 (–2.64)	–0.58 (–2.51)
<i>R<sup>2</sup> Value</i>	28.2%	28.3%	28.2%	28.2%	28.3%	28.4%

of the initial prospectus (especially in light of litigation risk faced by managers).

### 3.6. Link between word lists and post-IPO return volatility

Our central hypothesis is that uncertain, weak modal, and negative word frequencies proxy for ex ante uncertainty about the IPO's valuation. In an earlier paper, Beatty and Ritter (1986) use two proxies for ex ante uncertainty: (1) the log of one plus the number of different uses of proceeds mentioned in the IPO prospectus and (2) the inverse of IPO gross proceeds. Their logic for using the number of uses of proceeds is an artifact of SEC regulation. During their time period, the SEC required more speculative offerings to provide investors with detailed lists of what the firm intended to do with the net proceeds while more established IPOs could skip the details of their uses of proceeds.

We will not use either of the Beatty and Ritter (1986) uncertainty proxies in our paper for a few reasons. First, the SEC no longer treats established offerings differently from speculative deals in terms of detail in the number of uses of proceeds. Second, during the Internet bubble, several very speculative offerings had enormous proceeds from the IPO. Due to changes in SEC

enforcement and investor sentiment in speculative IPOs, Beatty and Ritter's proxies are a poor fit for our time period.

Although it involves looking into the future, we will use actual post-IPO daily stock return volatility as an ex post benchmark of uncertainty. To measure subsequent volatility, we calculate the market model root-mean square error for each IPO during days +5 to +64 (a 60-day period) relative to the IPO date. We exclude the first four days after the IPO to avoid unusual trading activity (i.e., abnormal flipping of shares) immediately following the offering.

This measure of uncertainty appears to be very reasonable. As reported in Table 2, post-IPO return volatility is positively correlated with first-day returns (0.367), supporting the assertions of Beatty and Ritter (1986) that IPOs with more uncertainty are linked with higher first-day returns. The frequencies of uncertain, weak modal, and negative words are also positively correlated with stock return volatility.

We find that the stock return volatility also seems to nicely sort IPOs by levels of uncertainty. Many IPOs during the Internet bubble (1999–2000) were speculative busts for investors. Ritter and Welch (2002) report that IPOs during 1999–2000 had average style-adjusted returns of



**Table 8**

Regressions with *post-IPO return volatility* as the dependent variable with levels of various word lists in the Form S-1 as explanatory variables, 1997–2010.

The dependent variable, *Post-IPO return volatility*, is the market model root-mean square error for each IPO over day +5 to day +64 relative to their IPO date. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, negative, positive, legal, and strong modal categories. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors adjusted for clustering within year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
<i>Log(days between S-1 and 424)</i>	–0.21 (–1.28)	–0.19 (–1.20)	–0.21 (–1.28)	–0.23 (–1.35)	–0.23 (–1.36)	–0.22 (–1.36)
% <i>Uncertain</i>	0.92 (4.95)					
% <i>Weak modal</i>		3.07 (4.34)				
% <i>Negative</i>			1.03 (3.04)			
% <i>Positive</i>				0.14 (0.31)		
% <i>Legal</i>					0.33 (0.95)	
% <i>Strong modal</i>						–0.34 (–0.63)
<b>Control variables</b>						
<i>Up revision</i>	0.01 (1.43)	0.01 (1.36)	0.01 (1.33)	0.01 (1.46)	0.01 (1.47)	0.01 (1.47)
<i>VC dummy</i>	0.15 (0.82)	0.03 (0.16)	0.09 (0.45)	0.16 (0.77)	0.18 (1.01)	0.16 (0.89)
<i>Top-tier dummy</i>	0.06 (0.31)	0.06 (0.33)	0.05 (0.24)	0.04 (0.22)	0.05 (0.24)	0.05 (0.25)
<i>Positive EPS dummy</i>	–0.36 (–2.62)	–0.40 (–2.70)	–0.25 (–2.32)	–0.36 (–2.63)	–0.35 (–2.59)	–0.35 (–2.65)
<i>Prior Nasdaq 15-day returns</i>	–0.00 (–0.32)	–0.01 (–0.37)	–0.00 (–0.21)	–0.00 (–0.24)	–0.00 (–0.25)	–0.00 (–0.21)
<i>Share overhang</i>	0.04 (1.80)	0.03 (1.00)	0.04 (1.55)	0.05 (2.04)	0.05 (2.21)	0.05 (2.48)
<i>Log(sales)</i>	–0.28 (–7.38)	–0.19 (–5.28)	–0.29 (–10.07)	–0.31 (–7.85)	–0.31 (–7.51)	–0.32 (–6.94)
<i>R<sup>2</sup> value</i>	46.7%	47.3%	46.9%	46.6%	46.6%	46.6%

–61.2%. In untabulated results, we sort our sample by post-IPO return volatility. Among the 50 IPOs with the lowest return volatility, none of these 50 firms went public during the Internet bubble (1999–2000). In contrast, 48 of the top 50 IPOs in terms of the highest post-IPO return volatility were from the bubble period. Thus, post-IPO return volatility seems to measure uncertainty in the sample.

