Does the Confidential IPO Registration Process Create Value?*

Mengnan Zhu†

August 18, 2021

*I especially thank my dissertation committee members: Debarshi Nandy (chair), Stephen Cecchetti, Davide Pettenuzzo, and Yang Sun. For helpful comments and discussions, I also thank Daniel Bergstresser, Thomas Chemmanur, Jonathan Cohn, Kathleen Hanley, Mark Kamstra, Karthik Krishnan, James Ji, Blake LeBaron, Michelle Lowry, Michael McKay, Aldo Musacchio, Jing Shi, Elif Sisli-Ciamarra, Suzanne Steele and Yinchu Zhu, and the seminar participants at Brandeis International Business School, and conference participants at UT Austin 2nd PhD Student Symposium on Financial Market Research & Policy Developments, 11th Financial Markets and Corporate Governance Conference and PhD Symposium (FMCG 2021), and 2021 Shanghai Lixin Conference on New Frontiers in the Interdisciplinary Research of Finance with Global Finance Journal. I am responsible for all errors and omissions. Sample codes in Python can be found at https://www.mengnanzhu.com.

†Brandeis International Business School, 415 South Street, MS 032, Waltham, MA 02454, USA. Email: mengnanzhu@brandeis.edu

Does the Confidential IPO Registration Process Create Value?

August 18, 2021

Abstract

Since the JOBS Act of 2012, 86% of U.S. firms have confidentially filed draft registration statement (DRS) to start their IPO registration process. Using textual analysis, this study investigates the information content of DRS and its evolution to the formal prospectus (Form S-1) to show that valuable information is produced during the confidential revision process between the firm and SEC. I construct an ex-ante proxy for the content of SEC comment letters, which leads to a 6.5% cumulative abnormal return. I find changes in word content indicate changes in firm's risk environment that can further explain offer price revision, underpricing, and withdrawal decision. Moreover, changes in the proportion of positive words reflect fundamental risk of the offering and are strongly associated with operating performance. Lastly, I show that the JOBS Act substantially improves disclosure readability and IPO completion rate while lengthening the entire registration process by 40 days.

JEL: G14, G18, G24
Keywords: IPOs, JOBS Act, Draft registration statement, Textual analysis

1 Introduction

Firms go public to meet their financial needs and to broaden opportunities for future access to the capital market. For any U.S. firm that undertakes an initial public offering (IPO), a registration statement (Form S-1) must be publicly filed with the Securities and Exchange Commission (SEC) under the Securities Act of 1933.[1] Thus, S-1 is considered as one of the first steps in a typical IPO registration process, as well as the first legal disclosure regarding the private firm going public. However, this has been substantially changed since the enactment of the Jumpstart Our Business Startups Act (JOBS Act) on April 5, 2012. The JOBS Act allows an Emerging Growth Company (EGC) to voluntarily submit a draft registration statement (Form DRS) to SEC for a confidential review.[2] In 2013-2016, about 76% of the firms who went public have started their IPO process with a confidential DRS. After the expansion of the JOBS Act to non-EGC firms in 2017, more than 96% of the IPO firms confidentially filed the DRS during 2017-2020. Overall, more than 86% of IPOs have taken this new approach to go public from 2013 to 2020. Therefore, the confidential IPO registration process with DRS filing has become the new standard for going public in the post-JOBS Act era. Once DRS is filed, IPO firms revise their draft filings based on comments received from SEC several times until S-1 is publicly filed. Therefore, the evolution of DRS to S-1, or the revise-and-resubmit (R&R) process, is essentially an information production process between the firm and SEC. Using textual analysis, this study sheds light on the financial economics of the confidential revision process on the pricing strategy of the offering, underpricing phenomenon, firm's fundamental risk, the rationale of withdrawal decision, as well as the role of regulator in the public market. See Section 2 for a detailed description of the confidential revision registration process.

There are considerable theoretical and empirical studies focusing on the firm going-public decision (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

[1] Form S-1 is also known as "investment prospectus" or "preliminary prospectus."

