Does the Confidential IPO Registration Process Create Value?

December 28, 2021

Abstract

This paper studies the confidential IPO registration process adopted by 86% of firms since the JOBS Act of 2012. Using textual analysis, I am the first to investigate the changes in information content between draft registration statements (Form DRS) and corresponding formal prospectus (Form S-1). I show that valuable information is produced during the confidential review process between the firm and SEC. I construct a novel proxy for the content of SEC comment letters before their release, which leads to a 4.5% cumulative abnormal return. I find that changes in the proportion of positive words are strongly associated with operating performance, reflecting the fundamental value of offerings. Moreover, changes in word content indicate changes in firm's information environment that can further explain offer price adjustment and underpricing. Lastly, the evidence suggests that the confidential review process significantly improved disclosure transparency, while SEC has remained neutral in the firm's withdrawal decision.

1   Introduction

Firms go public to meet their financial needs and broaden opportunities for future access to the capital market. However, the U.S. initial public offering (IPO) market has experienced a sharp decline in number of IPOs since the burst of dot-com bubble.[1] As a regulator's attempt to revive the IPO market, the Jumpstart Our Business Startups Act (JOBS Act) was signed into law in 2012, which substantially changed the registration process of going public. Typically, for any U.S. firm that undertakes an IPO, a registration statement (Form S-1) must be publicly filed with the Securities and Exchange Commission (SEC).[2] Under the JOBS Act of 2012, IPO firms are allowed to initiate a confidential review process with SEC by submitting a draft registration statement (Form DRS) before S-1 is publicly filed. Once the DRS is submitted, SEC provides feedback and raises concerns in private comment letters to the firm. Then, the firm revises and resubmits the draft through subsequent amendments for further review. Essentially, the evolution of DRS to S-1, or the revise-and-resubmit (R&R) process, is a confidential information production between IPO firm and SEC, which on average, repeats 2.6 rounds and takes 108 days until S-1 is publicly filed. In 2013-2016, more than 76% of completed IPOs adopted this new approach. After the expansion of the JOBS Act, 96% of IPO firms confidentially filed for IPO during 2017-2020, making it a new standard for going public.[3]

Despite of the wide adoption, there is no notable boost in number of IPOs since the JOBS Act. Therefore, understanding the financial economics of the confidential review process can address concerns on potential capital market dysfunction. One critical question is that whether valuable information is produced by SEC during the confidential review process. Specifically, whether the changes in firm's information environment have implications with regard to the pricing strategy of the offering, the underpricing phenomenon, firm's fundamental risk, the rationale of withdrawal decision, as well as the role of the regulator in the public market. Additionally, whether the outsiders (i.e., investors) can infer SEC's concerns on the offering before the IPO, given the fact that SEC comment letters are not

---

[1]See Gao, Ritter, and Zhu (2013) and Doidge, Kahle, Karolyi, and Stulz (2018).

[2]Form S-1, or investment prospectus, was introduced in Securities Act of 1933.

[3]The expansion in 2017 allows all firms to confidentially filed for IPO, rather than small size firms.

released at least 20 days after a completed IPO. Using textual analysis, this study is the first to investigate the evolution of information content from DRS to its formal prospectus (i.e., S-1) with the focus on both the changes in the words over time and the context in which the words are changed. I show that valuable information is produced during the confidential review process between the firm and the SEC. I construct a novel ex-ante proxy for the content of SEC comments, enabling investors to make informed investment decisions before the IPO and obtain a 4.5% cumulative abnormal return. I find changes in the proportion of positive words are strongly associated with operating performance, reflecting the fundamental value of offerings. Moreover, changes in word content indicate changes in firm's risk environment that can further explain offer price adjustment and underpricing. Lastly, the evidence from IPO withdrawal decision analysis shows that SEC has maintained a neutral role in determining firm's access to the public capital market.

There are considerable theoretical and empirical studies focusing on the firm's decision to go public (see Lerner, 1994; Chemmanur and Fulghieri, 1999; Maksimovic and Pichler, 2001; Chemmanur, He, and Nandy, 2010) and on the IPO pricing strategy (see Rock, 1986; Chemmanur, 1993; Beatty and Ritter, 1986; Benveniste and Spindt, 1989; Benveniste and Busaba, 1997; Loughran and Ritter, 2002). A few prominent studies examine the information content of the S-1 using textual analysis. For example, Hanley (1993) focuses on the proposed offer price range in the S-1. Hanley and Hoberg (2010, 2012) compare S-1 amendments. Loughran and McDonald (2013) link the tone of the S-1 to offer price adjustment, first-day returns and post-IPO return volatility. My analysis follows this strand of the literature and builds on these studies by performing textual analysis on the confidential review process. Although 86% of IPO firms have filed a DRS since the JOBS Act, so far no empirical research has focused on the information content of the DRS and the confidential review process defined in this study. I present the first analysis of the DRS and its evolution to the S-1, and I document the relationships between changes in content during the confidential review process and the firm's response to SEC comments, the offer price adjustment, first-day returns, operating performance, IPO withdrawal decision, as well as the overall impacts on the IPO market.

The objective of this study is to bridge the gap in the literature by addressing the critical question that has received minimal attention thus far: Is there any valuable information produced during the confidential review process between the firm and SEC? Addressing several key focuses in the IPO literature, this study uses five categories of Loughran-McDonald (LM) word lists to investigate the information content of the DRS and the effect of changes in content during the confidential review process on the firm's information environment. In addition, the study takes into account three major sections of the prospectus to consider the context of the words changed and to track changes by section over time. With a sample of 780 completed IPOs that went through the confidential review process during 2013-2020, this study links the changes in content of the prospectus, measured by content similarity and five categories of LM word lists, with offer price adjustment, first-day returns, pre- and post-IPO operating performance, the decision of IPO withdrawal, as well as SEC comment letters. In the robustness checks, the same set of analyses are conducted using the level of the LM word lists in the first confidential version of the prospectus (DRS) and the first public version of the prospectus (S-1), respectively. For each group of analyses, tests are also performed using specific sections of the prospectus: Risk Factors, Business, Management's Discussion and Analysis (MD&A) and the entire document (i.e., all sections in the main body, excluding pictures, attachments, exhibits, and XML content).

The six main findings of this study can be summarized as follows. First, the new evidence in the event study on the release of SEC comment letters shows that the changes in textual analysis measures during the confidential review process are an ideal ex-ante proxy for the content of comment letters on average, 51 days before the IPO, or 85 days before the release of comment letters. When grouping IPO firms based on percentage changes in the proportion of positive words (%Positive) in the Business section, this study finds different market reactions to the release of comment letters between the above-median and below-median group. Specifically, 20 days after the release, the differences in cumulative abnormal returns (CAR) between the above-median and below-median group are widened to 4.5%. Regression results show that changes in %Positive is strongly associated with CAR across 3-day, 10-day, and 20-day windows. Additional robustness tests show that the

results are consistent when using the negative word list or under several different grouping strategies. Therefore, the novel proxy proposed in this study is a valid approximation for SEC concerns, which enables investors to make informed investment decisions on IPOs, although the comment letters are not observable at the moment.

Second, the percentage changes in the proportion of negative words (%Negative) in the Risk Factor section during the confidential review process are negatively associated with offer price adjustment. Specifically, after controlling for known IPO-specific determinants, a one standard deviation increase of %Negative in the Risk Factor section (+7.7%) implies a 1% downward adjustment in the offer price relative to the midpoint of the filed price range. The fact that this finding is more prominent using the Risk Factor section than using the entire document suggests that the additional risks revealed by the SEC's scrutiny are captured mainly by the changes in the proportion of negative words in Risk Factor section. In addition, using the level of %Negative in the DRS or S-1, similar results are found supporting the arguments in Loughran and McDonald (2013).[4]

Third, the overall changes in word content during the confidential review process, measured by cosine similarity, is found to be strongly and positively correlated with first-day returns. It shows that a one standard deviation increase in the cosine similarity score (+4.5) leads to a 2% increase of first-day returns. Conditional on offer price adjustment, higher document similarity implies that fewer information is produced, and thus there is a higher probability that the firm is omitting the information that caused the price adjustment. To hedge against this increased liability risk due to omission, the firm chooses to underprice more in its offering. This finding supports the "Substitution Effect" explained in the empirical work of Hanley and Hoberg (2012). The interpretation is also in line with the costly information production theory of Chemmanur and Fulghieri (1999) and the confidential information release theory of Bhattacharya and Ritter (1983) and Maksimovic and Pichler (2001), suggesting the confidential revise-and-resubmit process should be viewed as regulation-driven information production that reduces the uncertainty in firm valuation.

Fourth, the analysis of operating performance documents the channel through which textual analysis measures capture the fundamental risks of the firm prior to its IPO. The

---

[4]See the Appendix for additional results.

4

findings show that the percentage changes in %Positive in the Business section during the confidential review process is a significant predictor of the real difference in post-IPO operating performance. For example, a one standard deviation increase of %Positive (+8.8%) in the Business section is associated with a 1.3% increase in asset-adjusted net income in 12 months since the IPO. In addition, a strong trend, almost monotonic rise in historical Sales, EPS, EBIT, and EBITDA is observed across %Positive quintiles using either DRS or S-1 filing. The findings suggest that it is the real effects that drives the return predictability of textual analysis measures.

Fifth, using a sample of firms that completed the confidential review process but eventually withdrew their IPO, this study further shows that firms with a high level of %Litigious and %Strong-modal words, and a low level of %Negative in their prospectus have a higher probability of withdrawing their IPO. Having a large number of legal words and using definitive language (e.g., always, definitely, never), along with a more optimistic tone, indicates an overconfident management of the IPO firm. This managerial hubris contributes to a subsequent withdrawal. Moreover, the evidence regarding the changes in the MD&A section during the confidential review process implies that the consistency in management's statement is negatively associated with the decision to withdraw an IPO.

Lastly, by comparing the changes in each firm's prospectus before and after the confidential review process, I find that the transparency and quality of disclosure has been notably improved, while SEC has stayed true to its neutral role in monitoring the public capital market. Specifically, there are 10% increases in the word content from DRS to S-1 after the confidential review process; meanwhile, the complexity of word content, measured by the average syllables per word, has been reduced. On the other hand, the changes made through the confidential review process are found insignificantly associated with the firm's withdrawal decision, suggesting there is no direct intervention by the regulator in determining the firm's access to the public capital. In terms of factors lead to an IPO withdrawal, it is not the SEC's scrutiny, but the intrinsic level of riskiness determines the value of the offering.

This study contributes to several streams of literature. First, it provides new empiri-

cal evidence for the IPO literature. In an asymmetric information-based setting between informed and uninformed investors including the firm itself, Rock (1986) suggests that a discount in offer price is necessary to solicit demand from uninformed investors due to "Winner's Curse". Chemmanur (1993) explains the IPO underpricing in an information production model in which underpricing in equilibrium is the compensation for outsiders who engage in the costly information production. Empirically, Beatty and Ritter (1986) show that the uncertainty about the firm's valuation, approximated by the uses and the amount of proceeds disclosed in the IPO prospectus, is positively related to first-day returns of the IPO. Chemmanur and Fulghieri (1999) suggest that the degree of information asymmetry, approximated by firm age and size, and industry characteristics affect the probability of going public, as the uninformed outsiders can evaluate the firm as a cost. Maksimovic and Pichler (2001) imply that the value of confidential information disclosed during the IPO process matters, especially among innovative firms (e.g. high-tech firms). Chemmanur, He, and Nandy (2010) examine the effect of firms' ex-ante product market characteristics on the going-public decision. This study provides new evidence that supports those findings by focusing on the DRS and the confidential review process. The results are robust in the settings in which content from the entire prospectus and from specific sections is used. In addition, the relationships are consistent whether using the draft version (DRS) or the formal version (S-1) of the prospectus.

This study extends and complements the literature that focuses on the information content of the IPO prospectus using textual analysis. Hanley (1993) documents a "partial adjustment" phenomenon on offer price with a focus on the proposed price range in S-1. Hanley and Hoberg (2010) decompose information in S-1 into standard and informative components. In their later work, Hanley and Hoberg (2012) compare S-1 amendments and find that issuers underprice their IPOs to hedge against the subsequent litigation risk. This study is likely closest in spirit to that of Loughran and McDonald (2013), who link the tone of S-1 to offer price adjustment, first-day returns and post-IPO return volatility. A common critique of textual analysis in finance is that the simple frequency of words used in a document may not reflect the writer's true intention. This study addresses

that concern in three dimensions: 1) Five word lists from Loughran and McDonald (2011) and Bodnaruk, Loughran, and McDonald (2015) are used to measure the proportion of specific categories (i.e., %Positive, %Negative, %Litigious, %Constraining, and %Strong-modal). All of these word lists are developed from firms' annual reports to approximate the tones in the disclosure, and thus are more suitable than other measures for analyzing IPO prospectuses.[5] 2) To consider the context of the words, textual analysis is conducted using the content of the entire prospectus, as well as the content of each major section. Since the objective of each section is different in the prospectus, focusing on one particular section at a time can produce a more accurate estimation of the meaning of the words. For example, measuring uncertainty using only the Risk Factors section may pick up subtle details that would otherwise be overlooked in the full prospectus, which typically contains more than 100,000 words. 3) Focusing on the changes in the content over time, the analyses are performed by comparing the S-1 to DRS in a word-list-by-word-list and section-by-section fashion. In this way, deviations from the original filing (the DRS), including both additions and/or deletions, can be better captured and tested, even though the changes are as small as a few words. This approach can largely avoid the situation in which the tone of the content is so ambiguous that a mixed or extraordinary relationship is observed.

This study also relates to the literature on the role of regulator in public capital market. Simon (1989) finds that the dispersion of abnormal returns was significantly lower following the Securities Act of 1933. Mahoney (1995) finds that external monitoring is necessary to prevent an upward-biased disclosure by issuers. In more-recent work, Dechow, Lawrence, and Ryans (2016) and Ryans (2020) examine SEC comment letters on firms' annual reports (10-K). Lowry, Michaely, and Volkova (2020) also analyze the communications between the SEC and the firms through comment letters using machine learning techniques. Although the SEC's feedback can be valuable and direct, its comment letters are not released until at least 20 business days after the IPO. This study takes advantage of the fact that both the DRS and S-1 are publicly available at the time of S-1 submission (prior to the IPO), regardless of the final outcome of the IPO (i.e., whether or not it is eventually withdrawn

---

[5]See Section 3 for a detailed discussion of textual analysis measures.

or goes public as planned).[6], and documents that the change in textual analysis measures during the confidential review process is an ideal ex-ante proxy for the content of comment letters prior to its release and before IPO. Additionally, this study shows SEC has remained independent and neutral in the firm's withdrawal decision.

Finally, this study makes contributions to textual analysis techniques by proposing an efficient method for extracting specific sections of structured financial documents in batches automatically. There are several practical issues that lead to inaccurate outcomes when extracting specific sections of firm disclosures, even though they are structured formal documents. As Loughran and McDonald (2016) point out in a comprehensive survey, in addition to misspellings, abbreviations, word representation of numbers, and the use of Roman numerals, errors in parsing may occur due to the inconsistency of the heading, the order, and the numbering of each section. Those errors usually produce misleading results in textual analysis. This study efficiently addresses those issues by taking advantage of several formatting tags in the HTML version of the prospectus, prior to converting to a cleaned full-text version.[7]

The remainder of the paper is organized as follows. Section 2 describes the IPO registration process and explains the confidential review process in detail. Section 3 describes sample construction and reports descriptive statistics. Section 4 discusses the empirical results of textual analysis on several aspects of IPO, including the proxy for SEC comments, offer price adjustment, first-day returns, operating performance, and withdrawal decision. Section 5 focuses on the policy impact and SEC's role in IPO market, and Section 6 concludes.

---

[6]For example, by the time this study is completed, the comment letters for WeWork still have not been released, even though it withdrew its IPO on September 30, 2019. However, its DRS and S-1 have been publicly available on EDGAR since August 14, 2019.

[7]See the Appendix for the detailed steps to extracting specific sections.

2   The IPO Registration Process

2.1   Post-JOBS Act IPO Registration Process

In the post-JOBS Act era, a typical IPO process starts with filing DRS for a confidential review. In return, the SEC will provide feedback in the form of a comment letter, which is also confidential prior to the IPO.[8] To address the concerns raised by SEC, the firm will subsequently file one or more revisions of the DRS. This confidential review-revise-resubmit process repeats on average, for 2.6 rounds with no fee required by SEC and lasts about 108 days until the firm files the S-1 publicly. Upon the submission of the S-1, all previously filed documents, including the DRS and its amendments but excluding the comment letters, are released on SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. At that time, the firm can start reaching out to potential investors by doing a "road show," usually followed by extensive media coverage. In the meantime, a similar but non-confidential review process on the S-1 continues until an Effective Notice is issued by SEC. The firm can go public as soon as the next business day once the Effective Notice is received. Ultimately, the final IPO prospectus, Form 424 or its variants (Form 424B3/424B4), is filed on the day of or within a few days after the IPO. Figure 1 illustrates the entire filing process for the IPO registration.

[Figure 1 here]

Even though almost every IPO firm files a DRS (96% since 2017) in its registration process, the content of the DRS has not generated extensive interest among investors and researchers, while the S-1 content has. One might suppose that the investor's inattention is due to the confidential nature of the DRS. However, this may not be the case because all DRS filings will be released at the same time when the S-1 is publicly filed. Moreover, three unique characteristics have made the DRS and its evolution to S-1 worth looking into. First, the DRS is currently the first document that contains ample amounts of intangible information regarding the private firm that is intending to go public. As with the S-1, the IPO firm is required to disclose a detailed description of its properties, business, security

---

[8]SEC comment letters are released at least 20 days after a completed IPO.

being offered and risks involved, among other matters through the DRS. Since the DRS is the draft version of the S-1, the highly structured content makes it possible to examine the qualitative information by comparing the DRS and S-1, both as whole documents and by specific sections. Second, the confidential revise-and-resubmit process, defined as the confidential review period between DRS submission and S-1 submission, almost perfectly separates the information production of the regulator (i.e., SEC) from that of the investors.[9] Private firms going public in a competitive industry, tend to disclose as little valuable information as possible, conditional on passing the SEC's scrutiny. Therefore, the changes made between the DRS and S-1 should reflect the firm's reaction to SEC comments. Third, since the changes during the confidential review process can be considered as regulation-driven revisions, it provides an opportunity to infer the information content of SEC comment letters prior to their release. Both the DRS and S-1 are publicly available at the time of S-1 submission, which is on average 51 days before the IPO, regardless of the completion of the IPO. Using textual analysis, it is feasible to create a proxy for SEC's concerns and make an informed decision regarding the IPO.

2.2 The Confidential Revise-and-Resubmit Process

[Figure 2 here]

For two reasons, this study focuses on the confidential review process, defined as the period between the submission date of the DRS and that of the S-1. First, it provides a unique opportunity to separate the investor's information production from that of the SEC. Before the public submission of S-1, all the information exchanged between the firm and the SEC is confidential. Therefore, examining the interaction between the firm and the SEC can shed light on the effectiveness of enforcing disclosure transparency and the role of the regulator in the public market. Second, there are material differences between the DRS and S-1, which allows for a comprehensive comparison. As shown in Figure 2, the median number of unique financial words trends up in both the DRS and S-1 after the

---

[9]JOBS Act allows certain firms to "test-the-waters" by reaching out to qualified institutional buyers and institutional accredited investors. However, firms have no incentives to disclose information learned from the demand-side of the market, which is also beyond the scope of SEC's scrutiny.

JOBS Act of 2012, while there are substantial differences between DRS and S-1 filings. For example, on average, more than 150 unique financial words were added during the confidential review process for 2020's IPOs. The purpose of adding those words and the meaning of that additional information have not been studied intensively.