Table 8 reports the regression results with post-IPO return volatility as the dependent variable. As before, days between the S-1 and 424 filings, the seven IPO control variables, an intercept, Fama and French 48-industry dummies, and calendar year dummies are included in all the regressions. Each of the six columns has a different word list as the main variable of interest. Across the six regressions, two IPO control variables are consistently statistically significant. IPOs with positive trailing EPS and higher trailing annual sales have lower levels of post-IPO return volatility.

Of the six word lists, only three are significantly linked with post-IPO return volatility. Column 1 reports a positive coefficient (*t*-statistic of 4.95) on % *Uncertain*; while in the second column, % *Weak modal* is also positive and

significant (*t*-statistic of 4.34); and in column 3, % *Negative* has a positive coefficient (*t*-statistic of 3.04). Thus, higher frequencies of uncertain, weak modal, and negative words in the S-1 are linked with higher levels of post-IPO return volatility.<sup>10</sup>

Loughran and McDonald (2011), using a large sample of 10-K filings, also find that their uncertain, weak modal, and negative word lists are significantly linked with subsequent return volatility in the year following a 10-K filing. In their case, they also find positive and strong modal word lists significant, which could be attributable to their focus on 10-Ks versus S-1s and their substantially larger sample size. Tetlock (2007) finds that pessimistic language in a popular newspaper column weakly predicts increases in Dow Jones Industrial Average volatility.

<sup>10</sup> If the word frequencies are generated from the Form 424 instead of the S-1, uncertain, weak modal, and negative word frequency coefficients are similarly positive and significant. A comparison of the regressions indicates that the final IPO prospectus, after incorporating information generated from the SEC and institutional investors, is not a better predictor of subsequent return volatility.

It appears that uncertain, weak modal, and negative word frequencies in the S-1 are good proxies for ex ante uncertainty surrounding the IPO's valuation. This result supports our hypothesis that document tone impacts underpricing through its relation to uncertainty in IPO pricing.

### 3.7. Link between firm characteristics and the uncertain and negative word lists

What types of firms have high uncertain, weak modal, or negative S-1 word frequencies? Table 9 regresses three different dependent variables (% *Uncertain*, % *Weak modal*, and % *Negative*) on the IPO control variables which are available at the time of the S-1 filing (i.e., when the dependent variables are observable).

Column 1 reports that firms with venture capital-backing and lower trailing annual sales have higher proportions of uncertain words. The first column regression reports a relatively high  $R^2$  value of 47.8%. The second column of Table 9 reports that IPOs with more frequent use of weak modal words are VC-backed and have lower sales. It should not be surprising that IPOs with lower past sales are more likely to use weasel words like *could*, *maybe*, and *possibly*, to describe current and future operations. The  $R^2$  value for the second regression is 63.9%.

In the third column, we find that VC-backed IPOs, lower sales, and firms with negative trailing annual earnings have more negative S-1 tone. The statistically significant negative coefficient on the *Positive EPS dummy* is consistent with managers using more negative words in describing weak historical performance. It is difficult to avoid using Loughran and McDonald (2011) negative words like *loss* and *losses*

when the firm has negative trailing EPS numbers. The regression in the third column has an  $R^2$  value of 49.4%. Thus, IPOs with lower trailing sales and venture capital-backing are associated with higher levels of the three word lists. Only the negative word list is positively linked with IPOs having negative trailing earnings.

To further examine the role that venture capitalists play in S-1 language, Panel A of Table 10 reports the percentage of uncertain, weak modal, and negative words across sales quartiles and VC-backing. This panel reports a strong negative association between sales and uncertain word usage. For the complete sample of IPOs in the lowest sales quartile (\$9 million or less in sales), the uncertain, weak modal, and negative percentages are 1.41%, 0.76%, and 1.53%. For IPOs in the largest sales quartile (trailing sales above \$130 million), the three respective word lists have significantly lower values (1.24%, 0.54%, and 1.31%). The differences between the low and high sales quartiles for each of the three word list averages are significant at the 1% level. Highly uncertain IPOs (as proxied by very low trailing sales) use more uncertain, weak modal, and negative words to reflect variability in the firm's valuation. Controlling for sales, VC-backed IPOs use a higher proportion of uncertain words in their S-1 than non-VC firms. For example, within the lowest sales quartile, VC-backed IPOs average 1.42%, 0.78%, and 1.57% for uncertain, weak modal, and negative frequencies compared to 1.39%, 0.71%, and 1.41% for non-VC-backed IPOs.