[2] One of the main criteria to be qualified as EGC is having total annual gross revenues of less than $1 billion during its most recently completed fiscal year.

Busaba, 1997; Loughran and Ritter, 2002). Recently, a few prominent researches 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 revisions, first-day returns and post-IPO return volatility. My analysis follows this strand of the literature and builds on these studies using textual analysis. Even though 86% of IPO firms have filed a DRS since 2013, there has been no empirical research so far focusing on the information content of the DRS, or the confidential revision process defined in this study. I present the first analysis of the DRS and its evolution to the S-1 and document the relationships between changes in content during confidential revision process and firm's response to SEC comments, the offer price revision, first-day returns, operating performance, IPO withdrawal decision, as well as the overall impacts of JOBS Act on U.S. IPO market.

The objective of this study is to bridge the gap in the literature by addressing the critical questions that has received minimal attention so far: does the confidential revise-and-resubmit (R&R) process create value? Specifically, is there any valuable information revealed during the confidential revision process between the firm and SEC? Focusing on several key focuses in IPO literature, this study investigates the information content of the DRS and the changes in content during the confidential revision process on the firm's information environment using seven categories of Loughran-McDonald (LM) word lists. In addition, the study takes into account three major sections of the prospectus to consider the context of the words and to track changes over time by section. For a sample of 780 completed IPOs who went through the confidential revision process during 2013-2020, this study links the changes in content of the prospectus, measured by seven categories of LM word lists, with offer price revisions, first-day returns, pre- and post-IPO operating performance, the decision of IPO withdrawal, as well as SEC comment letters. In the robustness checks, same set of analyses are conducted using the level of the LM word lists in the first confidential version of prospectus (DRS) and the first public version of prospectus (S-1), respectively. For each group of analyses, tests are also focused on

the specific sections of the prospectus: Risk Factors, Business, Management's Discussion and Analysis (MD&A) and entire document (i.e., all sections in the main body, excluding pictures, attachments, exhibits, and XML content).

A common critique of textual analysis in finance is that the simple frequency of words used in a document may not be able to reflect the true intention of the writer. This study addresses this concern in three dimensions: 1) Seven 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, %Uncertain, %Litigious, %Constraining, %Strong-modal, and %Risky). All those word lists are developed from firm annual reports to indicate the tones of the disclosure, and thus more suitable for analyzing IPO prospectus than others.[3] 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. Using a particular section of the prospectus can potentially produce a more accurate estimation of the meaning of the words. For example, measuring uncertainty using only Risk Factors section may pick up subtle details otherwise be overlooked in the full prospectus that 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, any deviation from the benchmark (the DRS) can be better captured and tested, even though the changes are as little as a few words. This approach can largely avoid the situation where the tone of the content is so ambiguous that a mixed or extraordinary relationship is observed.

The main findings of this study can be summarized as follows. First, the new evidence found in the event study on the release of SEC comment letters show that the changes in textual analysis measures during confidential revision process is 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 %Positive in Business section, this study finds opposite direction of market reactions for the highest and lowest decile upon release of comment letters. Specifically, in 20 days after the release, the differences in cumulative abnormal returns (CAR) for the highest and lowest

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

decile are widened to 6.5%. Results in regression framework provides further evidence that changes in %Positive is strongly associated with CAR across 3-day, 5-day, 10-day, and 20-day windows.[4] The novel proxy proposed in this study enables investors to make informed investment decision on IPO, although the comment letters are not observable at the moment.