Although other comparisons can be made using amendments or Form 424, none of them necessarily includes enough information to conduct a systematic comparison. Amendments to filings usually do not include the full content that the original filings contain. Each amendment works individually to modify parts of the original filing.[10] As a result, the length of each amendment varies, and thus the information contained is not comparable to a typical 100-page sectionalized DRS or S-1 filing. Similarly, the objective of Form 424 is to complete the information not available at the time that the S-1 was filed, such as the actual offer price and realized proceeds and fees. Since Form 424 is filed after the IPO, changing information in the prospectus exposes the firm to high litigation risk due to the "quiet period" rule.[11] Hence, the new information contained in Form 424 is expected to be limited and immaterial. Naturally, one should expect a very high similarity between the S-1 and Form 424. Loughran and McDonald (2013) find that S-1 and 424 filings have a very high correlation in the tone (greater than 0.9), measured using uncertain, negative, and weak modal word lists. Overall, by concentrating on the confidential revise-and-resubmit process, this paper is different from Hanley and Hoberg (2010, 2012), whose focus is on S-1 and S-1 amendments, and Ferris, Hao, and Liao (2013), who study Form 424, and Loughran and McDonald (2013), who examine S-1 and Form 424.

---

[10]For example, one amendment can be filed just to change the company name, while another amendment updates the latest contact information.

[11]The federal securities laws do not officially define the term "quiet period." In practice, it typically lasts for 40 days after the initial trading.

3 Data and Sample Construction

3.1 Matching DRS and S-1 Filings

The data construction procedure results in a sample of 903 firms that completed IPOs during 2013-2020, with 780 (or 86%) of them going through the confidential review process. The full-text version of DRS and S-1 filings are downloaded by specifying "Form Type" criterion on EDGAR. In order to compare the first draft version (DRS) and the first formal version (S-1) of the prospectus, a valid pair of the documents must be identified at the firm level. This matching process has several challenges. First, during the preparation for going public, a firm may change its ownership structure and name for various reasons, such as for tax purposes. In the original sample of 903 IPOs in 2013-2020, about 15% of the firms changed their names at least once during the registration process. Thus, matching DRS with S-1 simply by firm name may lead to potential errors. Second, there are inconsistencies between the entity filing the documents and the entity going public. One common strategy is to match by the Central Index Key (CIK), a 10-digit unique number that identifies each entity what files with EDGAR.[12] However, in practice, the firm usually hires a third-party entity to complete the submission on its behalf. As a result, there will be an inconsistency between the filer's CIK and the CIK of the actual firm that is going public.[13]

To allow accurate matching and reliable estimating in this study, the IPO firm's actual CIK (instead of the filer's CIK) is extracted from the header of the full-text version of the DRS and S-1, separately. Matching based on the extracted IPO firm's actual CIK, delivers 780 valid pairs for this study in the period of 2013-2020. For all identified pairs, the release date of the DRS is assured to be no later than the next business day after the S-1 submission date.[14] This paired sample is then supplemented with additional data sources for information related to the security offered, the firm fundamentals, and the post-IPO market information. Securities Data Company (SDC) Platinum is the source for IPO issuance data. Firm fundamentals and market information are from Compustat and CRSP,

---

[12]When the filings are downloaded, the CIK of the filer can be found in the name of the *.txt* file.

[13]For example, Uber's CIK is 0001543151; the filer of Uber's DRS has the CIK 0000950123; the filer of Uber's S-1 has the CIK 0001193125.

[14]Since Form S-1 is filed publicly on EDGAR, its submission date is the same as the release date.

respectively. Following the IPO literature, a standard screening process, as described in Gao, Ritter, and Zhu (2013), is applied to the matched sample. Specifically, IPOs with an offer price of less than $5 per share, financial firms (i.e., banks, savings, and loans), American Depository Receipts (ADRs), unit or rights offers, real estate investment trusts (REITs), natural resource limited partnerships and closed-end funds are excluded from the sample. The final sample contains 1560 documents (780 pairs), with all 780 firms going public through the confidential review process during 2013-2020.

In addition to examining the word content of the entire document, this study also focuses on three specific sections of the IPO prospectus: Risk Factors, Business, and Management's Discussion and Analysis (MD&A), which combined, represents about 47% of the entire content, on average, in terms of the total number of words. Those three sections are also the most common sections that investors focus on, as the content covers the downsides and upsides of the firm's and management's viewpoints, respectively. Hence, this study will focus on those three sections wherever a section-specific test is conducted. It is worth noting that when comparing the S-1 and DRS by section, it requires that the target section is available in both documents. Observations that fail to meet the requirements are excluded from those comparisons.

Loughran and McDonald (2016) note several practical issues that lead to inaccurate outcomes when extracting specific sections. My study efficiently addresses those issues by taking advantage of several formatting tags in the HTML version of the prospectus, prior to converting to a full-text version. See the Appendix for the detailed steps to break the entire document into sections.

3.2   Textual Analysis Variables

This study primarily uses five categories of word lists as its measures of tone in the IPO prospectus, while it uses the changes in the proportion of word lists and cosine similarity as the measure of the changes in content over time. In recent years, the implementation of textual analysis in finance has become one of the most important trends in both academia

and industry.[15] This study uses the set of word lists developed in Loughran and McDonald (2011) to measure tone in the DRS and S-1, as well as the change in tone over time. Unlike other measures,[16] the Loughran-McDonald word lists (LM word lists) are constructed specifically using firms' annual reports (i.e., Form 10-K), making the word lists perfectly suitable for gauging sentiment in mandated financial disclosures, such as IPO prospectus in this study.[17] With the addition of the Constraining word list introduced in Bodnaruk, Loughran, and McDonald (2015), there are eight categories of LM word lists available: Negative, Positive, Uncertainty, Litigious, Strong Modal, Weak Modal, Moderate Modal, and Constraining. Each category of word lists serves as a lexicon-based dictionary to measure the frequency of specific group of words that appear in the content of interest. Since the nature of the measure is objective, the frequency of word lists is not affected by potential misinterpretation. However, it is worth noting that there are overlaps among LM word lists. For example, all 27 words in Weak Modal are included in Uncertain, while there is a high correlation between Uncertain and Negative. Moreover, the boundary between Moderate Modal words and Strong Modal or Weak Modal words is subject to the user's discretion. Therefore, consistent with the suggestions in Loughran and McDonald (2011, 2013, 2016), only five categories of word lists are used in the analysis. In robustness checks, an additional Risky word list is constructed as an aggregate measure of uncertainty and negative words. The proportion of each word list used in the prospectus is measured in percentage scaled by the number of words in the content analyzed: %Positive, %Negative, %Uncertainty, %Litigious, %Strong-modal, %Constraining and %Risky. See the Appendix for the definitions and examples of words included in each LM word lists used.

Two measures have been constructed to capture the changes during the confidential review process. The first one is the Percentage Changes defined as the percentage changes in the proportion of each LM word list from DRS filings to the corresponding S-1. The second one is the Cosine Similarity taken from the natural language processing and information

---

[15]See Loughran and McDonald (2020) for an updated review of the literature.

[16]For example, Hu and Liu (2004) create a word list for sentiment analysis in social media. Chen, Despres, Guo, and Renault (2019) create a crypoto-specific lexicon including emojis, slang, and profanity.

[17]I thank Prof. Loughran and Prof. McDonald for generously making the word lists and the accompanying dictionary publicly available at https://sraf.nd.edu.

science literature. In textual analysis, cosine similarity measures the cosine angle between two document vectors on a unit sphere.[18]

Given two documents doc_1 and doc_2 that have been converted to word vectors x and y in terms of word counts, the cosine similarity for each word from $i = 1$ to N, is defined as:

$$CosineSimilarity(doc_1, doc_2) = \frac{\sum_i^N x_i y_i}{\sqrt{\sum_i^N x_i^2} \sqrt{\sum_i^N y_i^2}} \tag{1}$$

Since word counts are strictly non-negative, cosine similarity score is bounded between 0 and 1.[19] Intuitively, the higher the similarity between the two documents, the higher cosine similarity score is. The two constructed measures are complementary to each other. The percentage changes in LM word lists provide a disaggregated view at the word list level, while the cosine similarity score suggests an overall similarity at the document level.

3.3   IPO Control Variables

The first part of this study focuses on the IPO pricing activity, and hence the primary outcome variables of interest include: Offer Price Adjustment, defined as the percentage change in offer price from the midpoint of the filing price range; and First-day Returns, defined as the percentage change from the offer price to the closing price on the first trading day. Multiple papers in the IPO literature have documented the associations between certain variables and IPO pricing activity. Among them are Bradley and Jordan (2002), Loughran and Ritter (2004), Chemmanur, He, and Nandy (2010), Hanley and Hoberg (2012). Similar to Loughran and McDonald (2013), for all analyses conducted in this study, six variables in the traditional IPO literature and two variables related to the confidential review process are included as controls. Specifically, a VC-backed dummy is set to one if the IPO firm is backed by venture capital; a Top-tier Underwriter dummy is set to one if the lead underwriter has a ranking score of eight or higher based on an updated ranking in Loughran and Ritter (2004);[20] A Positive EPS dummy set to one if the IPO firm has

---

[18]See Brown and Tucker (2011) and Hanley and Hoberg (2012) for examples of measuring document similarity.

[19]Cosine similarity score has been scaled by 100 for easier interpretation.

[20]The original rankings are from Carter and Manaster (1990) and Carter, Dark, and Singh (1998).

positive EPS in the most recent fiscal year prior to IPO; Percentage of Shares Offered is included as a measure of the supply side of the IPO; Market sentiment is defined as the monthly returns of the CRSP Nasdaq value-weighted index prior to the IPO month; Firm Historical Operating Performance, is measured as annual sales in the most recent fiscal year prior to IPO. Finally, two variables related to the confidential review process are included: the length of the confidential review process, defined as the number of days from the day the DRS was filed to the day S-1 was filed; and the number of rounds of the confidential review process, defined as the number of times the DRS or DRS amendment was filed. See the Appendix for detailed variable definitions.

3.4   Model Specifications

To study the association between variables of interest in the IPO literature and the levels as well as the changes in textual analysis measures, the following OLS regression tests are performed.

Levels in textual analysis measures:

$$Y_{ijt} = \boldsymbol{\beta} LevelLM(5)_{ijt}$$
$$+ \boldsymbol{\theta} Controls(8)_{ijt} + YearFE_t + IndustryFE_j + \epsilon_{ijt} \qquad (2)$$

Changes in textual analysis measures:

$$Y_{ijt} = \boldsymbol{\beta} ChangeLM(5)_{ijt} + \gamma Similarity_{ijt} + \boldsymbol{\delta} LevelLM(5)_{ijt}$$
$$+ \boldsymbol{\theta} Controls(8)_{ijt} + YearFE_t + IndustryFE_j + \epsilon_{ijt} \qquad (3)$$

In both models, the dependent variable, Y_{ijt}, includes several variables of interest in the IPO literature, such as Offer Price Adjustment and First-day Returns as defined before. Five categories of LM word lists are used as main independent variables: %Positive, %Negative, %Litigious, %Strong-modal, and %Constraining. When focusing on the levels of textual analysis measures in Eq.(2), the contents from the DRS and S-1 are tested separately. To address the high correlation between %Negative and %Uncertain, as noted previously, the two word lists along with the aggregate measure, %Risky, are used one at

a time in separate regressions. Eq.(3) focuses on the changes of textual analysis measures during the confidential review process. The main independent variables are the cosine similarity and percentage changes for each of the five categories of LM word lists, while the levels of LM word lists are also included. The following eight variables are used in all tests as controls: VC-backed dummy; Top-tier Underwriter dummy; Positive EPS dummy; Annual sales; Shares offered; Market sentiment; the number of days in the confidential review process; and the number of rounds in the confidential review process. Previous papers, such as Hanley (1993) and Lowry and Schwert (2002), have documented a statistically significant relationship between offer price adjustment and first-day returns. Therefore, to be consistent, the percentage of offer price adjustment is added as an additional control variable when testing first-day returns. All regressions in this study also include an intercept, Fama-French 48-industry dummies and calendar year dummies. In all regression tables unless stated separately, the t-statistics are shown in parentheses and based on the standard errors clustered by industry and calendar year.

3.5 Sample Summary Statistics

There are 903 successful IPOs for the period of 2013-2020 in the constructed sample. Among them, 780 IPOs have a valid DRS-S-1 pair. This indicates that about 86% of the IPO firms (or 96% since 2017) went through the confidential review process in the post-JOBS Act era. Table 1 reports the summary of descriptive statistics for the 780 firms in the final sample.

[Table 1 here]

In Table 1 Panel A, the mean first-day returns during 2013-2020 are 23.6%, compared to 11.7% for the period of 2001-2012, as reported on Prof. Jay Ritter's website.[21] The mean offer price adjustment is -0.7%, with a standard deviation of 12.7%. Consistent with Gao, Ritter, and Zhu (2013), most of the IPO firms in the sample are relatively small, VC-backed and taken public by top-tier underwriters. Specifically, 58% of the IPO firms

---

[21]I thank Prof. Ritter for generously providing and maintaining IPO data. Data accessed from https://site.warrington.ufl.edu/ritter/ipo-data.

are funded by venture capital, while half of the firms hired a top-tier investment bank as the lead underwriter. Prior to IPO, only 22% of the IPO firms have positive trailing EPS, while the annual sales are $355 million on average. In terms of the length of the entire IPO registration process, it takes on average 160 calendar days and 4.4 amendments (including amendments to DRS and S-1) to complete an IPO.

Summary statistics for the levels of textual analysis measures in DRS and S-1 are reported in Table 1 Panel B. On average, S-1 has about 10% (or 9000) more LM words than the DRS. The proportions of each LM word list in the DRS and S-1 are at the level of 0.5%-1.8%, very similar to the level (i.e., 0.5%-1.4%) reported in Loughran and McDonald (2013).[22] Intuitively, as the objective of each section in a prospectus is different, LM word list proportions should vary across sections. This is confirmed in Table 1 Panel B comparing the levels in three major sections of the DRS or S-1, respectively. For example, in both the DRS and S-1, the Risk Factors section has the highest level of %Negative, %Uncertain and %Litigious, while the highest %Positive is found in the Business section, where the firm describes products or services, markets and the competitive landscape.

[Table 2 here]

Summary statistics for the changes in textual analysis measures in each of the three major sections are reported in the panels of Table 2, respectively. Each panel reports the percentage changes in the proportion of each LM word lists in one of the major sections. All seven categories of LM word lists decrease during the confidential review process, while %Uncertain and %Litigious decrease the most. For example, the decrease in %Uncertain is, on average, 2.6%, or 4.4 basis points. That amounts to about 44 fewer uncertain words, given a typical 100,000-word IPO prospectus. The overall decrease in LM word lists during the confidential review process is possibly due to SEC's Plain Writing Initiative. This guide, published by SEC, outlines the effort to reduce "the most common problems in disclosure documents," such as "passive voice," "legal and financial jargon," and "abstract words".[23]

---

[22]The sample period in Loughran and McDonald (2013) is 1997-2010, which is before the enactment of the JOBS Act. As a result, only S-1 filings are used for comparison. %Risky is excluded from the comparison, as it's not reported by the authors.

[23]See "A Plain English Handbook: How to Create Clear SEC Disclosure Documents", 1998, for more. Also see Loughran and McDonald (2014).

Although all LM word lists decrease when using the entire documents, there is heterogeneity across the three sections during the confidential review process. The Business section has the largest decrease in %Positive. The MD&A section has the largest decrease in %Uncertain, %Strong-modal and %Constraining. Despite the decrease in the MD&A section, %Uncertain increases by 1.3% in Business section. Unlike other sections, most of the LM word lists in the Risk Factors section increase slightly during the confidential review process. The variations in proportion of LM word lists across sections suggest that the context of the content matters when gauging the sentiment.

Table 2 also reports the cosine similarity scores by major section. As an additional measure of the changes during the confidential review process, cosine similarity score reflects the overall similarity based on the word content measured. Overall, the similarity for the entire document and the three sections are all at a very high level. The mean cosine similarity score for the entire document is 99.3. The high cosine similarity score indicates that the changes are small, relative to the large amount of word content measured. Note that having small changes does not necessarily mean that the impact of the changes is small. If, for example, a few sentences describing legal proceedings are added during the confidential review process, the impact of those changes could be significant, even though a high cosine similarity score will be observed. The Risk Factors section has the highest similarity, while the MD&A section has the lowest. This is expected since management's viewpoints are most likely to be concerned by SEC during the confidential review process. As mentioned previously, the interpretation of similarity across sections depends on each section's objective. For instance, an above-median cosine similarity score in Risk Factor section should be considered as honest disclosure of potential risks involved, and thus, a good signal to the firm.

4   Discussion of Empirical Results

In this section, I examine the associations between textual analysis measures and several variables of interest related to the pricing strategy of the offering, underpricing phenomenon, operating performance, IPO withdrawal decision, and the SEC comment letters. For each dependent variable, tests are conducted focusing on the confidential review process using the entire document and each of the three major sections: Risk Factors, Business, and MD&A. For the robustness of the study, the levels of the LM word lists in the DRS and S-1 are also tested separately. All regressions include an intercept, Fama-French 48-industry and calendar year dummies.[24]

4.1   Link Between the Confidential Review Process and SEC Comment Letter

As described in the previous discussion about the confidential review process, the evolution from DRS to S-1 is essentially an information production process by SEC. The regulator, the SEC, serves an important role in balancing information transparency and regulatory burden in the IPO market. In Lowry, Michaely, and Volkova (2020), the authors show that the comment letters issued by the SEC during the traditional registration process (i.e., from S-1 submission to IPO) contains valuable information that are relevant to firm valuation. However, SEC comment letters are not publicly available until at least 20 days after IPO. Although they know that the comment letters are informative, investors, unfortunately, have no access to the specific concerns raised by the regulator before making their investment decisions on IPO.

[Figure 3 here]

As an example, Figure 3 describes the confidential revise-and-resubmit process for Lyft, Inc. In the example of Lyft, SEC noticed the subtle details that may affect the credibility of the market share information provided in the business section of Lyft's DRS and required

---

[24]The results for intercept and control variables are omitted from some tables due to space limitations. See the Appendix for full regression results.

Lyft to revise accordingly in their subsequent revisions. As a result, Lyft addressed this concern by adding a paragraph to explain the methodology and potential conflict of interests regarding their market share. It is clear that this critical detail will increase the uncertainty of the firm's profitability. However, the SEC comment letter that pointed this out is not publicly available until 28 days after Lyft's IPO.

To propose a potential solution to the above issue, this study use textual analysis to construct a proxy for the content of SEC comment letters by the time of S-1 submission. The conjecture is that if the changes in the textual analysis measures during the confidential review process are a valid proxy for SEC's concerns, then such changes in the textual analysis measures should be able to capture the differences in market reactions when SEC comment letters are publicly released after the IPO. In this part of analysis, the paper takes advantage of the fact that the JOBS Act of 2012 created a unique opportunity that allows for extracting information contained in SEC comment letters by comparing the draft registration statement (DRS) and its formal version (S-1). The results show that it is feasible to approximate the content of SEC comment letters, on average, 51 days before the IPO, or 85 days before the their release.

In this section, a sample of SEC comment letters for all IPOs in 2013-2020 is constructed. Specifically, all the SEC comment letters regarding the IPO firm have been downloaded by selecting "UPLOAD" as the form type on EDGAR. Since only the SEC comment letters that related to the confidential review process are of interest, these letters must satisfy the following screening rules: First, the letter must be issued to the IPO firm with the same CIK. Second, the issue date of the letter (not be confused with the release date) must be between the filing date of the DRS and the filing date of the S-1. Third, the letter must be publicly available on EDGAR.[25] The final sample includes all IPO firms described in Section 3. In total, 2668 comment letters were issued to 770 IPOs firms during the confidential review process in 2013-2020.[26]

Since the objective of this analysis is to investigate whether changes in textual analysis

---

[25]It is unclear whether SEC comment letters can be requested before IPO under Freedom of Information Act (FOIA). In addition, it is unlikely that a regular investor will do so. Therefore, this study keeps the information set at the most realistic level.