In Panel B, when the IPO sample is sorted into EPS terciles and by VC-backing, IPOs with low trailing earnings use significantly higher proportions of uncertain, weak modal, and negative words. The differences between the extreme EPS terciles are all significant at the 1% level. As with trailing sales, more uncertain IPOs, as proxied by low earnings, use more uncertain language in their initial filing with the SEC. VC-backed IPOs, regardless of trailing EPS, use more uncertain/negative words.

VC-backed IPOs possibly load more on uncertain text because venture capitalists are often viewed as capital providers of last resort. As noted in Gompers and Lerner (2001), p. 11, "Companies that lack substantial tangible assets, those that have a large degree of uncertainty about their future, or those that will endure several years of losses prior to earning money are unlikely to receive bank loans. ... It is difficult for these firms to get the necessary outside financing to fund their projects from any other source of capital than venture capital." Since less risky, more profitable firms might be able to attract bank loans or use internally generated funds for investment activity, this would leave only the less desirable firms requesting funds from VCs. Given this relation, we should expect the documented link between VC funding and IPOs with high levels of uncertain S-1 text.

### 3.8. Withdrawn IPOs

Thus far in our analysis, we have examined only the S-1 filings of firms that successfully went public. Looking at firms that completely withdrew from the IPO process might provide insights into whether the word lists relate

**Table 9**

Regressions with % *uncertain*, % *weak modal*, and % *negative* words in the Form S-1 as the dependent variables, 1997–2010.

In column 1, the dependent variable is the percentage of uncertain words in the S-1 while in column 2, the left-hand side variable is the percentage of weak modal words in the S-1. Column 3 has the percentage of negative words in the S-1 as the dependent variable. The Loughran and McDonald (2011) word lists are used to classify words into uncertain, weak modal, and negative categories. See Appendix for the definitions of the four control variables which were available at the time of the S-1 filing. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The  $t$ -statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Independent variables	Dependent variable		
	% Of uncertain words in S-1 (1)	% Of weak modal words in S-1 (2)	% Of negative words in S-1 (3)
Control variables			
VC dummy	0.021 (2.02)	0.049 (4.77)	0.086 (6.93)
Top-tier dummy	−0.015 (−1.03)	−0.001 (−0.04)	0.008 (0.36)
Positive EPS dummy	0.001 (0.06)	0.013 (1.57)	−0.100 (−11.36)
Log(sales)	−0.034 (−8.52)	−0.037 (−9.69)	−0.018 (−2.23)
$R^2$ value	47.8%	63.9%	49.4%

**Table 10**

Percentage uncertain, weak modal, and negative words in the S-1 sorted by venture capital-backing, sales, and earnings categories.

The sample includes 1,887 US IPOs during 1997–2010 with an offer price of at least \$5 per share. The Loughran and McDonald (2011) word lists are used to classify words in the S-1 into uncertain, weak modal, and negative categories. In each panel, the first value in a cell is the percentage uncertain, the second is percentage weak modal, and the third relates to the percentage of negative words. In Panel A, the IPO sample is categorized into quartiles by trailing annual sales (group low: \$9 million or less; group 2: between \$9 million and \$37 million; group 3: between \$37.1 and \$130 million; and group high: above \$130 million). Panel B sorts the sample into three trailing EPS groups.