Second, the percentage changes in the proportion of %Negative in Risk Factor section during the confidential revision process are negatively associated with offer price revision. Specifically, after controlling for known IPO-specific determinants, one standard deviation increase of %Negative in Risk Factor section (+7.7%) implies a 1% downward revision in the offer price relative to the mid-point of suggested price range. Since this finding is more prominent using the Risk Factor section than using the entire document, it suggests that the additional risks revealed by SEC's scrutiny are mostly captured by the changes in the proportion of negative words in the specific section where risks involved in the offering are discussed. Using the level of %Negative in the DRS or S-1, similar relationship is found supporting the result in Loughran and McDonald (2013).[5]

Third, the overall change in word content during the confidential revision process, measured by cosine similarity, is found strongly and positively correlated with first-day returns. It shows one standard deviation increase in cosine similarity score (+4.5) leads to about 2% increase in first-day returns. Higher similarity implies no additional information is produced and thus there is a higher probability that the firm is omitting information. To hedge against this increased liability risk due to omission, the firm choose to underprice more in its offering. This finding also suggests that the R&R process should be viewed as a regulation-driven information production that associated with the value of the offering. As implied in the costly information production theory of Chemmanur and Fulghieri (1999), the confidential information release theory of Bhattacharya and Ritter (1983) and Maksimovic and Pichler (2001), as well as noted in empirical work in Hanley and Hoberg (2012), information produced by regulation-driven revisions reduces the uncertainty in firm valuation.

[4]The results are robust when using different word list or grouping strategy.

[5]See Appendix for additional results.

Fourth, the analysis with operating performance documents the channel where textual analysis measures capture the fundamental risks of the firm prior to its IPO. The findings show that the percentage changes in %Positive in Business section during the confidential revision process is a valid predictor of the real difference in post-IPO operating performance. For example, one standard deviation increase of %Positive (+8.8%) in Business section is associated with a 1.3% increase in asset-adjusted net income in 12 months since IPO. In addition, a monotonic rise in historical Sales, EPS, EBIT, and EBITDA is discovered 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 who completed the confidential revision process but eventually withdrew their IPO, this study further shows that firms with high level of %Litigious and %Strong-modal words, and low level of %Negative in their prospectus have higher probability to withdraw their IPO. Having large amount of legal words and using definitive language (e.g., always, definitely, never) along with a more optimistic tone indicate an overconfident management, or managerial hubris, in the IPO firm. Thus, it makes a subsequent withdrawal more likely. Moreover, the evidence focusing on the changes in the MD&A section during the confidential revision process imply the consistency in management's statement is associated with the decision-making process of IPO withdrawal.

Lastly, I show JOBS Act improves IPO disclosure redability and completion rate at a cost of lengthened registration process. It suggests that although the registration period after filing S-1 is significantly reduced by 66 days (or 55%) comparing to pre-JOBS Act era, the actual IPO process that starts with filing DRS is lengthened by 40 days (or 34%) on average. However, the post-JOBS Act IPO completion rate, defined as firms who went through the R&R process and successfully completed their IPO, is notably increased from 70.4% to 82.5%.

This study contributes to several streams of literature. First, it provides new empirical evidence to 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, where underpricing in equilibrium is the compensation for outsiders who engage in the costly information production. Empirically, Beatty and Ritter (1986) show the uncertainty about the firm's valuation, approximated by the uses and the amount of proceeds disclosed in 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 IPO process matters, especially among innovative firms (e.g. high-tech firms). Chemmanur, He, and Nandy (2010) examine the effect of firm's 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 R&R process. The results are robust in the settings of using content from the entire prospectus and by specific sections. In addition, the relationships are consistent no matter the draft version (DRS) or the formal version (S-1) of the prospectus is used.

This study extends and complements the literature that focuses on the information content of 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 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 revisions, first-day returns and post-IPO return volatility. In this study, I extend their approach to analyze the DRS, and build on their analysis by taking into account both the context of the words and their changes over time (i.e., the evolution of the DRS to S-1). In addition, this study also applies the textual analysis approach separately by major sections of the DRS and the S-1. Examining by section provides a more accurate estimation of the sentiment and additional insights into the source of the

sentiment. By comparing the changes of each category of word lists over time, this study sheds light on the rationale for the changes made and the value of the regulatory process.