[26]Ten firms had failed the screening rules described above at the time that the data were collected.

measures lead to different market reactions when the market sees the comment letters, an event study on the comment letter release is performed. Figure 4 plots the cumulative abnormal returns (CAR) over the window of ten trading days before and 20 trading days after the release day of comment letters. CAR is defined as the cumulative difference between a firm's stock return and corresponding size and book-to-market industry portfolio returns over the $[-10, +20]$ window relative to the release day of comment letters. Ryans (2020) and Lowry, Michaely, and Volkova (2020) show that revenue recognition or topics related to operating performance are the dominant concerns in the SEC comment letters. Therefore, I focus on changes in %Positive in the Business section in this part of analysis, as the results in Section 4.4 will show a strong association between positive words and operating performance. Next, all firms in the sample are divided into two groups by the median value of changes in %Positive in the Business section (i.e., below- and above-median groups). In Figure 4, the blue line represents the below-median group whose proportion of positive words are decreased by 5.3% on average. The red line, meanwhile, represents the above-median group whose proportion of positive words are increased by 1.5% on average. The black line represents all IPOs in the sample and serves as a benchmark, where proportion of positive words are decreased by 1.7% on average.

[Figure 4 here]

As expected, there is a clear difference in CAR since the release of SEC comment letter for the two groups. Firms in the above-median group, which have more positive words added during the confidential review process, experience an upward trend in CAR, while firms in the below-median group underperform relative to the above-median group, as well as to the benchmark. This observation supports my argument that the constructed ex-ante measure (i.e., changes in %Positive in the Business section during the confidential review process) captures the firm fundamental risks. Importantly, the textual analysis measures approximate the content of SEC comment letters and thus able to separate firms based on the level of concerns raised by the SEC. Prior to the release of SEC comment letters, there is no separation in CAR between the two groups. However, a market correction occurred

within two days after the comment letter release.[27] Then, the difference between below-median and above-median groups has widened consistently. By 20 trading days after the comment letter release, there is a 4.5% difference in CAR between the two groups.

[Table 3 here]

Next, I further explore the return predictability of the constructed ex-ante measure, or the proxy, in a regression framework. Table 3 presents the analysis of the link between the proxy constructed during the confidential review process and SEC comment letters . The dependent variable is the same CAR as in Figure 4 for various windows from (0, +5) to (0, +20). The results in Panel A show that percentage changes in %Positive in the Business section are strongly correlated with CAR in five, ten and 20 trading days after the release. Economically speaking, a one standard deviation increase of %Positive is associated with 2.8% cumulative abnormal return in 20 trading days since release. Panel B reports the regression results when splitting the sample into below- and above-median groups based on the percentage changes in %Positive in the Business section. The coefficient of %Positive in above-median group is strongly associated with CAR across the three tested windows. This indicates the more positive words added during the confidential review process, the better firm's stock performance will be. Additional robustness tests show that the results are consistent when using the negative word list or under several different grouping strategies. In summary, findings in Figure 4 and Table 3 provide new and strong evidence that the changes in textual analysis measures during the confidential review process are an ideal ex-ante proxy for the valuable information contained in the SEC comment letters, which are not publicly available prior to IPO. This means that, although the comment letters are not observable before the IPO, investors still can make informed investment decisions using the novel proxy proposed in this study.

---

[27]A delayed reaction is expected, as the comment letter is a specialized disclosure that does not have any announcement upon release.

4.2 Results with Offer Price Adjustment

Table 4 presents the regression results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit process, which is defined as the period between DRS submission and S-1 submission. Offer Price Adjustment is the percentage change in the offer price from the midpoint of the filing range. In untabulated results in Table 4, some known determinants that are used as control variables are found significantly correlated with IPO pricing activities as documented in the previous literature. For example, IPOs with top-tier lead underwriters and fewer offered shares have a higher offer price adjustment. VC-backed IPOs have a strong positive relationship with the offer price adjustment. Since the abovementioned results are consistent across different model specifications and the attention of this study is on textual analysis, interpretations of those IPO control variables are omitted in the remainder of the paper. See the Appendix for the full regression results.

[Table 4 here]

The first finding is that the level of %Constraining in S-1 is negatively associated, while the level of %Strong-modal is positively correlated with offer price adjustment. This is true in both baseline results and the results using entire document (column 1&2). For instance, when focusing on entire document, the coefficient on %Constraining is –15.5 (with t–statistic of –2.91). This implies that a 1% decrease in the proportion of constraining words in S-1 is associated with a 16% upward price adjustment. Following similar interpretation in Bodnaruk, Loughran, and McDonald (2015), having a low level of constraining words in the prospectus reflects that the firm is facing fewer constraints, which in turn indicates a higher degree of financial flexibility. Therefore, the offer price will be revised upward during the price discovery process, reflecting a firm's healthier financial condition. The results also show that a high level of strong modal words is an indicator of an upward price adjustment. This means that a confident tone may suggest a strong belief in the firm's value; therefore, a strong tendency to increase the offer price is observed. Interestingly, too many strong modal words may indicate over-confidence in management. The analysis of

the IPO withdrawal decision supports this interpretation, which will be discussed in detail in Section 4.5. On the other hand, none of the LM word lists has significant coefficient values (column 2). This means that after controlling for known determinants the changes in the proportions of LM word lists in the entire document cannot explain the adjustment in the offer price. This result is not surprising for two reasons. First, the majority of book-building process is not happening during the confidential review period. At this moment, firm's main goal is to fulfill the disclosure requirement. Therefore, the interaction with SEC should not be expected to have influence on firm's pricing strategy. Second, it is possible some changes might not be picked up by the textual analysis measures, given the large amount of word content in the entire document. In this case, as shown in column 3, focusing on major sections may overcome this issue and capture those subtle changes.

In column 3, a same test is performed using only the Risk Factor section. The results show a negative and significant correlation between the changes in %Negative and offer price adjustment. One standard deviation increases in %Negative (+7.7%) implies a 1% downward adjustment in offer price relative to the midpoint of the filed price range. As noted before, the confidential review process should be considered as an information production by the SEC. Thus, adding more negative words after SEC's review is considered as an unfavorable signal to the firm. It suggests either that some risks are omitted, or that the severity of the identified risks is underestimated in the original draft filing (DRS). For instance, the frequency of five representative negative words (i.e., "damage", "fail", "harm", "penalty", "unable") used in the Risk Factors section of We Co. (also known as WeWork) is significantly increased during the confidential review process. For example, in We Co.'s Risk Factors section, "fail" is used 60% more in the S-1 than in the DRS, while "harm" is used 22 times in the S-1 versus 14 times in the DRS. Moreover, after SEC's review, more than 90% of five representative negative words added are located in the Risk Factors section.[28] This supports the finding that the changes in %Negative in the Risk Factors section is precisely capturing firm risks. For the case of We Co., those changes (additions) of negative words during the confidential review process result in a downward adjustment of the firm's valuation (reflected by reducing the offer price) and eventually

---

[28]See the Appendix for details.

contribute to its withdrawal.

Although the changes in %Negative is found to be negatively correlated, the previous literature has shown that the levels of %Negative are positively associated with offer price adjustment. In column 1&2, the coefficient value on the level of %Negative is not statistically significant, while the positive sign indicates a result consistent with Loughran and McDonald (2013). Given the objective of the Risk Factors section and the findings on the %Negative, this study provides additional evidence regarding the interpretation of negative words in IPO filings. According to SEC's guidance, the Risk Factors section is supposed to identify any risk that could significantly impact the firm's performance. Therefore, a high level of %Negative should be expected in this section. Naturally, in the context of discussing risks involved in the business, the goal is to identify as many risks as possible, despite the pessimistic tone perceived by reader when more negative words are used. The hypothesis is that the more negative words used in the Risk Factor section, the more known risks are identified; and the more prepared the firm's risk management strategies are, the less vulnerable future cash flows are. In the risk management field, risks are classified based on the knowledge of occurrence and impact as "Known-knowns," "Known-unknowns," "Unknown-knowns" and "Unknown-unknowns". An example of "Known-knowns" could be seasonal fluctuations in demand or in raw material costs, while "Known-unknowns" usually refer to the risks that are known, but the damage or the level of severity is unknown at the time, such as climate disasters. It is highly unlikely that firms will discuss "Unknown-unknowns" in this section (or anywhere else in the prospectus) because it would be purely speculation, which is prohibited. Similarly, "Unknown-knowns" refers to hidden facts that can be simply addressed by hiring external specialists or professionals, and hence are not likely to be focused on. Normally, once a risk is identified in the prospectus, a corresponding strategy to mitigate its impact is proposed and explained. In addition, consistent with the arguments in Hanley and Hoberg (2012), sufficient disclosure of the underlying risks may also reduce the litigation risk that the firm would face after going public.

In summary, on the one hand, a high level of %Negative in the Risk Factors should be viewed as an indicator of adequate preparation for anticipated potential challenges. On the

other hand, the increase in the level of %Negative during the confidential review process suggests underestimated risk exposures.

4.3   Results with First-day Returns

This section presents the analysis of the underpricing phenomenon. The IPO literature has documented a strong and positive relationship between offer price adjustment and first-day returns (e.g., Hanley (1993) and Lowry and Schwert (2002)). To clearly identify the power of the textual analysis measures, offer price adjustment is added as a control when testing underpricing. Table 5 reports the regression results using the percentage changes in LM word list proportions during the confidential review process. First-day Returns is the percentage change from the offer price to the closing price on the first trading day. Consistent with previous studies, the results in Panel A show that offer price adjustment is positively correlated with first-day returns at the 1% confidence level in the four specifications across columns. In column 2&3, there is a negative relationship between %Constraining in the entire document and first-day returns, which implies the demand for firms with less financial flexibility is relatively low.

[Table 5 here]

In column 3, a strong positive relationship is found between cosine similarity score and first-day returns. This suggests that a one standard deviation increase in cosine similarity score (+4.5) leads to a 1.9% increase in first-day returns. Similar to the arguments in Hanley and Hoberg (2012) where S-1 and its amendments are compared, a high similarity can be interpreted as no additional information being produced during the confidential review process, and thus there is a high probability that the firm is omitting information. Column 4 performed the same test conditional on offer price adjustment to support this interpretation. When there is an offer price adjustment but no material changes in disclosure (reflected by high document similarity), it means a higher probability that the firm is omitting the information that caused the price adjustment. As a result of this omission, the firm faces a higher post-IPO litigation risk. According to Section 11 of the Securities Act of 1933,

investors can only sue issuers with evidence of material omissions in the prospectus AND losses below offer price. Therefore, to hedge against this increased liability risk, the firm need to underprice more if they choose not to provide additional disclosure. This is the positive relationship between document similarity and underpricing found in column 3&4. This finding also supports the "Substitution Effect" explained in Hanley and Hoberg (2012).

Table 5 Panel B reports the results using major sections of the prospectus. In the Business section (column 3), changes in %Negative have a significant coefficient of –0.1. This translates to a one standard deviation increase of %Negative in the Business section (+12.5%) associated with a 1.2% decreases in first-day returns. The addition negative words after SEC's scrutiny indicates that the firm was holding back some unfavorable information from its previous draft (DRS). Therefore, this additional negative information revealed during the confidential review process will lead to a downward adjustment of the firm valuation. The results in column 3 also show that the levels of %Negative is positively correlated with first-day returns and statistically significant. A strong negative tone after SEC's scrutiny implies a more pessimistic view regarding future cash flows, and thus it is riskier for investors to hold shares of the IPO firm, which leads to a higher amount of first-day returns through heavy underpricing. For example, a one percentage point increase in %Negative in Business section leads to 5.1% increase in first-day returns. This can explain almost 22% of the underpricing phenomenon, given a 23.6% average first-day returns in the sample. This finding supports several key theories from the IPO pricing literature in terms of sign, magnitude, and confidence level. Rock (1986) describes the "Winner's Curse" in IPO and argues that underpricing is necessary to attract uninformed investors in an asymmetric information-based setting. Beatty and Ritter (1986) provide empirical evidence showing that underpricing is positively related to ex-ante uncertainty of the firm. Loughran and McDonald (2013) also document a positive relationship between uncertain or negative word lists in S-1 filings and the first-day returns. Although there is no overlapping in the IPO sample periods, the results in this study confirm that the negative words are positively associated with first-day returns, suggesting a strong and persistent relationship over the past 20 years of the IPO market.

Interestingly, the cosine similarity score in MD&A section is negatively correlated with underpricing as shown in the last column of Panel B. This seemingly contradictory results actually follows the expectation for this particular section. Previous summary statistics show that MD&A section has the lowest similarity score comparing to other sections. This is because management's statement are most likely to be over-confident and optimistic, which in turn, concerned by SEC. Thus, high similarity in MD&A section suggests a consistent perspective in management, reflecting an observed negative relationship. Further tests in Section 4.5 using MD&A supports this interpretation and show that high similarity in MD&A section reduces probability of IPO withdrawal.

4.4 Results with Operating Performance

This paper also explores which changes in textual analysis measures are associated with the operating performance of the IPO firm. Unlike the Risk Factors section, which discusses many concerns, the Business section describes the firm's lines of business, its principal products or services and its markets. In addition, the competitive landscape, including a firm's advantages and disadvantages in relation to its competitors, significant suppliers and customers, among other matters, can be found in this section. The nature of the information related to the firm's operating performance in the Business section is usually forward-looking with a long horizon. Therefore, the hypothesis is that an overall optimistic tone in the Business section should be expected for "good quality" firms. Consistent with this expectation, %Positive is found with the highest level across all other major sections (as previously shown in Table 2). For example, the level of %Positive in the Business section is more than two times that in the MD&A section (1.46% compared to 0.67%). To further investigate the hypothesis that a high level of positive words is an indicator of good operating performance, this section starts with an univariate test in Table 6.

[Table 6 here]

In Table 6 Panels A and B, firms are placed into five quintiles based on %Positive in the Business section of the DRS or S-1, respectively. Four common measures of operating

29

performance are reported: Sales, EPS, EBIT, and EBITDA, with the mean and median values based on the most recent fiscal year prior to the IPO. In either DRS or S-1, a monotonic rise across quintiles is observed for all four operating performance measures, except for the lowest quintile.[29] There is a clear trend that implies higher %Positive is correlated with higher pre-IPO operating performance. For example, in Panel B although the mean of EPS reported in the S-1 is below zero for all quintiles, it increases from $–0.84 in the lowest to $–0.11 in the highest quintile. The difference of 73 cents in EPS is fairly large in the financial analysis, given that those IPO firms are small and less profitable in general. Similarly, Sales ranges from $57 millions to $527 millions across quintiles with an almost ten-time gap. Using DRS filings, Panel A shows the same patterns among operating performance measures in both the mean and median. Intuitively, given the objective of the Business section, more positive words indicates better operating performance at the time of IPO, thereby suggesting a more robust growth potential.

Table 7 further examines the predictability of %Positive in a regression framework as additional supports for above-memtioned findings. Using the percentage changes in LM word list proportions in the Business section, each column reports the regression results for common measures of post-IPO operating performance for various time periods after the firm went public (i.e., 6 months, 12 months). Post-IPO operating performance measures include Operating Income, and Net Income, which have been scaled by lagged total assets and winsorized at the 1% level. Fama-French 48-industry and calendar year fixed effects are included, while standard errors are clustered at the industry level.

[Table 7 here]

Consistent with the hypothesis that %Positive in the Business section is a valid predictor of a real difference in operating performance for the IPO firm, coefficient values of percentage changes in %Positive during the confidential review process are statistically significant and economically meaningful. For example, a one standard deviation increase of %Positive is associated with a 0.7% increase and a 1.3% increase in asset-adjusted net

---

[29]One possible explanation for the distinct observation in the lowest quintile of %Positive is firm's earning management behavior.

income in 6 months and 12 months since IPO, respectively. These findings highlight the fact that the changes in proportion of positive words after the SEC's reviewing process are a valid proxy for the fundamental risks of IPO firms. More importantly, these findings with operating performance document the channel whereby textual analysis measures capture firm risks as early as the time of S-1 submission. In a sense, this shows that SEC's review process reduced the information asymmetry regarding the offerings and improved the assessment of the intrinsic value of the firm.

4.5   Results with Withdrawal Decision

Thus far in this paper, I have investigated only the IPOs that went public successfully. A natural extension is to explore IPO withdrawal decisions, with a focus on the confidential review process. Unlike the SEC comment letters, the DRS is released at the time that the S-1 is publicly filed. In addition, the DRS and its amendments are available regardless of the final decision of the IPO process (i.e., withdraw or complete). Therefore, the confidential review process is ideal to study the firm's withdrawal decision because the counterfactual can be identified as those who filed a DRS initially, went through the confidential review process, but eventually decided to withdraw (e.g. We Co.). In this section, this paper examines the association between confidential review process and IPO withdrawal decision.

The withdrawn IPO sample is collected by downloading all Form RW (i.e., registration withdrawn) from EDGAR during 2013-2020. In order to have an unbiased comparison, withdrawn IPOs are required to have gone through the confidential review process. Firms who filed a DRS but never filed an S-1 or RW are excluded from the sample, because the real intentions of those firms are unclear. Additionally, firms that withdrew their IPO but successfully went public at a later date are excluded from this analysis. These screenings help to avoid selection biases. The final sample of this analysis contains 872 IPOs of which 719 were completed while 153 were withdrawn during 2013-2020. The sample summary statistics is reported in Table 8.

[Table 8 here]

31

Panel A shows that withdrawn IPOs take more time to respond to each round of revision. Although both completed and withdrawn IPOs go through the same number of rounds in confidential review process (about 2.5 rounds), it takes, on average, nine days more for withdrawn IPOs to respond to SEC's comments in each round. Intuitively, if the feedback from the SEC is critical or relates to many aspects of the firm, it will be more time-consuming for the IPO firm to revise the content or add supplemental information in their responses. Therefore, it is possible that the information production process contributes directly to the substantial difference in the final IPO decisions. Moreover, the comparison shows that withdrawn IPOs spend much more time before making a final decision on the offering and submit fewer S-1 amendments. One point to note here is that not all firms filed Form RW immediately after the management make the decision to withdraw. Some firms wait for up to nine months (or pick a more favorable date) before publicly giving up on their IPO. Thus, the total number of days spent on IPO registration process may be an inflated measure for the analysis of the withdrawal decision. To avoid this bias, this study focuses on the confidential review process that takes place prior to the S-1 submission .

Panel B presents summary statistics for LM word list levels in the S-1 filing using the entire document. There are clear distinctions between completed and withdrawn IPOs in LM word list levels. First, withdrawn IPOs contains a lower level of %Positive and a higher level of %Litigious in their prospectus. As shown in the previous analysis, this implies that the withdrawn IPOs are less profitable in terms of historical operating performance and expose to higher litigation risk. Second, the findings show that withdrawn IPOs use more strong modal words, and fewer negative and uncertain words in their disclosures. For example, on average, there are 16 basis points fewer %Risky for withdrawn IPOs. A low level of %Negative and a high level of %Strong-modal suggest an overconfidence among withdrawn firms' management. As documented in the IPO literature, managerial hubris is one of the critical factors that influences the successfulness of the IPO. This finding provides an intuitive way to measure managerial hubris through the definitive and overly optimistic tone in the prospectus.

Panel C presents summary statistics for the percentage changes in the proportions of LM

word lists during the confidential review process. Consistent with findings in the previous sections, SEC's scrutiny forces the withdrawn firm to disclose the negative information held back from their initial draft filings. During the confidential review process, there is a 1.1% increase in %Negative for withdrawn IPOs, while a 2% decrease in %Negative for completed IPOs. In addition, withdrawn IPOs have added much more content after the confidential review process, measured by total number of LM words. This implies that for IPO firms that submitted a less informative filing, the SEC has successfully "pushed out" more valuable information and reduced the degree of information asymmetry in their review, especially for withdrawn IPOs. To further explore the decision-making process of withdrawn IPOs, a series of logit regressions are tested focusing on the DRS, the S-1 and the confidential review process, respectively.