Panel A: Percentage uncertain, weak modal, and negative words by venture capital-backing across sales quartiles				
Item	Low sales	2	3	High sales
<u>VC</u>				
% <i>Uncertain</i>	1.42%	1.34%	1.34%	1.32%
% <i>Weak modal</i>	0.78%	0.69%	0.67%	0.66%
% <i>Negative</i>	1.57%	1.50%	1.46%	1.39%
	N=357	N=328	N=214	N=63
<u>Non-VC</u>				
% <i>Uncertain</i>	1.39%	1.26%	1.25%	1.23%
% <i>Weak modal</i>	0.71%	0.57%	0.55%	0.52%
% <i>Negative</i>	1.41%	1.28%	1.23%	1.30%
	N=113	N=146	N=258	N=408
<u>All IPOs</u>				
% <i>Uncertain</i>	1.41%	1.31%	1.29%	1.24%
% <i>Weak modal</i>	0.76%	0.66%	0.61%	0.54%
% <i>Negative</i>	1.53%	1.43%	1.34%	1.31%
	N=470	N=474	N=472	N=471
Panel B: Percentage uncertain, weak modal, and negative words by venture capital-backing and trailing EPS				
Item	Low trailing EPS	Middle trailing EPS	High trailing EPS	
<u>VC</u>				
% <i>Uncertain</i>	1.40%	1.33%	1.32%	
% <i>Weak modal</i>	0.76%	0.68%	0.66%	
% <i>Negative</i>	1.57%	1.48%	1.39%	
	N=493	N=317	N=152	
<u>Non-VC</u>				
% <i>Uncertain</i>	1.30%	1.27%	1.24%	
% <i>Weak modal</i>	0.61%	0.57%	0.54%	
% <i>Negative</i>	1.45%	1.33%	1.23%	
	N=137	N=316	N=472	
<u>All IPOs</u>				
% <i>Uncertain</i>	1.38%	1.30%	1.26%	
% <i>Weak modal</i>	0.73%	0.63%	0.57%	
% <i>Negative</i>	1.54%	1.40%	1.27%	
	N=630	N=633	N=624	

to mean effects and not just uncertainty. For example, the S-1 is created after a period of internal information gathering by managers and hired participants. An overabundance of negative S-1 text could lead to an IPO being negatively received during the information gathering process, thus making a subsequent withdrawal more likely (i.e., low mean effect).

We create our withdrawn sample from ipo.Nasdaq.com, which lists all withdrawn IPOs. We also require that the withdrawn IPO have filed a Form RW (registration withdrawal request) on EDGAR and did not subsequently issue a Form 424. Thus, we remove from the sample withdrawn firms that successfully went public at a later date. This data should offer an interesting contrast between completed IPOs (successful ones) and withdrawn IPOs (unsuccessful ones).

Table 11 reports the mean/median word frequencies and number of S-1 words for 1,887 completed IPOs (Panel A) and 793 withdrawn IPOs (Panel B). During 1996–2010, two stock market crashes and a substantial slowdown in IPO volume lead to a higher relative percentage of withdrawn

IPOs (30%) than in some previous papers.<sup>11</sup> During our sample period, two calendar years (2000 and 2008) actually had as many or even more S-1 filings by withdrawn IPOs than by completed IPOs. The best known withdrawn firm in the sample is due to its subsequent high profile bankruptcy: Solyndra, the failed solar panel manufacturer which notably received a \$535 million federal loan.

The mean and median word frequencies are generally only slightly higher for withdrawn IPOs than for the completed IPOs. For example, the mean uncertain, weak modal, and negative frequencies are, respectively, 1.34%, 0.66%, and 1.43% for withdrawn IPOs compared to 1.31%, 0.64%, and 1.41% for completed IPOs. Average positive frequencies are the same across the groups while withdrawn IPOs contain more S-1 words. The lack of a

<sup>11</sup> Busaba, Benveniste, and Guo (2001) report a withdrawn to all S-1 filings ratio of 21% while Edelen and Kadlec (2005) have 742 withdrawn IPOs compared to 4,605 IPO filings (a 16% ratio).

**Table 11**

Summary statistics and logit regressions for completed and withdrawn IPOs.

The table includes 2,680 US IPO S-1 filings of which 1,887 were completed while 793 were withdrawn. Financial firms, ADRs, REITs, and closed-end funds are excluded from the sample. Withdrawn IPOs are obtained from ipo.Nasdaq.com, must have filed a Form RW (registration withdrawn) on EDGAR, and must not have subsequently filed a Form 424. Form S-1 is the initial EDGAR filing for registering the IPO offering with the Securities and Exchange Commission (SEC). The Loughran and McDonald (2011) word lists are used to classify words into uncertain, weak modal, negative, positive, legal, and strong modal categories. In Panel C, a separate logit regression is run for each of the word lists. The dependent variable, *Withdrawn IPO dummy*, is set to one if the IPO filed a Form RW and never filed a Form 424, else zero. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The z-statistics are in parentheses with the standard errors clustered by year and industry. All logit regressions include 2,680 IPO S-1 observations during 1996–2010.