This study also relates to the literature on the information production of the regulator. Simon (1989) finds the dispersion of abnormal returns was significantly lower following the Securities Act of 1933. Mahoney (1995) finds that external monitoring is necessary for preventing an upward-biased disclosure by issuers. Recently, Dechow, Lawrence, and Ryans (2016), and Ryans (2020) examine SEC comment letters to firm annual reports (10-K). In a recent paper, Lowry, Michaely, and Volkova (2020) also analyze the communications between SEC and the firms through comment letters using machine learning techniques. Although SEC's feedback can be valuable and direct, SEC's comment letters are not released 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 or goes public as planned).[6], and documents that the change in textual analysis measures during R&R process is an ideal ex-ante proxy for the content of comment letters prior to its release and before IPO.

Finally, this study makes contributions to textual analysis techniques by proposing an efficient method for extracting specific sections of structured financial documents in batch automatically. There are several practical issues that lead to inaccurate outcomes when extracting specific sections of firm disclosures, despite of they are structured formal documents. Pointed out in a comprehensive survey in Loughran and McDonald (2016), in addition to misspellings, abbreviations, word representation of numbers, and using Roman numerals, errors in parsing may occur due to the inconsistency of the heading of each section, the order of each section, 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]

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

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

The remainder of the paper is organized as follows. Section 2 describes IPO registration process and visualizes the the confidential revision process in detail. Section 3 describes sample construction and reports descriptive statistics. Section 4 discusses empirical results of textual analysis on several aspects of IPO, including response to SEC comments, offer price revision, first-day returns, operating performance, and withdrawal decision. Section 5 focuses on the impact of the JOBS Act, and Section 6 concludes.

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, SEC will provide feedback in the form of comment letter, which is also confidential prior to IPO.[8] To address the concerns raised by SEC, one or more revisions to DRS will be filed subsequently. This confidential review-revise-resubmit process repeats on average 2.6 rounds with no fee required by SEC and lasts about 108 days until the firm files S-1 publicly. Upon the submission of S-1, all documents previously filed, including DRS and its amendments are released on SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, except the comment letters. Since then, 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 Effective Notice is received. At last, the final IPO prospectus, Form 424 or its variants (Form 424B3/424B4), is filed at or within a few days after the IPO. Figure 1 visualizes the whole filing process for the IPO registration.

[Figure 1 here]

Even though almost every IPO firm file DRS (96% since 2017) in their registration process, the information content of DRS has not generated extensive interest among investors

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

and researchers, as the S-1 has. One might suppose that the investor's inattention is due to the nature of confidentiality of 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 DRS and its evolution to S-1 worth looking into. First, DRS is currently the first document that contains ample amounts of intangible information regarding the private firm intended to go public. Same as S-1, the IPO firm is required to disclose a detailed description of its properties, business, security being offered and risks involved, among other matters through the DRS. Since DRS is the draft version of S-1, the highly structured content has made it possible to examine the qualitative information by comparing the DRS and S-1 both as a whole document and by specific section. Second, the confidential revise-and-resubmit (R&R) process, defined as the confidential review period between DRS submission and S-1 submission, almost perfectly separates the information production of the regulator (SEC) from that of the investors.[9] As a private firm going public in a competitive industry, it tends to disclose valuable information as little as possible, conditional on passing SEC's scrutiny. Thus, the changes during the confidential revision process can be considered as regulation-driven revisions. Therefore, the changes made between DRS and S-1 should reflect the firm's reaction to SEC comments. Third, since both DRS and S-1 are available up to several months before the IPO, the confidential revision process is an ideal ex-ante proxy for the information content of SEC comment letter prior to its release.