[Table 9 here]

Table 9 presents the results in the logit regression framework on the withdrawal decision. The dependent variable, RW dummy, is set to one for withdrawn IPOs and zero for completed IPOs. Each column reports the various specifications in which different documents are used in the textual analysis (i.e., DRS and S-1 in Panel A, and the "R&R" process in Panel B). All regressions include an intercept, Fama-French 48-industry dummies and calendar year dummies. First, consistent with the trends found in the summary statistics, more days spent and fewer rounds of reviews in the confidential review process indicate a higher probability of withdrawal: more time spent on the confidential review process means more issues need to be addressed before qualifying for IPO. Second, a high level of %Litigious leads to a high chance of withdrawal. Intuitively, having more legal words reflects a higher litigation risk, thus making a subsequent withdrawal more likely. Third, a high level of %Strong-modal contributes significantly to the withdrawal decision. The coefficient value on %Strong-modal is +4.1 and significant at the 1% confidence level. As an ex-ante measure of managerial hubris, the interpretation of the effect of strong modal words is that using a large amount of overconfident language (e.g., always, definitely, never) may be perceived as overselling the firm's potential from the investor's point of view. Fourth, a high level of %Negative words reduces the likelihood of IPO withdrawal. More negative words

used in the prospectus may lead to by a downward revision in offer price, thus making it much easier to sell the offered shares. This is in line with the "Winner's Curse" described in Rock (1986) and found in the previous analysis with underpricing: %Negative is positively correlated with underpricing. Moreover, more negative information can be seen as a more realistic and conservative view of firm management, and thus may reduce the post-IPO litigation risk.

In summary, litigation risk and managerial hubris are strong factors that contributes to the IPO withdrawn decision. Economically speaking, moving from 5th percentile to 95th percentile for %Litigious increases the probability of withdrawal from 12% to 25%. Similarly for %Strong-modal, the withdrawal probability increases from 11% to 27%. However, adding more negative words will reduce probability of withdrawal from 24% to 13%, when moving from 5th percentile to 95th percentile of %Negative.

So far, the analysis has shown that high litigation risk and managerial hubris contributes to the withdrawal decision. Since the decision is made by firm insiders (i.e., the management), it is interesting to see whether the word content in the MD&A, a section that contains mostly management's opinion, is informative in predicting IPO withdrawal. Table 9 Panel B presents the same series of tests that are conducted using the entire document or the MD&A section only. One of the reasons that the MD&A section attracts so much attention from investors and regulators is not only that it contains management's perspectives regarding firm's financial condition and results of operations, but also that it is discussed in narrative form. This narrative section should be the easiest way for outsiders to understand the risks incorporated with the shares issued.[30] However, the prior literature focusing on the MD&A section finds mixed results. Li (2010) finds no positive relation between the tone of the MD&A section in annual reports (10-K) and future performance, while Davis and Tama-Sweet (2012) find a significant linkage between the MD&A's tone in 10-K and future ROA. Using MD&A disclosures in 10-K, Hoberg and Lewis (2017) find

---

[30]In 2003, the SEC issued interpretive guidance for MD&A, which states the principal objectives of the section is "[t]o provide a narrative explanation of a company's financial statements that enables investors to see the company through the eyes of management...[and][t]o provide information about the quality of, and potential variability of, a company's earnings and cash flow, so that investors can ascertain the likelihood that past performance is indicative of future performance."

that the MD&A contains valuable information that can be used to detect fraud.

In both columns of Panel B, there is a strong negative relationship between cosine similarity score and withdrawal decision. Moreover, the effect is almost twice larger in the MD&A section than in the entire document. High similarity in the MD&A section indicates consistent statements before and after the SEC's review process. An intuitive explanation is that if the SEC finds management's story questionable, there will be material changes in content of MD&A between DRS and S-1, which should be captured by textual analysis measures. Thus, the inconsistency in MD&A should probably be considered a flaw in the statement or a change in the firm's risk environment. Either way, from an investor's perspective, this is a negative signal to the firm, which makes the offering less attractive.

In column 2 of Panel B, the findings suggest the levels of word list in the MD&A section is not able to explain withdrawal decision. The lack of predictive power of the LM word list levels in the MD&A section is possibly due to the fact that most statements in MD&A are carefully screened by the legal or public relation departments of the firm. The assumption is that all the statements provided in the MD&A are well prepared to avoid subsequent prosecution: what to say and how to say it in the MD&A section have been strategically designed to reduce the post-IPO litigation risk. Following this assumption, one should expect no significant results for text-based analysis using the MD&A section. For example, Loughran and McDonald (2011) use firms' annual reports to show the MD&A section does not produce better tone measures.

[Table 10 here]

To further support the "strategically designed statement" argument, Table 10 compares the level of LM word lists in MD&A with that in Risk Factors, Business and the Entire Document. Among the seven LM word lists examined, six have the lowest level for MD&A, with the exception of %Uncertain. Specifically, the level of %Litigious in the MD&A section of the S-1 is 0.5%, 46% lower than the level in the Business section, or 70% lower than the level in the Risk Factors section. It seems that firm management avoid mentioning legal matters in the MD&A section. Indeed, Hoberg and Lewis (2017) find that personal names are mentioned less frequently in the MD&A section to avoid becoming a defendant in

subsequent prosecutions. Chen et al. (2020) document that firms may alter their wordings such that a favorable output will be generated by machine and AI readers. Meanwhile, high %Uncertain makes it more challenging to interpret the exact meaning of the statements; thus, the statement will be less likely to be used as evidence against the managers in subsequent litigation. Therefore, it is reasonable to assume that statements in the MD&A section have been designed to reduce litigation risk against the management. In summary, although the levels of LM word lists in the MD&A section fail to explain withdrawal decisions, the changes during the confidential review process show that the consistency in management's statements, measured by cosine similarity, is associated with the decision-making process of withdrawal.

5 Policy Impact and SEC's Role in the IPO Market

In the final part of the study, I investigate the impact of the confidential revise-and-resubmit process on two aspects of the IPO market: the transparency of corporate disclosures and the role of regulator in the public capital market. When the JOBS Act of 2012 was signed into law, one of its main objectives was to revive the IPO market. Therefore, SEC, as the dominating regulator in the IPO market, plays an irreplaceable role in protecting investors, maintaining fair and efficient markets, and facilitating capital formation.[31] By comparing the changes in each firm's prospectus before and after the confidential revise-and-resubmit process, I find that the transparency and quality of disclosure has been improved, while SEC has stayed true to its neutral role in monitoring the public capital market. Specifically, there is a 10% increase in the word content from DRS to S-1 after the confidential revise-and-resubmit process; meanwhile, the complexity of word content, measured by the average syllables per word or by the average word length, has been significantly reduced. On the other hand, the changes made through the confidential revise-and-resubmit process are found insignificantly associated with the firm's withdrawal decision, suggesting there is no direct intervention by the regulator in determining the firm's access to the public capital.

---

[31]SEC's mission as stated on official webpage at https://www.sec.gov/about/what-we-do.

[Table 11 here]

Table 11 Panel A and B, show the comparison of U.S. IPO market during pre- and post-JOBS Act period in terms of the total time spent on the entire IPO registration process and the overall IPO completion rate, respectively. Pre-JOBS Act includes 2680 IPOs during 1997-2010, of which 1887 were completed while 793 were withdrawn.[32] The post-JOBS Act period includes 872 IPOs, of which 719 were completed while 153 were withdrawn during 2013-2020. IPO completion rate is defined as the fraction of firms that filed S-1 and successfully went public.

Prior to the JOBS Act, it took on average 117 calendar days to complete an IPO, while the number of days increased to 157 in the post-JOBS Act period. The increased time in the IPO process is primarily due to the confidential revise-and-resubmit process. When excluding the period of the confidential review process, the results suggest that the registration period after filing the S-1, or the public review period, is substantially reduced by 66 days (or 55%). However, the actual IPO process from the firm's perspective, which starts with filing the DRS confidentially, is lengthened by 40 days (or 34%) on average. Considering the expensive consulting and legal fees associated with, on average, 1.6 revisions to the original DRS, the firm's cost is expected to increase significantly. However, it seems the additional time and money spent on the confidential review process is not justified by higher IPO proceeds for the firm, as the mean of the post-JOBS Act first-day returns is 24%, compared to 11% during 2001-2011. On the other hand, the post-JOBS Act IPO completion rate is notably increased from 70.4% to 82.5%. It seems the lengthened registration time contributes to the improvement in the completion rate, but it is worth mentioning that the overall market condition may be different between the pre- and post-JOBS Act periods.

[Figure 5 here]

[Figure 6 here]

In the previous discussion of Figure 2, there is a clear trend that although the median number of unique financial words used in DRS has increased every year since 2013, the

---

[32]Pre-JOBS Act IPO data are based on Loughran and McDonald (2013).

corresponding S-1 consistently contains about 10% more unique financial words than that of DRS. On average, the total words used in S-1 is about 9000 more than in DRS and it is statistically significant at 1% level. This suggests that more information regarding the firm has been revealed after the confidential revise-and-resubmit process. Next, this study looks into the readability of those additional information, defined as the word complexity. In the linguistics literature, one common measure of the readability is the average syllables per word.[33] Figure 5 plots the median level of word complexity in the DRS and S-1 for the main sample used in this study, respectively. The comparison between the DRS and S-1 shows that the word complexity of prospectus has reduced consistently after SEC's review across the past eight years since the enactment of JOBS Act in 2012. On average, there is a 15% decrease in word complexity after the confidential review. In addition, the notable decreases in word complexity following confidential review is found across industries. Figure 6 plots the median of percentage changes in word complexity during the confidential review process by Fame-French 12-industry classification. Across all industries except utility and chemistry industries, it shows a consistent reduction in word complexity. Intuitively, the two exceptions are probably due to their characteristics, as more-complex words or terminologies are needed to clearly explain the complicated business operation involved in the utility and chemistry industries. Overall, the evidence show that the confidential review process improves the transparency of the IPO disclosure in terms of a 10% increase in content and a 15% reduction in word complexity. This improvement may have an indirect impact on reducing the issues of uncertainty and asymmetric information in equity issuance. Although the quantification of the exact impact is beyond this study's scope, in general, one should expect a positive effect on the proportion of small investors, the likelihood of seasoned equity issuance, and improved corporate governance.[34]

While the information quality and transparency has been improved after the confidential review, I further test whether this indicates an excessive intervention by regulator. There has been a long debate on whether the regulatory overburden contributes to the sluggish

---

[33]Other measures, such as FOG Index and SMOG Index, have been tested for robustness and the results are qualitatively the same.

[34]Loughran and McDonald (2010) have documented a positive impact of improved 10-K readability in the stock market.

IPO market. This study finds that the extra confidential review process with SEC does not directly determine the firm's access to the public capital. The results from the withdrawal analysis show that it is not the changes made during the confidential review process, but the intrinsic risks inherited in the offering contribute to the firm's withdrawal decision.

[Table 12 here]

Table 12 Panel A shows the comparison of determinants on the withdrawal decision in several specifications. In column 1, several levels of word list proportions are significantly associated with withdrawal decision as discussed in precious section. Column 2 uses the percentage change in word list proportions as main variable of interests, while the levels are included as control. The results show that four out of five word lists changes used are insignificantly correlated with withdrawal decision. Moreover, the magnitude of the only exception is negligible comparing to results with levels in column 1. Column 3 further tests the interaction terms in the same setup and finds consistent evidence.

Panel B reports the average marginal effects on withdrawal decision for the previous logit regressions. Comparing to the effects by levels of word list proportions, the effects by the changes in the word list proportions during confidential review process are mostly insignificant. This means going through the confidential review process neither makes a firm looks better nor worse from the disclosure perspective. Intuitively, it is not the changes made leads to an unsuccessful IPO, but the intrinsic risks of firm determines the value of the offerings. SEC's review simply reveals more content that are easier to understand and leaves it to the market for a fair valuation of the IPO. In summary, I find that after the confidential review process, there are a 10% increase in content and significant decrease in word complexity. Meanwhile, although there are more interactions with the IPO firm, SEC has maintained a neutral role in overseeing the public capital market.

6 Conclusion

One of the most critical stages in a firm's life cycle is going public. Since the enactment of the JOBS Act in 2012 and its expansion in 2017, a new confidential IPO registration process has been created and broadly adopted. This paper is the first to study the confidential revise-and-resubmit process in the post-JOBS Act era and strives to answer the vital question: Is there any valuable information produced during the confidential review process between the firm and the SEC? Specifically, using five categories of Loughran-McDonald (LM) word lists, this study documents the link between the changes in content during the evolution from DRS to S-1 and several critical matters related to IPOs, such as offer price adjustment, first-day returns, post-IPO performance, IPO withdrawal decision, as well as the link to the SEC comment letter and the overall impacts of the JOBS Act on the U.S. IPO market. In addition to using the entire contents of prospectuses, a series of analyses are conducted focusing on three specific sections to take into account the context of the word content: Risk Factors, Business, and Management's Discussion and Analysis (MD&A).

The main findings in this study shed light on the financial economics of the confidential review process for the pricing strategy of the offering, the underpricing phenomenon, the firm's information environment, the rationale for the withdrawal decision, market efficiency, and the role of the regulator. First, the event study on the release of SEC comment letters provides new evidence that the changes in textual analysis measures during confidential review process are an ideal ex-ante proxy for the content of SEC comment letters prior to their release and before IPO. Second, the evidence suggests that the additional risks revealed by the SEC's review process are mainly captured by the changes in the proportion of negative words in the Risk Factors section, where risks involved in the offering are discussed. Third, the overall change in word content during the confidential review process, measured by cosine similarity, is found to be strongly and positively correlated with first-day returns. Fourth, the analysis of operating performance documents that it is the real effects that drives the return predictability of the textual analysis measures. Fifth, the study shows that a high level of litigation risk and managerial hubris makes a subsequent

withdrawal more likely. Lastly, I find that through the confidential review process, the transparency and quality of disclosure has been significantly improved, while SEC has stayed true to its neutral role in monitoring the public capital market.

A natural extension of this study is to utilize the DRS filings and the confidential review process further. Since the DRS and its amendments are available regardless of the final IPO decision, they provide an opportunity to study the firm's choice of raising capital because the counterfactual can be identified as those who filed a DRS initially but later decided to pursue other alternatives (e.g., merger and acquisition, or via Special Purpose Acquisition Company). Moreover, examining the impact of the confidential review process on the IPO market has policy implications regarding the regulator's role in balancing information transparency and regulatory burden. For example, the costs-benefits of regulator's attempts to increase transparency for climate-change-related and cyber-security-related disclosures still needs further analysis.

References

R. P. Beatty and J. R. Ritter. Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15:213–232, 1986.

L. M. Benveniste and W. Y. Busaba. Bookbuilding vs. fixed price: An analysis of competing strategies for marketing ipos. *Journal of Financial and Quantitative Analysis*, pages 383–403, 1997.

L. M. Benveniste and P. A. Spindt. How investment bankers determine the offer price and allocation of new issues. *Journal of financial Economics*, 24(2):343–361, 1989.

S. Bhattacharya and J. R. Ritter. Innovation and communication: Signalling with partial disclosure. *The Review of Economic Studies*, 50(2):331–346, 1983.

A. Bodnaruk, T. Loughran, and B. McDonald. Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, pages 623–646, 2015.

D. J. Bradley and B. D. Jordan. Partial adjustment to public information and ipo underpricing. *Journal of Financial and Quantitative Analysis*, pages 595–616, 2002.

S. V. Brown and J. W. Tucker. Large-sample evidence on firms' year-over-year md&a modifications. *Journal of Accounting Research*, 49(2):309–346, 2011.

R. Carter and S. Manaster. Initial public offerings and underwriter reputation. *the Journal of Finance*, 45(4):1045–1067, 1990.

R. B. Carter, F. H. Dark, and A. K. Singh. Underwriter reputation, initial returns, and the long-run performance of ipo stocks. *The Journal of Finance*, 53(1):285–311, 1998.

T. J. Chemmanur. The pricing of initial public offerings: A dynamic model with information production. *The Journal of Finance*, 48(1):285–304, 1993.

T. J. Chemmanur and P. Fulghieri. A theory of the going-public decision. *The review of financial studies*, 12(2):249–279, 1999.

T. J. Chemmanur, S. He, and D. K. Nandy. The going-public decision and the product market. *The Review of Financial Studies*, 23(5):1855–1908, 2010.

C. Y.-H. Chen, R. Despres, L. Guo, and T. Renault. What makes cryptocurrencies special? investor sentiment and return predictability during the bubble. *Investor Sentiment and Return Predictability During the Bubble (June 3, 2019)*, 2019.

A. K. Davis and I. Tama-Sweet. Managers' use of language across alternative disclosure outlets: Earnings press releases versus md&a. *Contemporary Accounting Research*, 29 (3):804–837, 2012.

P. M. Dechow, A. Lawrence, and J. P. Ryans. Sec comment letters and insider sales. *The Accounting Review*, 91(2):401–439, 2016.

C. Doidge, K. M. Kahle, G. A. Karolyi, and R. M. Stulz. Eclipse of the public corporation or eclipse of the public markets? *Journal of Applied Corporate Finance*, 30(1):8–16, 2018.

S. P. Ferris, Q. Hao, and M.-Y. Liao. The effect of issuer conservatism on ipo pricing and performance. *Review of Finance*, 17(3):993–1027, 2013.

X. Gao, J. R. Ritter, and Z. Zhu. Where have all the ipos gone? *Journal of Financial and Quantitative Analysis*, 48(6):1663–1692, 2013.

K. W. Hanley. The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics*, 34:231–250, 1993.

K. W. Hanley and G. Hoberg. The information content of ipo prospectuses. *The Review of Financial Studies*, 23(7):2821–2864, 2010.

K. W. Hanley and G. Hoberg. Litigation risk, strategic disclosure and the underpricing of initial public offerings. *Journal of Financial Economics*, 103(2):235–254, 2012.

G. Hoberg and C. Lewis. Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance*, 43:58–85, 2017.

M. Hu and B. Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

J. Lerner. Venture capitalists and the decision to go public. *Journal of financial Economics*, 35(3):293–316, 1994.

F. Li. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. *Journal of Accounting Research*, 48(5):1049–1102, 2010.

T. Loughran and B. McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

T. Loughran and B. McDonald. Ipo first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics*, 109(2):307–326, 2013.

T. Loughran and B. McDonald. Regulation and financial disclosure: The impact of plain english. *Journal of Regulatory Economics*, 45(1):94–113, 2014.

T. Loughran and B. McDonald. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230, 2016.

T. Loughran and B. McDonald. Textual analysis in finance. *Annual Review of Financial Economics*, 12:357–375, 2020.

T. Loughran and J. Ritter. Why has ipo underpricing changed over time? *Financial management*, pages 5–37, 2004.

T. Loughran and J. R. Ritter. Why don't issuers get upset about leaving money on the table in ipos? *The Review of Financial Studies*, 15(2):413–444, 2002.

M. Lowry and G. W. Schwert. Ipo market cycles: Bubbles or sequential learning? *The Journal of Finance*, 57(3):1171–1200, 2002.

M. Lowry, R. Michaely, and E. Volkova. Information revealed through the regulatory process: Interactions between the sec and companies ahead of their ipo. *The Review of Financial Studies*, 2020.

P. G. Mahoney. Mandatory disclosure as a solution to agency problems. *The University of Chicago Law Review*, 62(3):1047–1112, 1995.

V. Maksimovic and P. Pichler. Technological innovation and initial public offerings. *The Review of Financial Studies*, 14(2):459–494, 2001.