*Panel A: Completed IPOs (N=1,887)*

	% Uncertain	% Weak modal	% Negative	% Positive	% Legal	% Strong modal	S-1 words
Mean	1.31	0.64	1.41	0.97	0.75	0.53	47,827
Median	1.32	0.64	1.42	0.97	0.71	0.52	42,027
Std. dev.	0.19	0.17	0.30	0.17	0.20	0.11	19,272

*Panel B: Withdrawn IPOs (N=793)*

	% Uncertain	% Weak modal	% Negative	% Positive	% Legal	% Strong modal	S-1 words
Mean	1.34	0.66	1.43	0.97	0.77	0.54	51,312
Median	1.34	0.67	1.46	0.96	0.73	0.52	44,098
Std. dev.	0.19	0.18	0.28	0.16	0.24	0.13	22,728

*Panel C: Six separate logit regressions with withdrawn IPO dummy as the dependent variable*

Independent variables	% Uncertain	% Weak modal	% Negative	% Positive	% Legal	% Strong modal
	1.22 (1.90)	1.28 (1.66)	0.59 (1.17)	−0.74 (−1.56)	0.15 (0.37)	1.47 (7.95)
FF dummies	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes

material difference between the groups is consistent with evidence from Busaba, Benveniste, and Guo (2001) and Edelen and Kadlec (2005) who find that completed and withdrawn IPOs have similar characteristics such as underwriter quality, filing amount, profitability, and sales.

To examine whether the word frequencies differ significantly between the samples, Panel C reports six different logit regressions with *Withdrawn IPO dummy* as the dependent variable. As before, we include calendar year and Fama-French industry dummies. The coefficients on both % *Uncertain* and % *Weak modal* are significant only at the 10% level. At this level of significance, higher levels of uncertain or weak modal words are associated with higher chances of the IPO being withdrawn. Notice that the coefficient on % *Negative* has a positive sign, but is not significant. The last column of Panel C reports that % *Strong modal* has a positive coefficient with a z-statistic of 7.95. A possible explanation for the effect of strong modal words is that firms with overconfident language (e.g., *always*, *best*, and *unparalleled*) may be seen, from an investor's perspective, as overselling their potential. In review, withdrawn IPOs do not have notably higher levels of uncertain/negative words in their S-1 filings. Yet, higher uncertain and weak modal word frequencies in the S-1 do marginally raise the probability of the IPO being withdrawn.

### 3.9. Principal component analysis of the three uncertainty word lists

From the various word lists of Loughran and McDonald (2011), the three word lists we find related to IPO underpricing—uncertain, weak modal, and negative—are

consistent with theories based on uncertainty. That these measures also are significantly related to subsequent price volatility further supports their use as an ex ante measure of uncertainty. In our empirical analysis, we examine the word lists separately because of their substantial overlap. Of the 291 uncertain words, 40 overlap with the negative list. All of the weak modal words are also uncertain words and, from Table 4, note that the two most frequently occurring uncertain words are also weak modal words. An important question arising from this overlap is whether these word lists are measuring characteristics unique to their targeted sentiment or whether they all proxy for the same attribute.

To address this question, we extract the first and second principal components from the uncertain, weak modal, and negative word lists (principal components are obtained from a 3 by 1,887 matrix, where 1,887 is the number of IPOs). The extracted principal components provide a sense of the dimensionality of the three lists. A key output from principal component analysis is the extent to which each IPO loads on the first and second components. If only the first principal component is significant in regressions on first-day returns or post-IPO return volatility, then the three word lists are most likely capturing the same attribute. That is, uncertain, weak modal, and negative word lists would proxy for the same effect.

In column 1 of Table 12, the dependent variable is first-day returns while the second column has post-IPO return volatility as the dependent variable. We use the same IPO control variables, calendar year dummies, and industry dummies as in previous regressions. The key independent

**Table 12**

Regressions with *first-day return* and *post-IPO return volatility* as the dependent variables with the first and second principal components as explanatory variables, 1997–2010.

In column 1, the dependent variable, *First-day return*, is the percentage change from the offer price to the closing price. In column 2, the left-hand side variable, *Post-IPO return volatility*, is the market model root-mean square error for each IPO over day +5 to day +64 relative to their IPO date. The Loughran and McDonald (2011) word lists are used to classify S-1 words into uncertain, weak modal, and negative categories. The first and second principal components are extracted from the three uncertainty word lists. See Appendix for control variable definitions. All regressions also include an intercept, Fama and French (1997) 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with the standard errors clustered by year and industry. All regressions include 1,887 US IPO observations during 1997–2010.