2.2 The Confidential Revise-and-Resubmit Process

[Figure 2 here]

This study defines and focuses on the confidential revise-and-resubmit (R&R) process as the period between the submission date of DRS and the submission date of S-1 for the following two reasons. First, upon release, it provides a unique opportunity to separate

[9]Although the JOBS Act allows EGCs to "test-the-waters" by reaching out to qualified institutional buyers and institutional accredited investors, this provision is difficult to engage in practice, due to the "gun-jumping" rules of Section 5(c), which prohibits any written or oral offers prior to the filing of a registration statement (Form S-1).

the investor's information production from that of SEC. Before the public submission of S-1, all the information exchanged between the firm and SEC is confidential. Therefore, examining the interaction played 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 DRS and S-1 which allows a comprehensive comparison. As shown in Figure 2, the median number of unique financial words trends up in both DRS and S-1 after JOBS Act of 2012, while there are substantial differences between DRS and S-1 filings. For example, there are on average more than 150 unique financial words added during the confidential revision process for 2020's IPOs. The purpose of adding those words and the meaning of those additional information has not been studied intensively. Although other comparisons can be made using amendments or Form 424, none of them necessarily include enough information to conduct a systematic comparison. Amendments usually do not include full content as the filing they amended. Each amendment works individually to modify parts of the original filing. For example, one amendment can be filed just to change the company name, while another amendment updates the latest contact information. 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 S-1 is filed, such as the actual offer price and realized proceeds and fees. Since Form 424 is filed after IPO, changing information in prospectus exposes to high litigation risk due to the "quiet period" rule.[10] Hence, the new information contained in Form 424 is expected to be limited and immaterial. Naturally, one should expect a very high similarity between S-1 and Form 424. Loughran and McDonald (2013) find that S-1 and 424 flings have a very high correlation (more than 0.9) on the tone measured using uncertain, negative, and weak modal word lists. Overall, by concentrating on the confidential revise-and-resubmit (R&R) 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 compare Form 424, and Loughran and McDonald (2013), who examine S-1 and Form 424.

[10]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 who completed IPO during 2013-2020, where 780 (or 86%) of them went through the confidential revision 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 as follows. 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 have changed their names at least once during the registration process. Thus, matching DRS with S-1 simply by firm name may lead to potential loss in observations. Second, there are inconsistencies between the entity filed the documents and the entity going public. Another commonly used strategy is to match by the Central Index Key (CIK), a 10-digit unique number that identifies each entity who files with EDGAR.[11] 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 who is going public.[12]

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 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 releasing date of DRS is assured to be no later than the next business day of the S-1 submission date.[13] This paired sample is then supplemented with additional data sources for information related to the security offered, firm fundamentals, and post-IPO market information. Securities Data Company (SDC) Platinum is the source for IPO issuance data.

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

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

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

Firm fundamentals and market information are from Compustat and CRSP, respectively. Following the IPO literature, a standard screening process as described in Gao, Ritter, and Zhu (2013) has been 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. At last, there are 1560 documents (780 pairs) in the final sample, where all 780 firms went public through the confidential revision process.

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 stands for about 47% of the entire content on average in terms of total number of words. Those three sections are also the most common sections focused by investors, as the content covers the downsides and upsides of the firm and management's viewpoint, 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 in those comparisons.

In Loughran and McDonald (2016), the authors note several practical issues that lead to inaccurate outcomes when extracting specific sections. 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 full-text version. See Appendix for the detailed steps to break the entire document into sections.

3.2 Textual Analysis Variables

This study primarily uses seven categories of word lists as the measures of tone in the IPO prospectus, while 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 world.[14] This study uses the set of word lists developed in Loughran and McDonald (2011) to measure the tone in the DRS and S-1, as well as the change of the tone over time. Unlike other measures,[15] the Loughran-McDonald word lists (LM word lists) are constructed specifically using firm annual reports (i.e., Form 10-K), making it perfectly suitable for gauging sentiment in mandated financial disclosures, such as IPO prospectus in this study.[16] 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 six categories of word lists are used in the analysis and 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 Appendix for the definitions and examples of words included in each of the seven LM word lists used.

To capture the changes during the confidential revision process, two measures have been constructed. 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

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

[15]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.