K. Rock. Why new issues are underpriced. *Journal of financial economics*, 15(1-2):187–212, 1986.

J. P. Ryans. Textual classification of sec comment letters. *Review of Accounting Studies*, pages 1–44, 2020.

C. J. Simon. The effect of the 1933 securities act on investor information and the performance of new issues. *The American Economic Review*, pages 295–318, 1989.
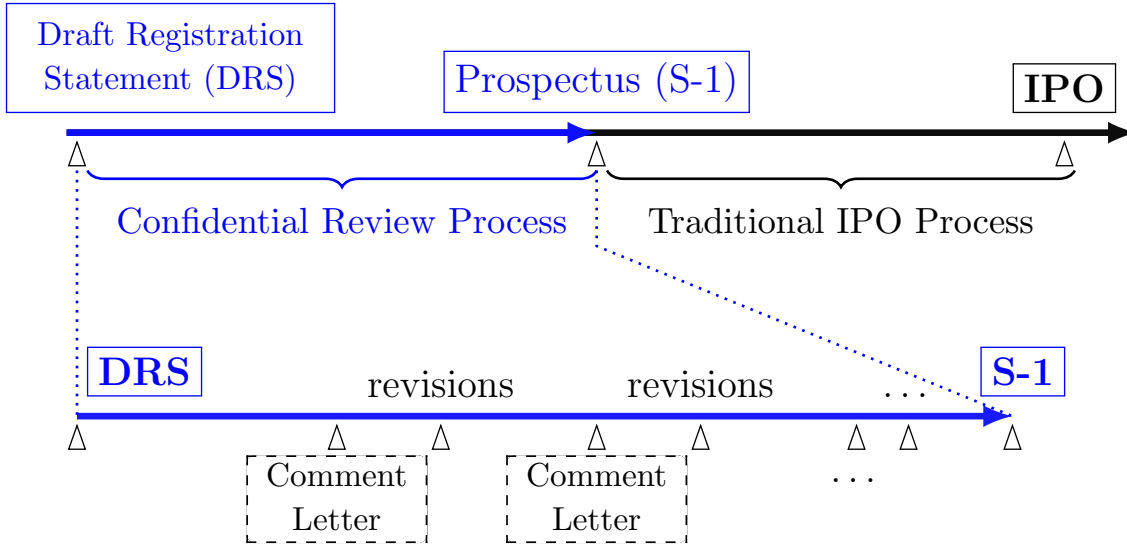
Figure 1: Post-JOBS Act IPO Registration Process

The entire timeline on the top describes a typical IPO registration process in the post-JOBS Act era. Before the JOBS Act, the traditional IPO process started with publicly filing an IPO prospectus (S-1), as illustrated in the black part of the timeline. The blue part of the timeline shows the confidential revise-and-resubmit process created by the JOBS Act of 2012. The timeline on the bottom further describes the interactions between the firm and SEC in detail. The revisions are amendments to the original draft (DRS). Comment Letter is the feedback provided by the SEC regarding its concerns in the firm's DRS. The dashed box indicates that comment letters are not observable before the IPO.

In the example of Lyft, Inc., the IPO registration process is as follows.

- 2018/12/6: Confidentially filed DRS
- 2019/2/7: Confidentially filed the first revision to DRS
- 2019/2/25: Confidentially filed the second revision to DRS
- 2019/3/1: Publicly filed S-1 (DRS and its revisions are released)
- 2019/3/29: IPO at $72 on NASDAQ (closing at $78.29)
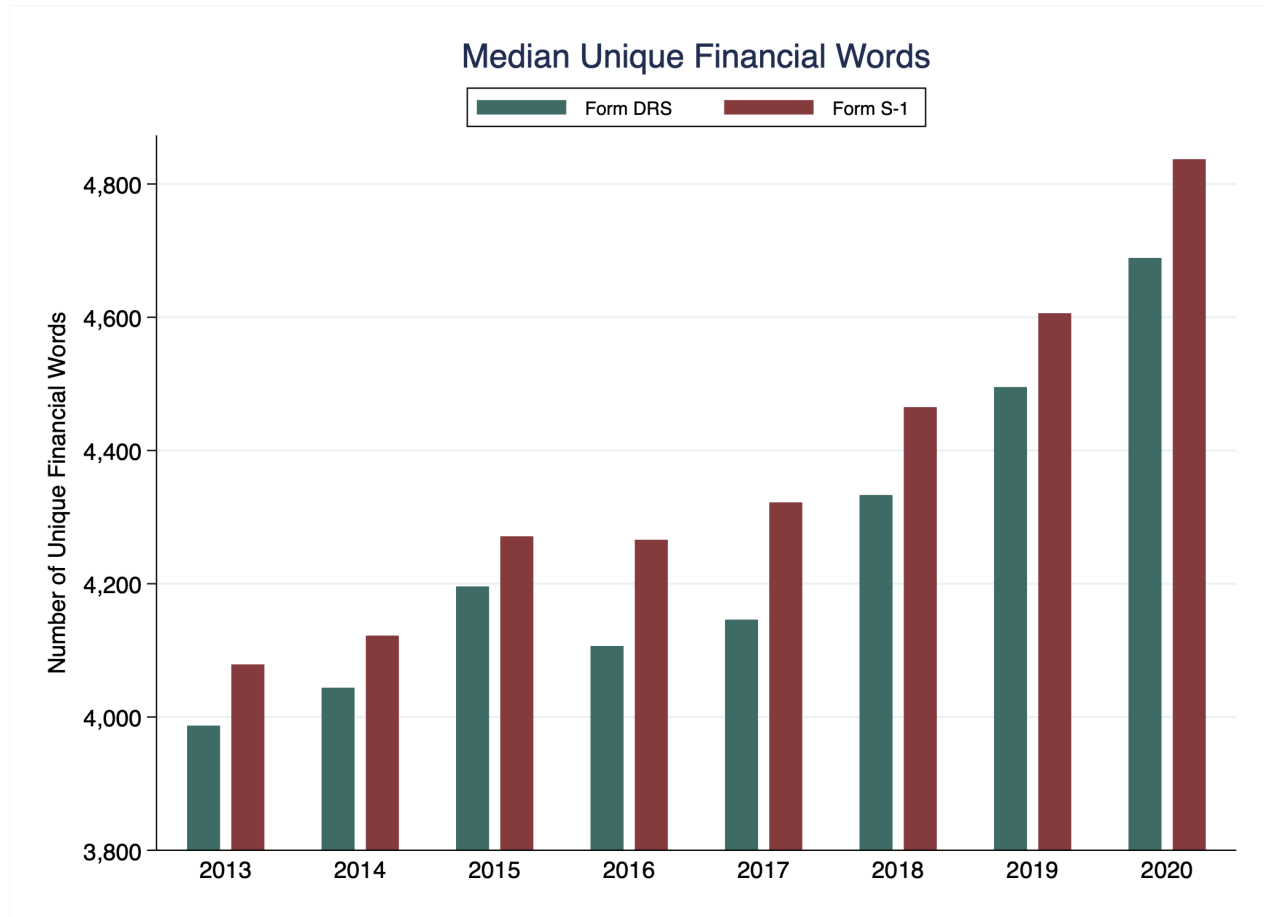
Figure 2: Unique Financial Words in DRS and S-1 Filings

This figure plots the median number of unique financial words in the Draft Registration Statement (DRS) and prospectus (S-1) during 2013-2020. The sample includes 780 IPOs that went through the confidential revise-and-resubmit process (defined as from DRS to S-1). IPOs with an offer price of less than $5 per share, financial firms, ADRs, units or rights offers, REITs, natural resource limited partnerships, closed-end funds, and stocks not listed on NYSE, NASDAQ, and AMEX are excluded from the sample. Unique Financial Words is defined as the words included in the word list dictionary created in Loughran and McDonald (2011).

Excerpts from the Business Section of Lyft's DRS
(observable upon S-1 submission)

Our values, brand, innovation and focused execution have driven significant growth in market share and in the number of users on our platform. As ridesharing becomes more mainstream, we believe that users are increasingly choosing a ridesharing platform based on brand affinity and value alignment. Our U.S. ridesharing market share was 39% in December 2018, up from 22% in December 2016.[40] This growth comes from both new drivers and riders as well as increased ride frequency. For the quarter ended December 31, 2018, we had 18.6 million Active Riders and over 1.1 million drivers who provided rides.

⬇

Excerpts from the SEC's Comment Letter
(not observable before IPO)

In several footnotes throughout this filing, you state that market share figures are based on the number of rides and were gathered by Slice Technologies, Inc., doing business as Rakuten Intelligence, and that an entity affiliated with Rakuten currently holds more than 5% of your outstanding capital stock. Please disclose this information in this section and file the consent of Rakuten as an exhibit. Please also discuss the methodology, including material assumptions and estimates, used by Rakuten Intelligence to derive market share figures.

⬇

Excerpts from the Business Section of Lyft's S-1
(observable upon S-1 submission)

Certain U.S. ridesharing market share figures contained in this prospectus are based on the number of rides provided by drivers using Lyft or Uber in the United States, and such ride data was collected by Rakuten Intelligence. Rakuten Intelligence's ride data was collected using its proprietary technology that identifies Lyft and Uber e-receipts across a panel of over four million users of Rakuten Intelligence products. Rakuten Intelligence then normalized such ride data to account for certain overrepresented and underrepresented characteristics of its panel. Rakuten is the parent company of Rakuten Intelligence, and entities affiliated with Rakuten currently hold more than 5% of our outstanding Class A common stock. For more information, see the sections titled "Principal Stockholders" and "Certain Relationships and Related Party Transactions."

Figure 3: Example of the SEC's Information Production Process

This figure describes the confidential revise-and-resubmit process for Lyft, Inc. The excerpts are from the business section of Lyft's Draft Registration Statement (DRS), SEC's comment letter, and Lyft's S-1 (the last revised version of DRS), respectively. SEC comment letters are not observable until 28 days after Lyft's IPO, while both DRS and S-1 are publicly available at the time of the S-1 submission.
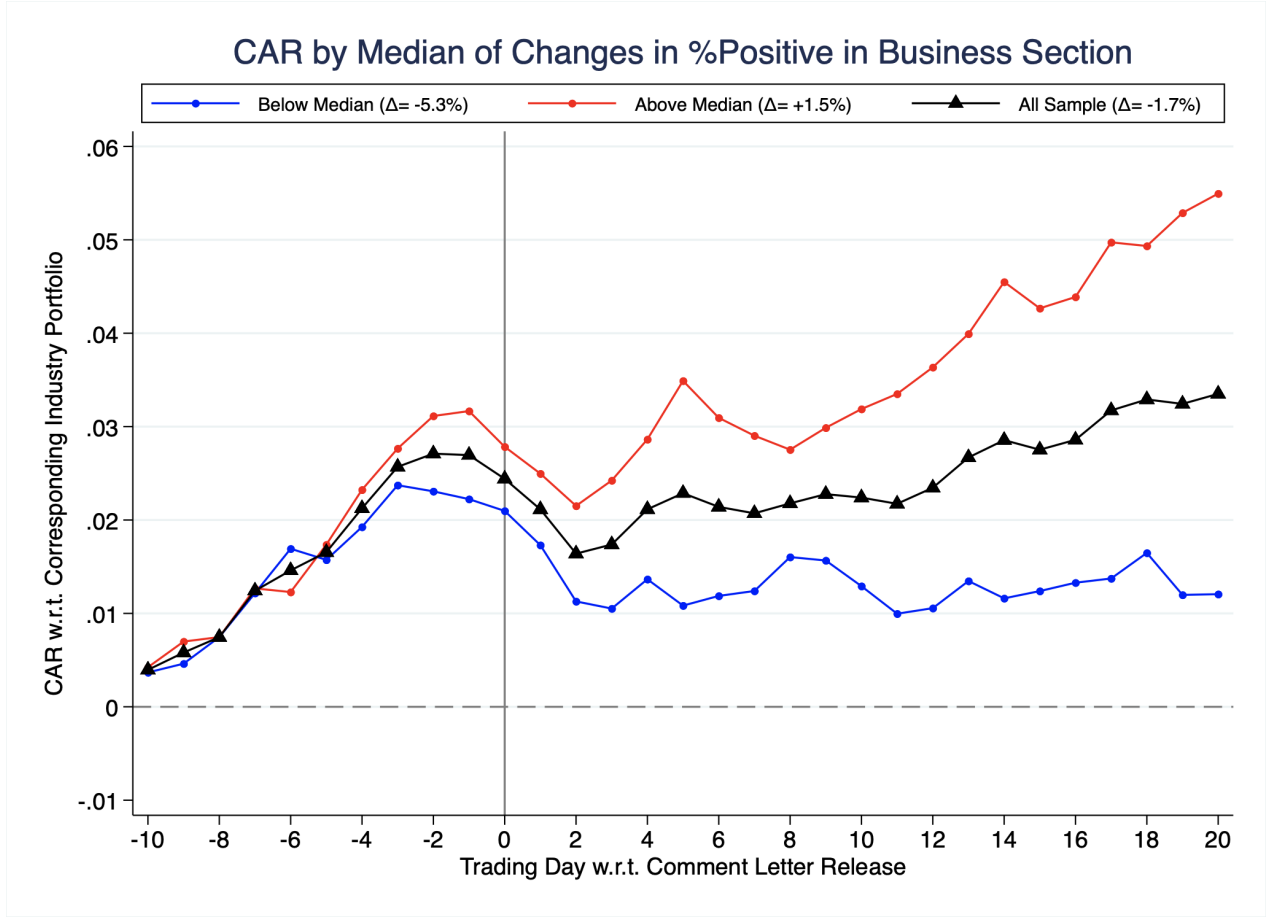
Figure 4: CAR by Median of Changes in Positive Words During R&R

This figure plots the cumulative abnormal returns (CAR) for two groups of IPOs, as well as a benchmark. The two groups are constructed based on the median value of the percentage changes in %Positive in the Business section during the confidential revise-and-resubmit process (from DRS to S-1). The blue line represents the below-median group whose proportion of positive words is decreased by 5.3% during the confidential review process. In contrast, the red line represents the above-median group whose proportion of positive words is increased by 1.5% on average. The black line is the benchmark that represents all IPOs in the sample. CAR is defined as the cumulative difference between a firm's stock return and corresponding size and book-to-market industry portfolio returns over the $[-10, +20]$ days window that is relative to the release day of comment letters.
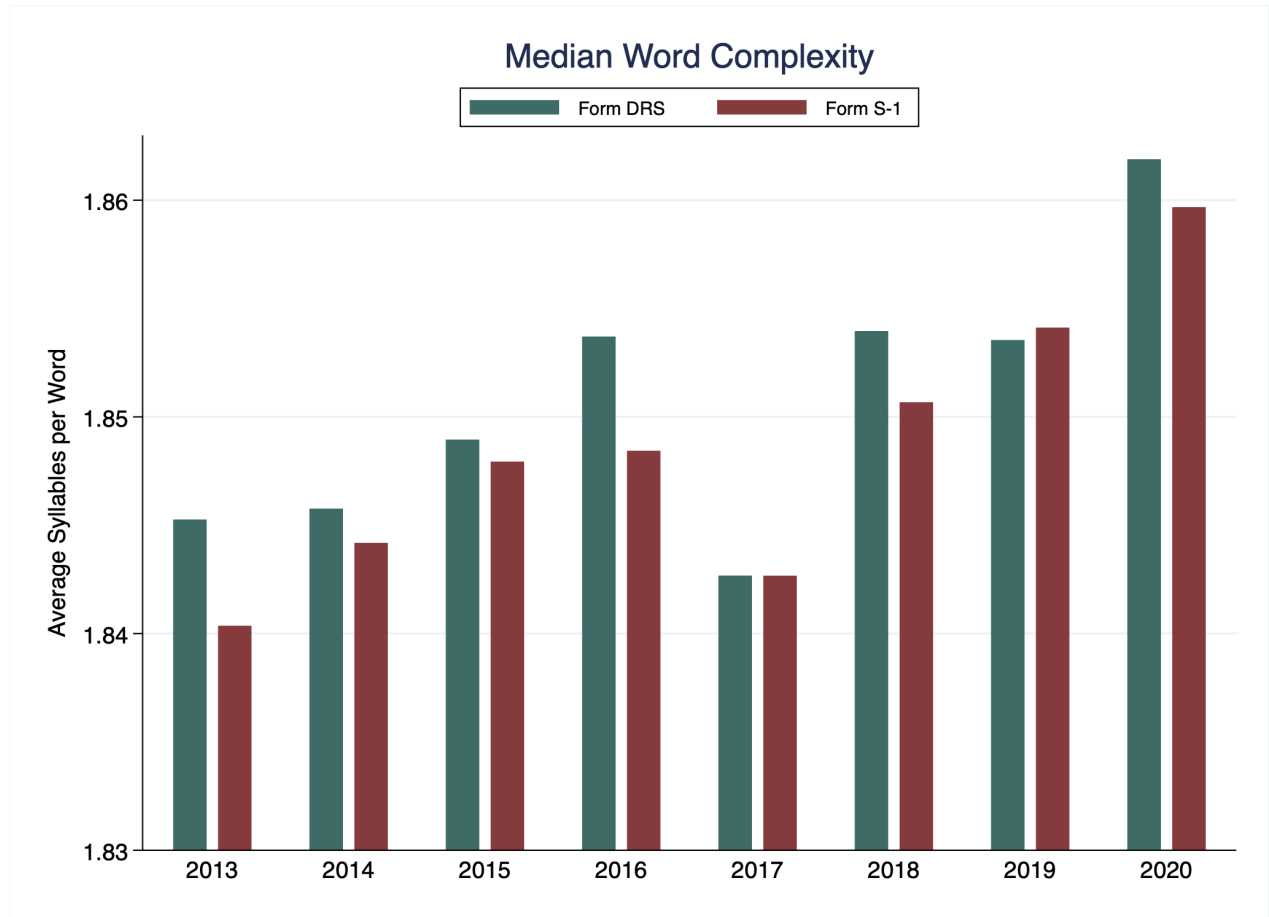
Figure 5: Word Complexity in DRS and S-1 Filings

This figure plots the median value of word complexity in the Draft Registration Statement (DRS) and the prospectus (Form S-1) for 780 IPOs that went through the confidential revise-and-resubmit process (defined as from DRS to S-1) in 2013-2020, respectively. Word Complexity is defined as the average syllables per word, as in Loughran and McDonald (2011).
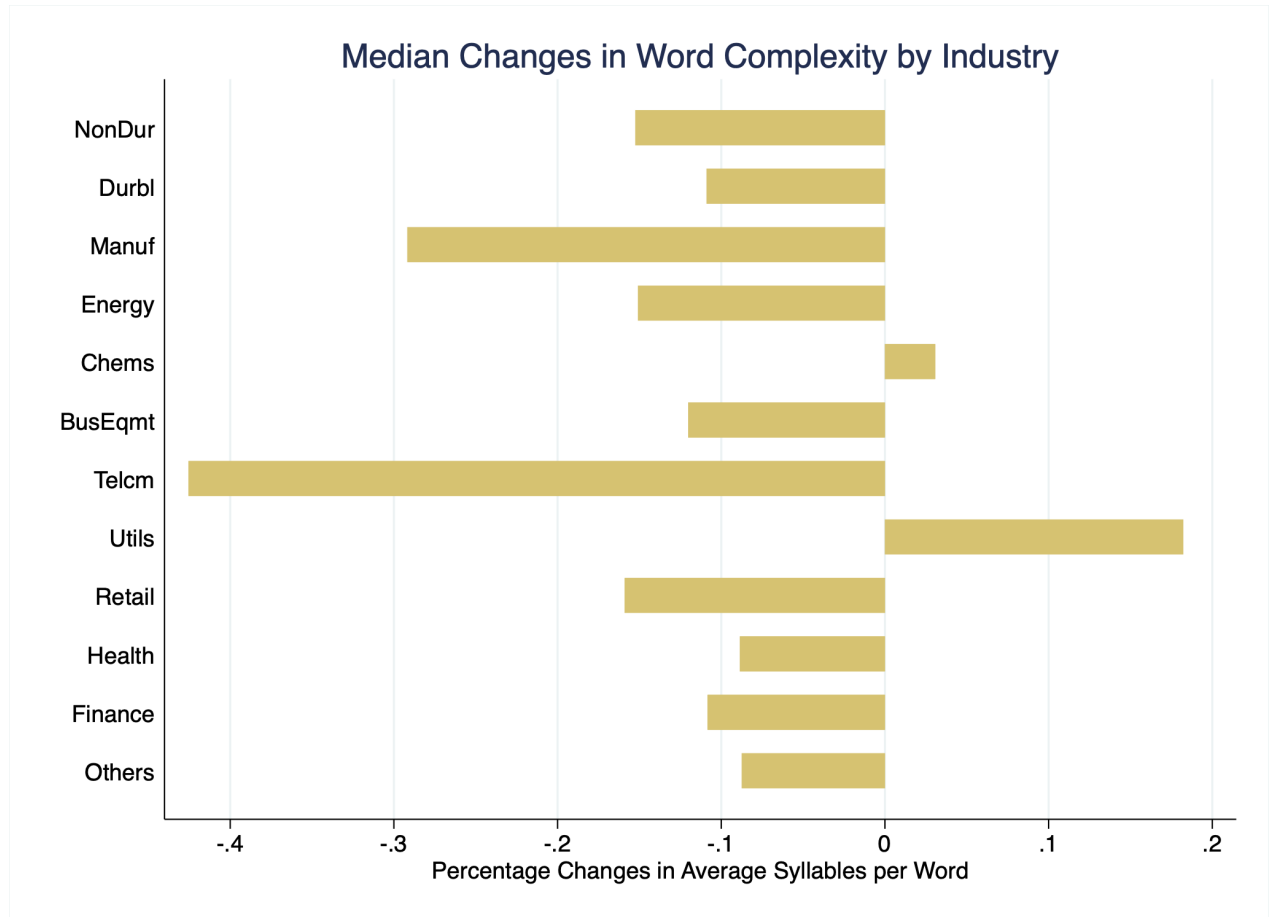
Figure 6: Changes in Word Complexity by Industry

This figure plots the median value of percentage changes in word complexity during the confidential revise-and-resubmit process by industry. The sample includes 780 IPOs that went through the confidential revise-and-resubmit process (defined as from DRS to S-1). Industry is defined using Fama-French 12-industry classification. Word Complexity is defined as the average syllables per word, as in Loughran and McDonald (2011).