Variables	Dependent variables	
	First-day return (1)	Post-IPO return volatility (2)
<i>Log(days between S-1 and 424)</i>	−4.80 (−2.17)	−0.19 (−1.19)
<i>First principal component</i>	3.42 (3.33)	0.28 (4.44)
<i>Second principal component</i>	1.69 (1.10)	0.11 (1.32)
<b>Control variables</b>		
<i>Up revision</i>	1.68 (19.79)	0.01 (1.31)
<i>VC dummy</i>	1.26 (0.47)	0.06 (0.28)
<i>Top-tier dummy</i>	2.05 (1.00)	0.06 (0.33)
<i>Positive EPS dummy</i>	3.91 (2.81)	−0.29 (−2.43)
<i>Prior Nasdaq 15-day returns</i>	0.96 (3.23)	−0.00 (−0.31)
<i>Share overhang</i>	3.99 (2.75)	0.03 (1.22)
<i>Log(sales)</i>	−0.65 (−0.68)	−0.24 (−7.12)
<i>R<sup>2</sup> value</i>	56.3%	47.1%

variables are the first and second principal components. In both regressions, the *First principal component* is positive and statistically significant, while the coefficient on the *Second principal component* is insignificant. This is consistent with the premise that uncertain, weak modal, and negative word lists are one dimensional in nature.<sup>12</sup>

Along with these regressions, we find that the first principal component generally has much higher correlations with firm characteristics than the second component. For example, the first component has higher correlations with first-day returns (0.164 versus 0.038), absolute value of offer price revisions (0.200 versus 0.058), and VC dummy (0.411 versus 0.077) than the second principal component.

<sup>12</sup> If *Withdrawn IPO dummy* is the dependent variable (in the presence of the first and second principal components, calendar, and industry dummies) in a logit regression, the *First principal component* is positive and significant (*z*-statistic of 2.48) while the *Second principal component* is insignificant (*z*-statistic of −0.10).

## 4. Conclusion

We find that more ex ante uncertainty about an IPO's valuation, as proxied by high uncertain, weak modal, and negative word frequencies, produces higher first-day returns and larger aftermarket volatility for a large sample of IPOs during 1997–2010. Although our evidence does not rule out other theories—e.g., information asymmetries, lawsuit avoidance, analyst lust, allocation of shares (see Ritter and Welch, 2002)—our results are consistent with Beatty and Ritter's (1986) theory of ex ante uncertainty and indicate that a nontrivial portion of underpricing can be attributed to this effect. Our word lists provide better measures of uncertainty because they are direct, ex ante, firm-specific, measured well before the IPO, and are linked with ex post stock return volatility.

In our paper, we also relate S-1 tone to two other explanations of first-day returns: prospect theory and bookbuilding. Consistent with prospect theory, we find that higher weak modal/negative S-1 word frequencies (possibly reflecting a weak informational position by issuers) are linked with upward offer price revisions. We report that IPOs with higher weak modal or negative word frequencies tend to have higher absolute revisions in their offer prices, which is consistent with the information production hypothesis.

We also find that changes in tone between the S-1 and 424 have limited impact on upward revisions in the offer price. It appears that the initial uncertainty contained in the S-1 is long-lived and not substantively changed by the time of the Form 424 filing. This is consistent with the notion that once the information providers have been attracted by the uncertain text, managers (possibly worried about litigation risk) elect not to alter the prospectus tone.

In sum, one of the more appealing conceptual frameworks for IPO underpricing emphasizes the role of ex ante uncertainty. Lacking direct proxies for this measure, the uncertainty framework has received less attention in the evolving literature on IPOs. We provide a means of measuring ex ante uncertainty that is validated by its significant relations with ex post observations. Our principal component analysis and tests where the word lists are combined suggest that an aggregate measure based on the uncertain, weak modal, and negative word lists of Loughran and McDonald (2011) provides a useful proxy for uncertainty in the context of the IPO offering.

## Appendix. Definitions of the variables used in the paper

<i>First-day returns</i>	Defined as the percentage change from the offer price to the closing price.
<i>Words in S-1</i>	Total number of words contained in the S-1 (initial IPO filing on EDGAR). A token is considered a word if it appears in our Master Dictionary (available at <a href="http://www.nd.edu/~mcdonald/Word_Lists.html">http://www.nd.edu/~mcdonald/Word_Lists.html</a> ).
<i>Words in 424</i>	Total number of words contained in the 424 (IPO prospectus).
<i>% Uncertain</i>	Percentage of words within the S-1 or 424 that are classified as uncertain using the Loughran and McDonald (2011) word list. Examples