Table 1: Summary Statistics for IPO Sample, 2013-2020

The sample includes 780 IPOs that went through the confidential revise-and-resubmit (R&R) process (defined as from DRS to S-1) in 2013-2020. Panel A reports the summary statistics of IPO related variables; Panel B reports the summary statistics of the levels of seven Loughran and McDonald (2011) word lists by major sections of IPO prospectus. IPOs with an offer price of less than $5 per share, financial firms, ADRs, units or rights offers, REITs, natural resource limited partnerships, closed-end funds, and stocks not listed on NYSE, NASDAQ and AMEX are excluded from the sample. Form S-1 is the registration statement for initial public offering filed publicly with the Securities and Exchange Commission (SEC). Form DRS is the Draft Registration Statement filed confidentially with SEC prior to Form S-1. Offer Price Adjustment is the percentage change in the offer price from the mid-point of the filing range; First-day Returns is the percentage change from the offer price to the closing price on the first trading day. See Appendix for other variable definitions.

Panel A: IPO Variables

	Obs.	Mean	SD	5th	25th	Median	75th	95th
Offer Price Adjustment (in %)	780	-0.69	12.70	-27.27	-6.46	0.00	7.14	17.65
First-day Returns (in %)	780	23.63	36.86	-16.96	0.00	13.69	38.89	93.46
VC-backed dummy	780	0.58	0.49	0	0	1	1	1
Top-tier Underwriter dummy	780	0.49	0.50	0	0	0	1	1
Positive EPS dummy	780	0.22	0.42	0	0	0	0	1
Annual Sales (in millions)	780	355.47	2390.82	0.00	0.19	51.74	198.11	1021.64
Prior Month Nasdaq Return (in %)	780	2.10	4.26	-5.00	-1.16	2.39	4.86	9.30
Share Offered (in %)	780	26.51	16.49	9.34	16.66	22.92	30.49	64.11
Rounds of Confidential R&R	780	2.55	1.23	1	2	2	3	5
Period of Confidential R&R (in days)	780	108.48	106.53	33	52	72	117	320
Period from S-1 Submission to IPO (in days)	780	51.95	70.32	21	26	31	44	155
Entire IPO Registration Process (in days)	780	160.44	126.81	63	87	117	180	413
Number of DRS Amendments	780	1.55	1.23	0	1	1	2	4
Number of S-1 Amendments	780	2.85	1.58	1	2	3	4	6
Number of All Amendments	780	4.40	1.88	2	3	4	5	8

Panel B: Summary Statistics for the Levels of LM Word Lists

	Entire Document			Risk Factors			Business			MD&A		
	DRS	S-1	Diff.	DRS	S-1	Diff.	DRS	S-1	Diff.	DRS	S-1	Diff.
Size of Raw File (in MB)	0.79	0.87	0.08	0.16	0.16	0.00	0.11	0.11	0.01	0.08	0.10	0.02
Number of LM Words	94920	104197	9277	22445	23061	616	13762	14912	1150	9081	10542	1460
LM Word Lists Levels	(%)	(%)	(bps)	(%)	(%)	(bps)	(%)	(%)	(bps)	(%)	(%)	(bps)
%Positive	0.85	0.84	-1.45	1.17	1.16	-0.69	1.49	1.46	-2.85	0.68	0.67	-1.63
%Negative	1.84	1.81	-3.86	3.89	3.88	-1.60	1.54	1.53	-1.29	1.05	1.05	-0.20
%Uncertain	1.63	1.59	-4.41	3.17	3.16	-1.81	1.32	1.33	0.99	1.42	1.36	-6.12
%Litigious	1.14	1.11	-2.75	1.69	1.70	0.51	0.96	0.94	-1.22	0.52	0.51	-1.19
%Strong Modal	0.54	0.54	-0.46	0.61	0.61	0.07	0.41	0.41	-0.22	0.34	0.32	-1.49
%Constraining	0.79	0.77	-1.74	1.25	1.25	-0.48	0.68	0.67	-1.30	0.66	0.62	-3.29
%Risky	3.44	3.36	-8.21	7.00	6.97	-3.33	2.85	2.84	-0.30	2.41	2.35	-6.04

Table 2: Summary Statistics of Changes in Textual Analysis Measures by Section

Table 2 presents summary statistics for the percentage changes in Loughran and McDonald (2011) word list proportions during the confidential revise-and-resubmit (R&R) process (defined as from DRS to S-1) by major sections of IPO prospectus: Entire Document, Risk Factors, Business, and MD&A. Cosine Similarity and seven Loughran and McDonald (LM) word lists are used: Positive, Negative, Uncertain, Litigious, Strong Modal, Constraining, and Risky. See Appendix for variable definitions. ***,** and * denote significance of t-test of mean difference at the 1%, 5%, and 10% level, respectively.

Panel A: Changes in LM Word List Proportions in Entire Document

	Obs.	Mean	SD	5th	25th	Median	75th	95th
PctChng: Size of Raw File in MB	780	10.92***	13.57	-1.32	3.48	7.89	15.37	30.25
PctChng: Count of Total LM Words	780	10.11***	12.09	0.43	3.85	7.59	13.32	25.66
PctChng: %Positive LM Words	780	-1.62***	4.63	-8.49	-3.75	-1.55	0.37	4.84
PctChng: %Negative LM Words	780	-1.96***	3.62	-7.03	-3.79	-2.03	-0.39	3.24
PctChng: %Uncertain LM Words	780	-2.57***	4.32	-8.46	-4.61	-2.35	-0.54	2.31
PctChng: %Litigious LM Words	780	-2.28***	4.13	-8.82	-4.61	-2.18	-0.14	4.17
PctChng: %Strong Modal LM Words	780	-0.61***	6.14	-8.75	-3.62	-0.91	1.91	9.34
PctChng: %Constraining LM Words	780	-2.10***	3.72	-7.67	-4.08	-2.08	-0.25	3.32
PctChng: %Risky LM Words	780	-2.28***	3.32	-7.06	-4.01	-2.12	-0.62	2.14
Cosine Similarity - Entire Document	780	99.29***	4.54	98.77	99.69	99.86	99.95	99.99

Panel B: Changes in LM Word List Proportions in Risk Section

	Obs.	Mean	SD	5th	25th	Median	75th	95th
PctChng: Size of Raw File in MB	677	3.11***	8.86	-1.78	0.37	1.41	3.98	11.11
PctChng: Count of Total LM Words	677	3.61***	14.12	-1.22	0.44	1.53	3.97	10.63
PctChng: %Positive LM Words	677	-0.40**	5.74	-4.70	-1.55	-0.41	0.27	3.23
PctChng: %Negative LM Words	677	-0.10	7.69	-3.45	-1.18	-0.38	0.14	2.33
PctChng: %Uncertain LM Words	677	0.30	21.65	-3.93	-1.28	-0.46	0.00	1.64
PctChng: %Litigious LM Words	677	0.42	7.46	-3.98	-1.00	-0.12	1.19	7.03
PctChng: %Strong Modal LM Words	677	0.26	4.32	-4.45	-1.04	-0.10	1.14	6.19
PctChng: %Constraining LM Words	677	-0.17	8.38	-4.04	-1.36	-0.45	0.19	3.63
PctChng: %Risky LM Words	677	-0.20	8.05	-3.23	-1.05	-0.43	0.01	1.59
Cosine Similarity - Risk Factors	677	99.62***	3.33	99.45	99.89	99.97	99.99	100.00

Table 2 (Cont.): Summary Statistics of Changes in Textual Analysis Measures by Section

Panel C: Changes in LM Word List Proportions in Business Section

	Obs.	Mean	SD	5th	25th	Median	75th	95th
PctChng: Size of Raw File in MB	622	19.44**	209.22	-0.54	2.38	5.38	10.95	27.89
PctChng: Count of Total LM Words	622	20.98**	246.34	0.00	2.65	5.33	11.09	27.69
PctChng: %Positive LM Words	622	-1.64***	8.84	-10.02	-4.11	-1.74	0.02	5.02
PctChng: %Negative LM Words	622	0.66	12.49	-9.68	-3.24	-0.95	1.10	13.21
PctChng: %Uncertain LM Words	622	1.24***	10.06	-6.93	-1.90	0.05	2.66	10.81
PctChng: %Litigious LM Words	622	2.94	54.38	-12.72	-4.76	-2.09	0.84	17.68
PctChng: %Strong Modal LM Words	622	1.05	19.64	-13.28	-4.88	-1.12	1.94	14.74
PctChng: %Constraining LM Words	622	-0.45	18.91	-11.88	-4.78	-1.95	0.41	10.91
PctChng: %Risky LM Words	622	0.51**	6.48	-6.82	-2.05	-0.42	1.31	9.71
Cosine Similarity - Business	622	99.23***	3.26	98.50	99.55	99.83	99.95	99.98

Panel D: Changes in LM Word List Proportions in MD&A Section

	Obs.	Mean	SD	5th	25th	Median	75th	95th
PctChng: Size of Raw File in MB	499	21.32***	28.10	-2.18	3.48	15.38	30.18	69.13
PctChng: Count of Total LM Words	499	17.82***	23.46	-1.09	4.22	12.38	25.43	56.77
PctChng: %Positive LM Words	499	-1.27**	12.92	-18.45	-8.22	-2.16	2.42	23.73
PctChng: %Negative LM Words	499	0.71	11.97	-14.08	-5.47	-1.26	3.55	23.06
PctChng: %Uncertain LM Words	499	-3.71***	10.31	-19.38	-9.43	-3.37	0.86	10.50
PctChng: %Litigious LM Words	499	-0.62	16.60	-20.52	-8.67	-2.12	4.03	28.81
PctChng: %Strong Modal LM Words	499	-2.52***	18.10	-24.58	-12.50	-3.84	2.56	27.55
PctChng: %Constraining LM Words	499	-4.11***	12.49	-22.14	-10.93	-3.62	0.87	14.93
PctChng: %Risky LM Words	499	-2.04***	8.20	-13.37	-6.80	-2.53	1.42	12.55
Cosine Similarity - MD&A	499	98.74***	2.59	95.66	98.60	99.43	99.85	99.99

Table 3: Market Reactions to Comment Letter Release

Table 3 Panel A presents the regression results using the percentage changes in LM word list proportions in the Business section during the confidential revise-and-resubmit (R&R) process (from DRS to S-1). Panel B reports the subgroup regression results based on the median percentage changes in %Positive in the Business section. Results in each column show cumulative abnormal return (CAR) for various windows since the release of the comment letter (i.e., 5-days, 10-days, and 20-days). CAR is defined as the cumulative difference between firm stock return and corresponding size and book-to-market industry portfolio returns. Due to space limitations, results for control variables and constant are omitted. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

Panel A: Regressions with Cumulative Abnormal Return						
	CAR(0,+5)		CAR(0,+10)		CAR(0,+20)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cosine Similarity	0.0216	-0.0305	-0.0724	-0.0915	0.187*	0.105
	(0.27)	(-0.62)	(-0.73)	(-1.27)	(1.90)	(0.95)
PctChng: %Positive LM Words	0.127***	0.108*	0.156*	0.126	0.355***	0.315**
	(4.12)	(2.28)	(2.14)	(1.45)	(3.62)	(2.57)
PctChng: %Litigious LM Words	0.0134	0.00675	0.00636	-0.00417	-0.00319	-0.0140
	(1.61)	(0.72)	(0.82)	(-0.24)	(-0.38)	(-0.85)
PctChng: %Strong Modal LM Words	-0.0180	-0.0128	-0.0226	-0.0122	-0.0302	-0.0265
	(-1.08)	(-0.44)	(-0.83)	(-0.41)	(-0.71)	(-0.50)
PctChng: %Constraining LM Words	-0.0386**	-0.0337	-0.0154	-0.00396	0.0565***	0.0685**
	(-3.47)	(-1.77)	(-0.96)	(-0.11)	(7.13)	(2.49)
PctChng: %Negative LM Words	0.0210	-0.000160	0.0986**	0.0750	0.129	0.0905
	(0.56)	(-0.00)	(2.55)	(1.21)	(1.28)	(0.85)
S1: %Positive LM Words		0.516		1.892		2.037
		(0.37)		(0.87)		(0.99)
S1: %Litigious LM Words		2.182		2.263		4.141
		(0.94)		(0.80)		(1.16)
S1: %Strong Modal LM Words		3.215		-0.246		-4.027
		(0.68)		(-0.04)		(-0.40)
S1: %Constraining LM Words		-8.086*		-4.440		-9.112
		(-1.93)		(-0.78)		(-1.73)
S1: %Negative LM Words		-0.147		0.700		-1.519
		(-0.11)		(0.20)		(-0.39)
IPO Controls	No	Yes	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	594	594	594	594	594	594
R^2	0.050	0.088	0.038	0.070	0.068	0.098

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3 (Cont.): Market Reactions to Comment Letter Release

Panel B: Regressions by Median Percentage Changes in %Positive in the Business Section

	(1) CAR(0,+3)	(2) CAR(0,+10)	(3) CAR(0,+20)
Cosine Similarity	0.0549	-0.00851	0.130
	(0.55)	(-0.09)	(0.85)
PctChng:%Positive × Above Median	0.126***	0.251***	0.353**
	(3.74)	(4.26)	(2.64)
PctChng:%Positive × Below Median	0.0281	-0.0569	0.260
	(0.22)	(-0.32)	(1.17)
PctChng: %Litigious LM Words	0.00427	-0.00258	-0.0136
	(0.75)	(-0.19)	(-0.82)
PctChng: %Strong Modal LM Words	-0.0268	-0.0246	-0.0302
	(-1.74)	(-0.98)	(-0.63)
PctChng: %Constraining LM Words	-0.00658	-0.0128	0.0659**
	(-0.32)	(-0.32)	(2.47)
PctChng: %Negative LM Words	-0.00429	0.0697	0.0889
	(-0.12)	(1.64)	(0.68)
S1: %Positive LM Words	1.335	2.262	2.149
	(1.32)	(1.04)	(1.09)
S1: %Litigious LM Words	1.185	2.194	4.120
	(0.59)	(0.78)	(1.28)
S1: %Strong Modal LM Words	4.159	-0.873	-4.217
	(1.28)	(-0.14)	(-0.38)
S1: %Constraining LM Words	-6.359	-4.083	-9.004
	(-1.71)	(-0.69)	(-1.73)
S1: %Negative LM Words	-0.188	0.661	-1.531
	(-0.15)	(0.25)	(-0.39)
IPO Controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	594	594	594
R^2	0.102	0.072	0.098

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Regressions with Offer Price Adjustment

Table 4 presents the regression results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit (R&R) process (from DRS to S-1) and the levels of the proportions in S-1 filings. Results in each column show various sections of IPO prospectus used in the textual analysis. Offer Price Adjustment is the percentage change in the offer price from the mid-point of the filing range. Due to space limitations, results for control variables and constant are omitted. See Appendix for full regression results. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

| | Offer Price Adjustment (in %) | | |
| | (1) | (2) | (3) |
	Baseline	Entire Document	Risk Factors
Cosine Similarity		0.0451	-0.263
		(0.39)	(-1.56)
PctChng: %Positive LM Words		-0.101	-0.0310
		(-1.14)	(-0.30)
PctChng: %Litigious LM Words		-0.197	-0.00699
		(-1.13)	(-0.06)
PctChng: %Strong Modal LM Words		-0.0497	0.0806
		(-0.40)	(0.60)
PctChng: %Constraining LM Words		0.184	-0.0802*
		(1.19)	(-1.99)
PctChng: %Negative LM Words		0.0711	-0.128*
		(0.40)	(-2.33)
S1: %Positive LM Words	8.465	10.12	4.402
	(1.46)	(1.81)	(1.11)
S1: %Litigious LM Words	0.910	2.270	-0.550
	(0.39)	(0.71)	(-0.34)
S1: %Strong Modal LM Words	17.31**	18.23**	1.590
	(2.99)	(3.04)	(0.21)
S1: %Constraining LM Words	-13.83**	-15.51**	-0.705
	(-2.46)	(-2.91)	(-0.17)
S1: %Negative LM Words	1.755	1.563	2.425
	(0.61)	(0.52)	(1.28)
IPO Controls	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	773	773	673
R^2	0.229	0.235	0.233

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Regressions with First-day Returns

Table 5 presents the regression results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit (R&R) process (from DRS to S-1) and the levels of the proportions in S-1 filings. Results in each column show various sections of IPO prospectus used in the textual analysis. First-day Returns is the percentage change from the offer price to the closing price on the first trading day. Due to space limitations, results for control variables and constant are omitted. See Appendix for full regression results. The t-statistics are in parentheses with the standard errors clustered by the Fama-French 48-industry and calendar year.

Panel A: Using Entire Document				
	First-day Returns (in %)			
	(1)	(2)	(3)	(4)
Offer Price Adjustment (in %)	0.973***	0.952***	0.928***	0.262**
	(9.11)	(8.07)	(8.70)	(3.15)
Cosine Similarity			0.410***	
			(4.02)	
Cosine Similarity × Upward Price Adj.				0.554***
				(7.77)
Cosine Similarity × Downward Price Adj.				0.285***
				(5.99)
PctChng: %Positive LM Words			-0.0842	-0.0275
			(-0.29)	(-0.09)
PctChng: %Litigious LM Words			-0.531	-0.300
			(-1.75)	(-0.93)
PctChng: %Strong Modal LM Words			-0.253	-0.257
			(-1.48)	(-1.78)
PctChng: %Constraining LM Words			0.510	0.501
			(1.06)	(1.18)
PctChng: %Negative LM Words			-0.169	-0.235
			(-0.46)	(-0.56)
S1: %Positive LM Words		-9.789	-6.173	-9.052
		(-0.86)	(-0.53)	(-0.67)
S1: %Litigious LM Words		5.938	9.997	9.377
		(0.59)	(0.94)	(0.90)
S1: %Strong Modal LM Words		15.99	19.30	15.93
		(1.04)	(1.23)	(1.01)
S1: %Constraining LM Words		-37.76*	-42.97*	-44.12
		(-2.27)	(-1.93)	(-1.89)
S1: %Negative LM Words		16.85	16.68	15.54
		(1.75)	(1.75)	(1.81)
IPO Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	773	773	773	773
R^2	0.302	0.312	0.321	0.373

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 (Cont.): Regressions with First-day Returns

Panel B: Using Major Sections

	First-day Returns (in %)			
	(1)	(2)	(3)	(4)
	Entire Document	Risk Factors	Business	MD&A
Offer Price Adjustment (in %)	0.928***	0.931**	0.919***	1.105***
	(8.70)	(3.23)	(12.30)	(6.39)
Cosine Similarity	0.410***	-0.118	0.199	-0.781*
	(4.02)	(-0.41)	(0.50)	(-2.04)
PctChng: %Positive LM Words	-0.0842	0.335	0.145	0.116
	(-0.29)	(0.85)	(0.83)	(0.54)
PctChng: %Litigious LM Words	-0.531	-0.257	-0.0195	-0.0218
	(-1.75)	(-1.40)	(-0.87)	(-0.04)
PctChng: %Strong Modal LM Words	-0.253	0.119	0.0962	-0.165
	(-1.48)	(0.41)	(1.06)	(-1.28)
PctChng: %Constraining LM Words	0.510	-0.133	0.0994	-0.0623
	(1.06)	(-0.93)	(1.68)	(-0.15)
PctChng: %Negative LM Words	-0.169	-0.227	-0.0983*	-0.0426
	(-0.46)	(-1.01)	(-2.16)	(-0.18)
S1: %Positive LM Words	-6.173	-8.636	2.399	-9.826
	(-0.53)	(-0.73)	(0.52)	(-0.40)
S1: %Litigious LM Words	9.997	-1.315	-4.209	-0.651
	(0.94)	(-0.19)	(-1.52)	(-0.03)
S1: %Strong Modal LM Words	19.30	-7.688	7.384	31.22
	(1.23)	(-0.61)	(0.54)	(1.46)
S1: %Constraining LM Words	-42.97*	-20.75	-4.563	-10.23
	(-1.93)	(-1.54)	(-0.85)	(-0.49)
S1: %Negative LM Words	16.68	8.582	5.084**	5.752
	(1.75)	(1.46)	(2.60)	(0.77)
IPO Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	773	673	617	494
R^2	0.321	0.352	0.348	0.358

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Pre-IPO Operating Performance by %Positive Quintiles in Business Section

Table 6 Panel A and B present the comparison of pre-IPO operating performance based on %Positive quintiles in the Business section of the DRS and the S-1, respectively. Both panels report the mean and median of Annual Sales, EPS, EBIT, and EBITDA from the most recent fiscal year before the IPO. %Positive is defined as the percentage of words in DRS or S-1 that are classified as positive based on Loughran and McDonald (2011) word lists. See the Appendix for other variable definitions.

Panel A: Pre-IPO Operating Performance by %Positive Quintiles in the Business section of DRS

	Lowest		Q_2		Q_3		Q_4		Highest	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
DRS: %Positive LM Words	0.94	0.98	1.19	1.18	1.37	1.36	1.68	1.67	2.27	2.19
Annual Sales (in millions)	57.12	1.94	53.64	0.68	142.06	9.89	322.93	113.13	468.33	172.86
Trailing FY EPS	-0.82	-0.72	-0.87	-0.87	-0.61	-0.75	-0.34	-0.32	-0.19	-0.24
Trailing FY EBIT (in millions)	-9.43	-13.74	-19.41	-19.78	-10.92	-16.53	-3.22	-7.20	4.47	-1.73
Trailing FY EBITDA (in millions)	2.36	-11.83	-15.82	-19.24	-2.81	-15.02	12.33	-2.25	43.86	8.26

Panel B: Pre-IPO Operating Performance by %Positive Quintiles in the Business section of S-1

	Lowest		Q_2		Q_3		Q_4		Highest	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
S1: %Positive LM Words	0.92	0.96	1.16	1.16	1.34	1.34	1.65	1.63	2.24	2.17
Annual Sales (in millions)	57.26	2.33	45.54	0.05	146.99	10.65	266.97	124.89	527.28	171.20
Trailing FY EPS	-0.84	-0.73	-0.82	-0.81	-0.65	-0.76	-0.41	-0.39	-0.11	-0.16
Trailing FY EBIT (in millions)	-10.05	-14.45	-19.33	-19.49	-12.02	-16.67	-14.78	-10.05	17.65	0.80
Trailing FY EBITDA (in millions)	1.56	-11.09	-17.83	-19.24	-1.57	-14.76	-0.59	-3.94	58.68	10.06

Table 7: Regressions with Post-IPO Operating Performance Using Changes in the Business Section

Table 7 presents the regression results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit (R&R) process (from DRS to S-1) in the Business section. Results in each column show measures of post-IPO operating performance for various periods. All operating performance measures have been scaled by lagged total assets and winsorized at the 1% level. Due to space limitations, results for control variables and constant are omitted. See Appendix for full regression results. The t-statistics are in parentheses with the standard errors clustered by the Fama-French 48-industry.

	EBITDA/Asset (in %)		Net Income/Asset (in %)	
	(1)	(2)	(3)	(4)
	6 months	12 months	6 months	12 months
Cosine Similarity	-0.0476	-0.0447	-0.0339	0.0364
	(-0.57)	(-0.43)	(-0.53)	(0.37)
PctChng: %Positive LM Words	0.0722*	0.100*	0.0795**	0.151***
	(1.90)	(1.99)	(2.60)	(3.26)
PctChng: %Litigious LM Words	-0.00175	-0.000461	-0.00481	0.00610
	(-0.16)	(-0.05)	(-0.63)	(0.77)
PctChng: %Strong Modal LM Words	-0.0314	-0.00599	-0.0313	-0.0166
	(-1.21)	(-0.45)	(-1.23)	(-1.26)
PctChng: %Constraining LM Words	0.0236	0.00327	0.0348***	0.00109
	(1.20)	(0.17)	(3.04)	(0.06)
PctChng: %Negative LM Words	0.0322	0.0155	0.00956	0.00271
	(1.13)	(0.60)	(0.74)	(0.10)
S1: %Positive LM Words	2.948*	2.546	2.710*	2.108
	(1.70)	(1.50)	(1.81)	(1.18)
S1: %Litigious LM Words	3.825**	3.601**	1.183*	1.335*
	(2.51)	(2.74)	(1.71)	(1.75)
S1: %Strong Modal LM Words	-3.233	-4.933	-2.652	-1.774
	(-0.70)	(-1.42)	(-0.50)	(-0.42)
S1: %Constraining LM Words	-2.795	-3.006	-1.670	0.0445
	(-0.87)	(-1.00)	(-0.76)	(0.02)
S1: %Negative LM Words	-1.355	-1.760	0.175	-0.149
	(-0.83)	(-1.19)	(0.27)	(-0.17)
Constant	-3.400	1.635	-7.372	-9.787
	(-0.26)	(0.13)	(-0.82)	(-0.85)
IPO Controls	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	559	512	559	512
R^2	0.495	0.485	0.435	0.413

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Summary Statistics for Withdrawn IPOs, 2013-2020

The sample includes 872 IPOs, of which 719 were completed while 153 were withdrawn in 2013-2020. All IPOs have gone through the confidential revise-and-resubmit (R&R) process (defined as from DRS to S-1). Withdrawn IPOs are defined as those that have filed Form RW (registration withdrawn) and have not subsequently received an Effective Notice from SEC. Panel A reports summary statistics for the overall IPO registration process. Panel B presents summary statistics of Loughran-McDonald (LM) word list levels in S-1 filings. Panel C presents summary statistics of the percentage changes in LM word list proportions during the R&R process. The comparison of mean differences between withdrawn and completed IPOs is shown in columns. ***,**, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A: IPO Registration Process

	Obs.	Completed IPO	Withdrawn IPO	Differences
Number of R&R round	872	2.5	2.4	-0.1
Number of days per R&R round	872	41.1	49.9	8.8***
Days DRS to S1	872	106.0	119.4	13.4
Days S1 to IPO	872	51.0	290.6	239.6***
Number of DRS Amendments	872	1.5	1.4	-0.1
Number of S1 Amendments	872	2.8	2.2	-0.6***

Panel B: LM Word List Levels in S-1 Filings

	Obs.	Completed IPO	Withdrawn IPO	Differences
S1: %Positive LM Words	872	0.84	0.77	-0.07***
S1: %Litigious LM Words	872	1.12	1.16	0.04**
S1: %Strong Modal LM Words	872	0.54	0.60	0.06***
S1: %Constraining LM Words	872	0.77	0.78	0.01
S1: %Negative LM Words	872	1.82	1.72	-0.10***
S1: %Uncertain LM Words	872	1.60	1.54	-0.06***
S1: %Risky LM Words	872	3.39	3.23	-0.16***

Panel C: Changes in LM Word List Proportions During the R&R Process

	Obs.	Completed IPO	Withdrawn IPO	Differences
PctChng: Size of Raw File in MB	872	10.55	58.88	48.33**
PctChng: Count of Total LM Words	872	9.72	61.04	51.32**
PctChng: %Positive LM Words	872	-1.65	-0.83	0.81*
PctChng: %Litigious LM Words	872	-2.27	-1.34	0.93*
PctChng: %Strong Modal LM Words	872	-0.52	0.17	0.69
PctChng: %Constraining LM Words	872	-2.19	-2.55	-0.36
PctChng: %Negative LM Words	872	-2.02	1.12	3.14**
PctChng: %Uncertain LM Words	872	-2.54	-0.86	1.68*
PctChng: %Risky LM Words	872	-2.29	0.01	2.30**

Table 9: Logit Regressions with Withdrawal Decision

The sample includes 872 IPOs, of which 719 were completed while 153 were withdrawn in 2013-2020. All IPOs have gone through the confidential revise-and-resubmit (R&R) process (defined as from DRS to S-1). Firms that went public after previous withdrawn are excluded. Panel A reports logit regression results using the level of word list proportions in the entire document of DRS or S-1. Panel B reports the results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit (R&R) process. RW dummy is set to one for withdrawn IPO and zero for completed IPO. See the Appendix for other variable definitions. All regressions include an intercept, the Fama-French 48-industry and calendar year fixed effects. The z-statistics are in parentheses with the standard errors clustered by withdrawn year.

Panel A: Using LM Word List Levels		
	Withdrawal Dummy	
	(1)	(2)
	Form DRS	Form S-1
	(Entire Document)	(Entire Document)
Ln Number of R&R round	-0.784**	-0.840***
	(-2.37)	(-2.86)
Ln Days DRS to S1	0.567**	0.616***
	(2.44)	(2.73)
%Positive LM Words	-1.575	-1.905
	(-1.30)	(-1.38)
%Litigious LM Words	1.883***	1.947***
	(3.10)	(2.88)
%Strong Modal LM Words	4.088***	4.098***
	(3.17)	(3.21)
%Constraining LM Words	-1.200	-1.883
	(-0.80)	(-1.03)
%Negative LM Words	-1.162**	-1.243*
	(-2.28)	(-1.89)
Constant	-6.554***	-5.923***
	(-3.78)	(-3.46)
IPO Controls	No	No
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	872	872
Pseudo R^2	0.207	0.209

$* \ p < 0.10, \ ** \ p < 0.05, \ *** \ p < 0.01$

Table 9 (Cont.): Logit Regressions with Withdrawal Decision

Panel B: Using Changes in Word List Proportions During R&R		
	Withdrawal Dummy	
	(1)	(2)
	Entire Document	MD&A
Ln Number of R&R round	-0.744**	-0.235
	(-2.00)	(-0.45)
Ln Days DRS to S1	0.496*	0.301
	(1.82)	(0.84)
Cosine Similarity	-0.038***	-0.062***
	(-2.63)	(-2.89)
PctChng: %Positive LM Words	0.007	0.010
	(0.37)	(0.91)
PctChng: %Litigious LM Words	0.012	-0.009
	(0.93)	(-0.81)
PctChng: %Strong Modal LM Words	-0.016	-0.006
	(-0.87)	(-1.18)
PctChng: %Constraining LM Words	-0.037*	0.008
	(-1.88)	(0.50)
PctChng: %Negative LM Words	0.007	-0.016
	(0.73)	(-1.08)
S1: %Positive LM Words	-1.665	-0.042
	(-1.24)	(-0.05)
S1: %Litigious LM Words	1.752***	1.072
	(2.58)	(1.35)
S1: %Strong Modal LM Words	4.057***	0.069
	(3.08)	(0.04)
S1: %Constraining LM Words	-1.479	1.113
	(-0.74)	(1.34)
S1: %Negative LM Words	-1.141*	0.577
	(-1.75)	(0.74)
Constant	-1.993	2.514
	(-0.71)	(1.26)
IPO Controls	No	No
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	872	481
Pseudo R^2	0.220	0.181

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Comparison of LM Word List Proportions by Section

Table 10 compares the levels of word list proportions in MD&A to that of in Risk Factors, Business and the Entire Document. Each category of LM word lists is defined as the percentage of words that are classified in that category based on Loughran and McDonald (2011) word lists.

	MD&A Mean (in %)	Risk Factors Mean (in %)	Business Mean (in %)	Entire Document Mean (in %)
In DRS Filings				
%Positive	0.68	1.17	1.49	0.85
%Negative	1.05	3.89	1.54	1.84
%Uncertain	1.42	3.17	1.32	1.63
%Litigious	0.52	1.69	0.96	1.14
%Strong Modal	0.34	0.61	0.41	0.54
%Constraining	0.66	1.25	0.68	0.79
%Risky	2.41	7.00	2.85	3.44
In S-1 Filings				
%Positive	0.67	1.16	1.46	0.84
%Negative	1.05	3.88	1.53	1.81
%Uncertain	1.36	3.16	1.33	1.59
%Litigious	0.51	1.70	0.94	1.11
%Strong Modal	0.32	0.61	0.41	0.54
%Constraining	0.62	1.25	0.67	0.77
%Risky	2.35	6.97	2.84	3.36

Table 11: The Impacts of JOBS Act on IPO Market

Table 11 Panel A and B show the impacts of the confidential revise-and-resubmit (R&R) process on IPO market in terms of IPO process time and completion rate, respectively. Pre-JOBS Act includes 2680 IPOs, of which 1887 were completed while 793 were withdrawn during 1997-2010. The Post-JOBS Act period includes 872 IPOs, of which 719 were completed while 153 were withdrawn during 2013-2020. Confidential Review Period is defined as the number of calendar days from filing DRS to filing S-1 with SEC; Public Review Period is defined as the number of calendar days from filing S-1 to receiving Effective Notice issued by SEC; Entire IPO Registration Process is the sum of the confidential and public review period.

Panel A: Impact on IPO Registration Process Time

(in calendar days)	Pre-JOBS Act (1997-2010)				Post-JOBS Act (2013-2020)				Diff.
	Obs.	Mean	SD	Median	Obs.	Mean	SD	Median	
Period of Confidential Review	1887	-	-	-	719	106	102	71	+106
Period of Public Review	1887	117	98	88	719	51	64	31	-66
Entire IPO Registration Process	1887	117	98	88	719	157	118	115	+40

Panel B: Impact on IPO Completion Rate

	Obs.	Completed	Withdrawn	Completion Rate
Pre-JOBS Act (1997-2010)	2680	1887	793	70.4%
Post-JOBS Act (2013-2020)	872	719	153	82.5%

Table 12: Determinants of Withdrawal Decision

Table 12 Panel A shows the comparison of logit regressions on the withdrawal decision using the levels of word list proportions, the percentage change in word list proportions, and the interactions of both. Panel B reports the average marginal effects of previous logit regressions. See the Appendix for other variable definitions. The z-statistics are in parentheses with the standard errors clustered by withdrawn year.

<div align="center">Panel A: Logit Regressions Results</div>

	Withdrawal Dummy		
	(1)	(2)	(3)
	Levels in S-1	Changes in R&R	With Interactions
Ln Number of R&R round	-0.840***	-0.744**	-0.694*
	(-2.86)	(-2.00)	(-1.96)
Ln Days DRS to S1	0.616***	0.496*	0.467*
	(2.73)	(1.82)	(1.77)
S1: %Positive	-1.905	-1.665	-1.821
	(-1.38)	(-1.24)	(-1.39)
S1: %Litigious	1.947***	1.752***	1.795***
	(2.88)	(2.58)	(2.59)
S1: %Strong Modal	4.098***	4.057***	4.016***
	(3.21)	(3.08)	(2.92)
S1: %Constraining	-1.883	-1.479	-1.137
	(-1.03)	(-0.74)	(-0.52)
S1: %Negative	-1.243*	-1.141*	-1.029**
	(-1.89)	(-1.75)	(-2.13)
Cosine Similarity		-0.038***	-0.036***
		(-2.63)	(-3.03)
PctChng: %Positive		0.007	0.100
		(0.37)	(0.56)
PctChng: %Litigious		0.012	0.022
		(0.93)	(0.21)
PctChng: %Strong Modal		-0.016	0.010
		(-0.87)	(0.07)
PctChng: %Constraining		-0.037*	-0.162
		(-1.88)	(-0.95)
PctChng: %Negative		0.007	-0.121
		(0.73)	(-0.51)
S1 Level × PctChng: %Positive			-0.116
			(-0.48)
S1 Level × PctChng: %Litigious			0.071
			(0.46)
S1 Level × PctChng: %Strong Modal			-0.006
			(-0.08)
S1 Level × PctChng: %Constraining			-0.046
			(-0.21)
S1 Level × PctChng: %Negative			0.161
			(0.76)
IPO Controls	No	No	No
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	872	872	872
pseudo R^2	0.209	0.220	0.222

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12 (Cont.): Determinants of Withdrawal Decision

Panel B: Average Marginal Effects		
	(1)	(2)
	Levels in S-1	Changes in R&R
Ln Number of R&R round	-0.095***	-0.083**
	(-3.02)	(-2.03)
Ln Days DRS to S1	0.070***	0.055*
	(2.88)	(1.83)
S1: %Positive	-0.216	-0.186
	(-1.39)	(-1.25)
S1: %Litigious	0.221***	0.195***
	(2.99)	(2.62)
S1: %Strong Modal	0.466***	0.452***
	(3.26)	(3.21)
S1: %Constraining	-0.214	-0.165
	(-1.03)	(-0.73)
S1: %Negative	-0.141*	-0.127*
	(-1.95)	(-1.79)
Cosine Similarity		-0.004***
		(-2.66)
PctChng: %Positive		0.001
		(0.37)
PctChng: %Litigious		0.001
		(0.92)
PctChng: %Strong Modal		-0.002
		(-0.89)
PctChng: %Constraining		-0.004*
		(-1.81)
PctChng: %Negative		0.001
		(0.73)
IPO Controls	No	No
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	872	872

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix A: Variable Definitions

Variable Name	Definition
Offer Price Adjustment	The percentage change in the offer price from the mid-point of the filing price range.
First-day Returns	The percentage change from the offer price to the closing price on the first trading day.
VC-backed Dummy	Dummy variable set to one if the IPO is backed by venture capital, else zero.
Top-tier Underwriter Dummy	Dummy variable set to one if the lead underwriter of the IPO has an updated ranking score of eight or more, as in Loughran and Ritter (2004).
Positive EPS Dummy	Dummy variable set to one if the EPS in most recent fiscal year is positive.
Annual Sales	Firm annual sales in most recent fiscal year prior to IPO, in millions of dollars.
EPS	Firm earnings per share in most recent fiscal year prior to IPO, in dollars.
EBIT	Firm earnings before interest and taxes in most recent fiscal year prior to IPO, in millions of dollars.
EBITDA	Firm earnings before interest, taxes, depreciation, and amortization in most recent fiscal year prior to IPO, in millions of dollars.
Share Offered	The number of shares offered divided by the total number of shares outstanding after offering, in percentage.
Market Returns	The monthly returns of the CRSP Nasdaq value-weighted index prior to the IPO month, in percentage.
Period of Confidential Review	The number of calendar days between filing DRS and filing S-1 on EDGAR.
Period of Public Review	The number of calendar days between filing S-1 and receiving Effective Notice on EDGAR.
Entire IPO Registration Process	The number of calendar days between filing DRS and receiving Effective Notice on EDGAR.
Number of DRS Amendments	The number of DRS amendments associated with the initial DRS filed on EDGAR during the IPO registration process.
Number of S-1 Amendments	The number of S-1 amendments associated with the initial S-1 filed on EDGAR during the IPO registration process.
Number of IPO Amendments	The total number of all DRS and S-1 amendments filed on EDGAR during the IPO registration process.