	include <i>believe</i> , <i>pending</i> , <i>approximate</i> , <i>uncertain</i> , and <i>uncertainty</i> .
% Weak modal	Percentage of weak modal words in the S-1 or 424 using the Loughran and McDonald (2011) word list. Examples include <i>could</i> , <i>might</i> , <i>nearly</i> , <i>maybe</i> , and <i>possibly</i> .
% Negative	Percentage of words within the S-1 or 424 that are classified as negative using the Loughran and McDonald (2011) word list. Examples of negative words include <i>loss</i> , <i>failure</i> , <i>decline</i> , <i>bankruptcy</i> , and <i>difficult</i> .
% Positive	Percentage of words within the S-1 or 424 that are classified as positive using the Loughran and McDonald (2011) word list. Examples of positive words include <i>beneficial</i> , <i>successful</i> , <i>good</i> , and <i>strong</i> .
% Legal	Percentage of words within the S-1 or 424 that are classified as legal using the Loughran and McDonald (2011) word list. Examples of legal words include <i>regulations</i> , <i>contracts</i> , <i>settlement</i> , <i>plaintiffs</i> , and <i>litigation</i> .
% Strong modal	Percentage of strong modal words in the S-1 or 424 using the Loughran and McDonald (2011) word list. Examples include <i>must</i> , <i>never</i> , and <i>will</i> .
Change in % uncertain	Defined as the % <i>Uncertain</i> in the Form 424 minus the % <i>Uncertain</i> in the S-1 filing.
Change in % weak modal	Defined as the % <i>Weak modal</i> in the Form 424 minus the % <i>Weak modal</i> in the S-1 filing.
Change in % negative	Defined as the % <i>Negative</i> in the Form 424 minus the % <i>Negative</i> in the S-1 filing.
Change in % positive	Defined as the % <i>Positive</i> in the Form 424 minus the % <i>Positive</i> in the S-1 filing.
Change in % legal	Defined as the % <i>Legal</i> in the Form 424 minus the % <i>Legal</i> in the S-1 filing.
Change in % strong modal	Defined as the % <i>Strong modal</i> in the Form 424 minus the % <i>Strong modal</i> in the S-1 filing.
Days between S-1 and 424	The number of calendar days between the S-1 filing and the filing of the Form 424 on EDGAR.
Up revision	Percentage upward revision in the offer price from the mid-point of the filing range if the offer price is greater than the mid-point, ((offer price-mid-point)/mid-point) × 100 if offer price > mid-point, else zero.
Absolute (revision)	Absolute value of the percentage revision in the offer price from the mid-point of the filing range.
VC dummy	Dummy variable set to one if IPO is backed by venture capital, else zero.
Top-tier dummy	Dummy variable set to one if the lead underwriter of the IPO has an updated Carter and Manaster (1990) rank of eight or more, else zero.
Positive EPS dummy	Dummy variable set to one if trailing EPS is positive at the time of the IPO, else zero.
Prior Nasdaq 15-day returns	The buy-and-hold returns of the CRSP Nasdaq value-weighted index on the 15-trading days prior to the IPO date, ending on day $t-1$ .
Share overhang	Defined as the number of shares retained divided by the number of shares in the initial offering.
Sales	Trailing annual firm sales in millions of dollars at the time of the IPO.
Post-IPO return volatility	The market model root-mean square error for each IPO over day +5 to day +64 relative to their IPO date. The value is multiplied by 1,000.
Withdrawn IPO dummy	Dummy variable set to one if the IPO (listed on ipo.Nasdaq.com) filed a Form RW (registration withdrawn) on EDGAR and never subsequently filed a Form 424, else zero.