Variable Name	Definition
File Size	The size of the *.txt* file, in megabyte (MB).
Number of LM Words	The number of words appear in the Loughran and McDonald (2011) word list dictionary.
Word Complexity	The average syllables per word for the words appear in the Loughran and McDonald (2011) word list dictionary.
%Positive	Percentage of words in DRS, or S-1, or specific section that are classified as positive based on Loughran and McDonald (2011) word list. Examples of positive words include accomplish, enable, good, stable, success.
%Negative	Percentage of words in DRS, or S-1, or specific section that are classified as negative based on Loughran and McDonald (2011) word list. Examples of negative words include loss, failure, closing, terminate, deficit.
%Uncertain	Percentage of words in DRS, or S-1, or specific section that are classified as uncertain based on Loughran and McDonald (2011) word list. Examples of uncertain words include approximate, uncertain, possible, unsure, assumed.
%Litigious	Percentage of words in DRS, or S-1, or specific section that are classified as legal based on Loughran and McDonald (2011) word list. Examples of legal words include settlement, plaintiff, bail, investigate, testify.
%Strong-modal	Percentage of words in DRS, or S-1, or specific section that are classified as strong-modal based on Loughran and McDonald (2011) word list. Examples of strong modal words include always, clearly, definitely, must, never.
%Constraining	Percentage of words in DRS, or S-1, or specific section that are classified as constraining based on Loughran and McDonald (2011) word list. Examples of constraining words include bond, covenant, forbid, impose, restrict.
%Risky	Percentage of words in DRS, or S-1, or specific section that are classified as negative or uncertain based on Loughran and McDonald (2011). Defined as %Negative plus %Uncertain.
Change in %LM Word Lists	Defined as the %LM word list in S-1 minus the %LM word list in DRS for each category of LM word lists in DRS, or S-1, or specific section, in basis points.
Cosine Similarity	Defined as the cosine angle between two document vectors on a unit sphere, calculated using Eq.1.

Appendix B: Regressions with Offer Price Adjustment and First-day Returns Using Changes of Word List Proportion During R&R Process

This table presents the regression results using the percentage changes in LM word list proportions during the confidential revise-and-resubmit (R&R) process (from DRS to S-1). Results in each column show various sections used in textual analysis. See Appendix A for variable definitions. All regressions include an intercept, Fama-French 48-industry and calendar year fixed effects. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

	Offer Price Adjustment (in %)				First-day Returns (in %)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Entire Document	Risk Factors	Business	MD&A	Entire Document	Risk Factors	Business	MD&A
Cosine Similarity	0.0451	-0.263	-0.191	0.0970	0.410***	-0.118	0.199	-0.781*
	(0.39)	(-1.56)	(-1.22)	(0.42)	(4.02)	(-0.41)	(0.50)	(-2.04)
PctChng: %Positive LM Words	-0.101	-0.0310	-0.0566	-0.0405	-0.0842	0.335	0.145	0.116
	(-1.14)	(-0.30)	(-0.62)	(-0.90)	(-0.29)	(0.85)	(0.83)	(0.54)
PctChng: %Litigious LM Words	-0.197	-0.00699	-0.0254**	-0.0529	-0.531	-0.257	-0.0195	-0.0218
	(-1.13)	(-0.06)	(-2.48)	(-1.02)	(-1.75)	(-1.40)	(-0.87)	(-0.04)
PctChng: %Strong Modal LM Words	-0.0497	0.0806	-0.0571	-0.00138	-0.253	0.119	0.0962	-0.165
	(-0.40)	(0.60)	(-1.70)	(-0.06)	(-1.48)	(0.41)	(1.06)	(-1.28)
PctChng: %Constraining LM Words	0.184	-0.0802*	0.101***	0.0733	0.510	-0.133	0.0994	-0.0623
	(1.19)	(-1.99)	(3.67)	(0.98)	(1.06)	(-0.93)	(1.68)	(-0.15)
PctChng: %Negative LM Words	0.0711	-0.128*	0.0000710	-0.0394	-0.169	-0.227	-0.0983*	-0.0426
	(0.40)	(-2.33)	(0.00)	(-0.48)	(-0.46)	(-1.01)	(-2.16)	(-0.18)
S1: %Positive LM Words	10.12	4.402	5.647**	2.524	-6.173	-8.636	2.399	-9.826
	(1.81)	(1.11)	(2.82)	(1.28)	(-0.53)	(-0.73)	(0.52)	(-0.40)
S1: %Litigious LM Words	2.270	-0.550	1.395	-1.589	9.997	-1.315	-4.209	-0.651
	(0.71)	(-0.34)	(1.02)	(-0.62)	(0.94)	(-0.19)	(-1.52)	(-0.03)
S1: %Strong Modal LM Words	18.23**	1.590	10.27*	8.182	19.30	-7.688	7.384	31.22
	(3.04)	(0.21)	(2.10)	(0.92)	(1.23)	(-0.61)	(0.54)	(1.46)
S1: %Constraining LM Words	-15.51**	-0.705	-4.601	-5.603	-42.97*	-20.75	-4.563	-10.23
	(-2.91)	(-0.17)	(-1.31)	(-1.20)	(-1.93)	(-1.54)	(-0.85)	(-0.49)
S1: %Negative LM Words	1.563	2.425	1.255	-2.516	16.68	8.582	5.084**	5.752
	(0.52)	(1.28)	(1.11)	(-0.94)	(1.75)	(1.46)	(2.60)	(0.77)
Offer Price Adjustment (in %)					0.928***	0.931**	0.919***	1.105***
					(8.70)	(3.23)	(12.30)	(6.39)
VC-backed dummy	2.463**	2.730	3.068***	2.279*	6.757*	5.710	7.686	5.521
	(3.44)	(1.78)	(3.59)	(2.14)	(2.09)	(1.22)	(1.44)	(0.69)
Top-tier Underwriter dummy	3.305***	3.069***	2.418**	3.154*	1.822	1.089	2.769	1.337
	(4.07)	(4.48)	(3.30)	(2.24)	(0.48)	(0.38)	(0.87)	(0.37)
Positive EPS dummy	0.661	0.879	0.501	-0.377	1.268	0.345	-0.899	2.454
	(0.47)	(0.45)	(0.34)	(-0.22)	(0.40)	(0.06)	(-0.21)	(0.89)
Annual Sales (in logs)	-0.0411	-0.00228	-0.0270	-0.0808	-0.0181	-0.227	-0.229	-0.00218
	(-0.55)	(-0.04)	(-0.38)	(-0.88)	(-0.07)	(-0.71)	(-1.21)	(-0.01)
Prior Month Nasdaq Return (in %)	-0.0330	-0.0509	-0.103	-0.109	0.703	0.986	0.934	0.889
	(-0.28)	(-0.32)	(-0.86)	(-1.04)	(1.41)	(1.75)	(1.68)	(1.49)
Share Offered (in %)	-0.0547**	-0.0446**	-0.0222	-0.0176	-0.152	-0.131	-0.169	-0.127
	(-2.79)	(-2.96)	(-0.73)	(-0.76)	(-1.44)	(-1.45)	(-1.81)	(-1.29)
Ln Days DRS to S1	-1.349	-1.721	-0.889	-2.683	1.585	4.048**	4.237**	2.501
	(-1.57)	(-1.27)	(-0.77)	(-1.71)	(0.70)	(2.58)	(2.47)	(0.28)
Ln Number of R&R round	-0.184	-0.00577	-0.829	1.092	-2.004	-5.184	-4.963*	-3.827
	(-0.14)	(-0.00)	(-0.54)	(0.99)	(-0.58)	(-1.89)	(-1.97)	(-0.45)
Constant	-12.57	17.15	7.736	1.373	-38.34	30.52***	-19.48	90.14***
	(-0.83)	(0.86)	(0.45)	(0.05)	(-1.28)	(3.63)	(-0.46)	(5.42)
Observations	773	673	617	494	773	673	617	494
R^2	0.235	0.233	0.261	0.279	0.321	0.352	0.348	0.358

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C: Regressions with Offer Price Adjustment and First-day Returns Using Word List Proportions in DRS

This table presents the regression results using LM word list proportions in the DRS filings. Results in each column show various sections used in textual analysis (i.e., Entire Document, Risk Factors, Business, and MD&A). Offer Price Adjustment is the percentage change in the offer price from the mid-point of the filing range. First-day Returns is the percentage change from the offer price to the closing price on the first trading day. See Appendix for other variable definitions. All regressions include an intercept, Fama-French 48-industry and calendar year fixed effects. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

	Offer Price Revision (in %)				First-day Returns (in %)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Entire Document	Risk Factors	Business	MD&A	Entire Document	Risk Factors	Business	MD&A
DRS: %Positive	9.348	4.685	5.570*	2.764*	-7.166	-9.082	1.857	-8.232
	(1.85)	(1.34)	(2.27)	(2.02)	(-0.57)	(-1.41)	(0.46)	(-1.40)
DRS: %Litigious	3.081	0.0516	1.764	1.645	10.67	-1.641	-2.824	5.099
	(0.97)	(0.03)	(1.13)	(0.55)	(1.07)	(-0.99)	(-0.88)	(0.55)
DRS: %Strong Modal	17.03**	1.409	9.910	7.424	25.48*	-9.076	1.409	40.15
	(2.66)	(0.23)	(1.74)	(0.91)	(2.06)	(-0.74)	(0.10)	(1.78)
DRS: %Constraining	-15.01**	-1.291	-4.971	-6.844	-41.27**	-18.54	-8.653	-7.595
	(-3.22)	(-0.33)	(-1.49)	(-1.35)	(-2.48)	(-1.65)	(-1.24)	(-1.23)
DRS: %Negative	1.146	2.468	1.222	-2.958	16.57	8.944**	5.426**	5.634
	(0.49)	(1.54)	(1.07)	(-1.25)	(1.86)	(2.56)	(2.48)	(1.10)
Offer Price Adjustment (in %)					0.935***	0.931***	0.922***	1.093***
					(8.62)	(11.07)	(8.97)	(9.15)
VC-backed dummy	2.531***	2.553**	2.874***	2.094	7.103**	5.695	7.754	6.668
	(4.60)	(2.75)	(4.57)	(1.84)	(2.37)	(1.37)	(1.37)	(1.23)
Top-tier Underwriter dummy	3.254***	3.129***	2.663***	2.966*	1.904	1.156	2.677	1.175
	(4.19)	(4.46)	(3.55)	(2.35)	(0.48)	(0.39)	(0.85)	(0.35)
Positive EPS dummy	0.615	1.058	0.547	-0.434	0.885	0.0813	-1.357	0.966
	(0.49)	(0.73)	(0.35)	(-0.32)	(0.24)	(0.02)	(-0.27)	(0.35)
Annual Sales (in logs)	-0.0316	-0.00192	-0.0361	-0.0929	-0.00415	-0.232	-0.269	-0.0208
	(-0.49)	(-0.03)	(-0.56)	(-1.21)	(-0.02)	(-1.47)	(-1.26)	(-0.06)
Prior Month Nasdaq Return (in %)	-0.0429	-0.0430	-0.0865	-0.107	0.686	0.994*	0.934	0.864
	(-0.38)	(-0.40)	(-0.80)	(-1.16)	(1.43)	(2.01)	(1.73)	(1.77)
Share Offered (in %)	-0.0525**	-0.0418**	-0.0195	-0.0132	-0.145	-0.131	-0.165	-0.143
	(-2.67)	(-2.40)	(-0.72)	(-0.53)	(-1.40)	(-1.60)	(-1.10)	(-1.26)
Ln Days DRS to S1	-1.508	-1.704	-0.868	-3.035*	1.118	3.624*	3.888**	2.397
	(-1.80)	(-1.39)	(-0.76)	(-2.04)	(0.52)	(1.99)	(2.44)	(1.42)
Ln Number of R&R round	0.0671	0.0411	-0.820	1.475	-0.983	-4.625*	-4.659	-3.105
	(0.04)	(0.02)	(-0.64)	(1.51)	(-0.33)	(-2.35)	(-1.12)	(-0.46)
Constant	-6.843	-9.939	-11.50	11.76	-0.549	17.75	5.317	4.778
	(-0.77)	(-0.94)	(-1.31)	(1.26)	(-0.02)	(0.79)	(0.59)	(0.71)
Observations	773	673	617	494	773	673	617	494
R^2	0.232	0.231	0.248	0.274	0.316	0.350	0.344	0.354

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D: Regressions with Offer Price Adjustment and First-day Returns Using Word List Proportions in S-1

This table presents the regression results using LM word list proportions in the S-1 filings. Results in each column show various sections used in textual analysis (i.e., Entire Document, Risk Factors, Business, and MD&A). Offer Price Adjustment is the percentage change in the offer price from the mid-point of the filing range. First-day Returns is the percentage change from the offer price to the closing price on the first trading day. See Appendix for other variable definitions. All regressions include an intercept, Fama-French 48-industry and calendar year fixed effects. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

	Offer Price Revision (in %)				First-day Returns (in %)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Entire Document	Risk Factors	Business	MD&A	Entire Document	Risk Factors	Business	MD&A
S1: %Positive	8.465	4.094	5.759**	2.410	-9.789	-9.085	2.810	-9.537
	(1.46)	(1.11)	(2.80)	(1.63)	(-0.86)	(-1.19)	(0.57)	(-1.41)
S1: %Litigious	0.910	-0.223	1.180	-2.616	5.938	-1.207	-3.464	-0.440
	(0.39)	(-0.15)	(0.77)	(-1.01)	(0.59)	(-0.46)	(-1.13)	(-0.04)
S1: %Strong Modal	17.31**	3.017	10.11*	7.436	15.99	-8.146	7.526	25.24
	(2.99)	(0.47)	(1.93)	(1.01)	(1.04)	(-0.66)	(0.59)	(1.10)
S1: %Constraining	-13.83**	-0.932	-3.903	-3.888	-37.76*	-21.89	-4.870	-12.80*
	(-2.46)	(-0.22)	(-1.43)	(-0.94)	(-2.27)	(-1.89)	(-0.82)	(-2.08)
S1: %Negative	1.755	2.473	1.112	-2.969	16.85	8.051*	4.920**	5.544
	(0.61)	(1.33)	(0.98)	(-1.17)	(1.75)	(2.23)	(2.57)	(0.92)
Offer Price Adjustment (in %)					0.952***	0.936***	0.917***	1.098***
					(8.07)	(10.97)	(9.51)	(7.84)
VC-backed dummy	2.418***	2.678**	2.880***	2.297*	7.038*	5.495	7.775	5.869
	(4.81)	(2.73)	(4.13)	(2.33)	(2.23)	(1.38)	(1.40)	(1.18)
Top-tier Underwriter dummy	3.315***	3.206***	2.723***	3.090*	1.935	1.028	2.701	0.960
	(4.17)	(4.60)	(3.97)	(2.31)	(0.51)	(0.35)	(0.89)	(0.25)
Positive EPS dummy	0.632	0.911	0.538	-0.157	0.501	-0.0208	-1.172	0.824
	(0.50)	(0.59)	(0.35)	(-0.11)	(0.15)	(-0.01)	(-0.26)	(0.28)
Annual Sales (in logs)	-0.0392	0.00280	-0.0371	-0.0817	-0.0253	-0.226	-0.253	-0.0332
	(-0.60)	(0.04)	(-0.57)	(-1.32)	(-0.10)	(-1.44)	(-1.25)	(-0.10)
Prior Month Nasdaq Return (in %)	-0.0313	-0.0434	-0.0813	-0.0809	0.702	0.995*	0.942	0.912
	(-0.25)	(-0.43)	(-0.66)	(-0.82)	(1.45)	(1.97)	(1.75)	(1.61)
Share Offered (in %)	-0.0536**	-0.0432**	-0.0201	-0.0176	-0.142	-0.127	-0.164	-0.131
	(-2.68)	(-2.61)	(-0.79)	(-0.83)	(-1.34)	(-1.54)	(-1.25)	(-1.11)
Ln Days DRS to S1	-1.463	-1.652	-0.951	-2.880*	1.031	3.448	3.677**	2.311
	(-1.62)	(-1.40)	(-0.85)	(-1.93)	(0.47)	(1.86)	(2.71)	(1.55)
Ln Number of R&R round	0.104	0.0657	-0.635	1.308	-0.776	-4.504*	-4.629	-3.236
	(0.07)	(0.04)	(-0.48)	(1.51)	(-0.26)	(-2.15)	(-1.40)	(-0.52)
Constant	-5.951	-10.46	-11.51	11.57	9.247	24.86	1.054	17.65**
	(-0.63)	(-0.84)	(-1.54)	(1.50)	(0.34)	(1.00)	(0.09)	(2.53)
Observations	773	673	617	494	773	673	617	494
R^2	0.229	0.230	0.247	0.272	0.312	0.349	0.343	0.350

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E: Examples of Changes in Negative Words Used

This table reports the frequency of representative negative words (i.e., "damage", "fail", "harm", "penalty", "unable") used in DRS and S-1 filings of We Co. The number of each word used in the entire document and in the Risk Factors section are shown in the columns for both DRS and S-1 filings. The changes in the number of each word used after the confidential revise-and-resubmit (R&R) process are reported in the last two columns.

Frequency of Representative Negative Words Used for We Co.

Negative Words Used	DRS		S-1		Changes after R&R	
	Entire Document	Risk Factors	Entire Document	Risk Factors	Entire Document	Risk Factors
damage	8	6 (75%)	15	13 (87%)	+7	+7 (100%)
fail	17	15 (88%)	27	24 (89%)	+10	+9 (90%)
harm	16	14 (88%)	24	22 (92%)	+8	+8 (100%)
penalty	5	3 (60%)	7	5 (71%)	+2	+2 (100%)
unable	19	19 (100%)	26	26 (100%)	+7	+7 (100%)

Appendix F: Steps to Extract Specific Sections in the Prospectus

The following steps describe and explain the approach used in this study to efficiently break the entire document into sections. This approach addresses several practical issues mentioned in Loughran and McDonald (2016) with relatively high accuracy and speed. Note that this approach requires the prospectus to have a unique "Table of Contents" that contains page numbers. Further improvements may be achieved by additional manual checks and a combination of multiple screening criteria.

- **Step 1: Extract "Table of Contents" from the prospectus**
  The "Table of Contents" containing the starting page number of each section is unique and usually in table format. It appears at the beginning part of the prospectus after the list of underwriters. Although there are multiple ways to locate and extract the "Table of Contents," regular expression is recommended.

- **Step 2: Extract section name and its starting page number**
  Once "Table of Contents" is extracted, parsing by rows will generate a list of section names, as well as the corresponding starting page numbers. The ending page number of each section can be obtained using the starting page number of the next section minus one. It is worth noting that the sequence of parsing matters. Otherwise, the page numbers may be mismatched.

- **Step 3: Construct a reference table**
  A section reference table can be constructed using the extracted section names, starting page number, and ending page number. Given necessary cleaning, the reference table will indicate the section name and its page range and be in a searchable format.

- **Step 4: Reconstruct pages**
  Since pagination does not align with the psychical pages in a typical prospectus, it is useful to break the entire prospectus into several "text blocks" and then reconstruct the actual pages. This approach utilizes a formatting tag in HTML to efficiently divide the entire prospectus. In HTML, $\langle hr \rangle$ represents a "thematic break" between paragraph-level elements to indicate a change of scene in a story or a topic shift within a section. Dividing the entire prospectus through the $\langle hr \rangle$ tag will result in several text blocks with the numerical page number at or near each block's end. This approach avoids practical issues when extracting page numbers, as parsing within a shortened text block area increases the accuracy of identifying page numbers.

- **Step 5: Reconstruct sections**
  Given necessary cleaning, there should be text blocks that have page numbers. Then, each section can be reconstructed according to the reference table created in Step 3. Specifically, corresponding text blocks can be selected and concatenated according to the page range of a particular section.

- **Note:** After all steps, the final outputs for a given prospectus should include a section reference table, text blocks with page numbers, and reconstructed sections with names and text content.