## References

- Beatty, R., Ritter, J., 1986. Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics* 15, 213–232.
- Benveniste, L., Spindt, P., 1989. How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics* 24, 343–362.
- Benveniste, L., Wilhelm, W., 1990. A comparative analysis of IPO proceeds under alternative regulatory environments. *Journal of Financial Economics* 28, 173–208.
- Bradley, D., Jordan, B., 2002. Partial adjustment to public information and IPO underpricing. *Journal of Financial and Quantitative Analysis* 37, 595–616.
- Busaba, W., Benveniste, L., Guo, R., 2001. The option to withdraw IPOs during the premarket: empirical analysis. *Journal of Financial Economics* 60, 73–102.
- Carter, R., Dark, F., Singh, A., 1998. Underwriter reputation, initial returns, and the long-run performance of IPO stocks. *Journal of Finance* 53, 285–311.
- Carter, R., Manaster, S., 1990. Initial public offerings and underwriter reputation. *Journal of Finance* 45, 1045–1068.
- Chen, H., De, P., Hu, Y., Hwang, B., 2011. Customers as advisors: the role of social media in financial markets. Unpublished working paper. City University of Hong Kong and Purdue University.
- Clayton, M., Hartzell, J., Rosenberg, J., 2005. The impact of CEO turnover on equity volatility. *Journal of Business* 78, 1779–1808.
- Doidge, C., Karolyi, A., Stulz, R., 2012. Financial globalization and the rise of IPOs outside the U.S. Unpublished working paper. University of Toronto, Cornell University, and Ohio State University.
- Dougal, C., Engelberg, J., Garcia, D., Parsons, C., 2012. Journalists and the stock market. *Review of Financial Studies* 25, 639–679.
- Edelen, R., Kadlec, G., 2005. Issuer surplus and the partial adjustment of IPO prices to public information. *Journal of Financial Economics* 77, 347–373.
- Epstein, L., Schneider, M., 2008. Ambiguity, information quality, and asset pricing. *Journal of Finance* 63, 197–228.
- Fama, E., French, K., 1997. Industry costs of equity. *Journal of Financial Economics* 43, 153–193.
- Feldman, R., Govindaraj, S., Livnat, J., Segal, B., 2010. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* 15, 915–953.
- Ferris, S., Hao, Q., Liao, M., 2012. The effect of issuer conservatism on IPO pricing and performance. *Review of Finance*, forthcoming.
- Field, L., Karpoff, J., 2002. Takeover defenses of IPO firms. *Journal of Finance* 57, 1857–1889.
- Field, L., Lowry, M., Shu, S., 2005. Does disclosure deter or trigger litigation? *Journal of Accounting and Economics* 39, 487–507.
- Gao, X., Ritter, J., Zhu, Z., 2011. Where have all the IPOs gone? Unpublished working paper. University of Hong Kong, University of Florida, Chinese University of Hong Kong.
- Gompers, P., Lerner, J., 2001. *The Money of Invention: How Venture Capital Creates New Wealth*. Harvard Business School Press, Boston.
- Gurun, U., Butler, A., 2012. Don't believe the hype: local media slant, local advertising, and firm value. *Journal of Finance* 67, 561–597.
- Hanley, K., 1993. The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics* 34, 231–250.
- Hanley, K., Hoberg, G., 2012. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics* 103, 235–254.
- Hanley, K., Hoberg, G., 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23, 2821–2864.
- Hoberg, G., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. *Review of Financial Studies* 23, 3773–3811.
- Jegadeesh, N., Wu, A., 2011. Word power: a new approach for content analysis. Unpublished working paper. Emory University and University of Pennsylvania.
- Jenkinson, T., Ljungqvist, A., 2001. *Going Public: The Theory and Evidence on How Companies Raise Equity Finance*. Oxford University Press, Oxford.
- Kothari, S., Li, X., Short, J., 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. *Accounting Review* 84, 1639–1670.
- Larcker, D., Zakolyukina, A., 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50, 495–540.
- Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 221–247.
- Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66, 35–65.
- Loughran, T., Ritter, J., 2002. Why don't issuers get upset about leaving money on the table? *Review of Financial Studies* 15, 413–443.

- Loughran, T., Ritter, J., 2004. Why has IPO underpricing changed over time? *Financial Management* 33, 5–37.
- Lowry, M., Schwert, W., 2002. IPO market cycles: bubbles or sequential learning? *Journal of Finance* 57, 1171–1200.
- Lowry, M., Schwert, W., 2004. Is the IPO pricing process efficient? *Journal of Financial Economics* 71, 3–26.
- Lowry, M., Shu, S., 2002. Litigation risk and IPO underpricing. *Journal of Financial Economics* 65, 309–355.
- Mayew, W., Venkatachalam, M., 2012. The power of voice: managerial affective states and future firm performance. *Journal of Finance* 67, 1–44.
- Ritter, J., 1984. The ‘hot issue’ market of 1980. *Journal of Business* 57, 215–240.
- Ritter, J., Welch, I., 2002. A review of IPO activity, pricing, and allocations. *Journal of Finance* 57, 1795–1828.
- Rock, K., 1986. Why new issues are underpriced. *Journal of Financial Economics* 15, 187–212.
- Shabani, R., 2011. Corporate news, asset prices, and the media. Unpublished working paper. University of California, Berkeley.
- Sherman, A., Titman, S., 2002. Building the IPO order book: underpricing and participation limits with costly information. *Journal of Financial Economics* 65, 3–29.
- Tetlock, P., 2007. Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance* 62, 1139–1168.
- Veronesi, P., 1999. Stock market overreaction to bad news in good times: a rational expectations equilibrium model. *Review of Financial Studies* 12, 975–1007.