[16]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.[17]

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. Intuitively, the higher similarity between the two documents, the higher cosine similarity score. The two measures constructed are complementary to each other. The percentage changes in LM word lists provides a disaggregated view at 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 variable of interests includes: Offer Price Revision, defined as the percentage change in offer price from the mid-point of the filing price range; First-day Returns, defined as the percentage change from the offer price to the closing price on the first trading day. Multiple papers in IPO literature have documented the associations between certain variables and the IPO pricing activity. To name a few, see 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 traditional IPO literature and two variables related to the confidential revision process are included as controls. Specifically, VC-backed dummy is set to one if the IPO firm is backed by venture capital; 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);[18] Positive EPS dummy set to one if the IPO firm has positive

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

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

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, measured as annual sales in most recent fiscal year prior to IPO. At last, two variables related to the confidential revisoin process are included: the length of the confidential revision process, defined as the number of days from the day DRS filed to the day S-1 filed; the number of rounds of the confidential R&R process, defined as the number of times DRS or DRS amendment is filed. See Appendix for detailed variable definitions.

3.4 Model Specifications

To study the association between variable of interests in IPO and the levels as well as the changes of textual analysis measures, the following two OLS regression models are conducted.

Levels in textual analysis measures:

$$\begin{aligned} Y_{ijt} = & \boldsymbol{\beta} LevelLM(5)_{ijt} \\ & + \boldsymbol{\theta} Controls(8)_{ijt} + YearFE_t + IndustryFE_j + \epsilon_{ijt} \end{aligned} \tag{2}$$

Changes in textual analysis measures:

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

In both models, the dependent variable, Y_{ijt}, includes several variables of interests in IPO literature, such as Offer Price Revision and First-day Returns as defined before. The seven categories of LM word lists are used as main independent variables: %Positive, %Negative, %Uncertain, %Litigious, %Strong-modal, %Constraining, and %Risky. When focusing on the levels of textual analysis measures in Eq.(2), the content from the DRS and S-1 is tested separately. To address the high correlation between %Negative and %Un-

certain 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 revision process. Main independent variables are the cosine similarity and percentage changes for each of the seven 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; and the number of days and rounds of the confidential revision process. Previous papers, such as Hanley (1993) and Lowry and Schwert (2002), have documented a statistically significant relationship between offer price revision and first-day returns. Therefore, to be consistent, the percentage of offer price revision 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

During the period of 2013-2020, there are 903 successful IPOs in the constructed sample. Among them, 780 IPOs have a valid DRS-S-1 pair. This indicates about 86% of the IPO firms (or 96% since 2017) went through the confidential revision 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 the period of 2013-2020 is 23.6%, compared to 11.7% for the period of 2001-2012 as reported on Prof. Jay Ritter's website.[19] The mean offer price revision 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,

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

VC-backed and taken public by top-tier underwriters. Specifically, 58% of the IPO firms 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 entire IPO registration process, it takes on average 160 calendar days and 4.4 amendments (including DRS/A and S-1/A) 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).[20] Intuitively, as the objective of each section in the prospectus is different, the proportion of LM word lists should vary across sections. This is confirmed in Table 1 Panel B when comparing the levels in three major sections of the DRS or S-1, respectively. For example, in both the DRS and S-1, Risk Factors section has the highest level of %Negative, %Uncertain and %Litigious, while the highest %Positive are found in Business section where the firm describes products or services, markets and competitive landscape.

[Table 2 here]

Summary statistics for the changes of 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 of the proportion of each LM word lists in one of the major sections. All seven categories of LM word lists decrease during the confidential revision process, while %Uncertain and %Litigious decrease the most. For example, the decrease in %Uncertain on average is 2.6%, or 4.4 basis points. That is about 44 uncertain words less, given a typical 100,000-word IPO prospectus. The overall decrease in LM word lists during the confidential revision process is possibly due to SEC's Plain Writing Initiative. In the guide published by SEC, it outlines the effort in reducing "the most common problems in

[20]The sample period in Loughran and McDonald (2013) is 1997-2010, which is before the enactment of 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.

disclosure documents", such as "passive voice", "legal and financial jargon", and "abstract words".[21]

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

Table 2 also reports the cosine similarity scores by major section.[22] As an additional measure of the changes during the confidential revision process, cosine similarity reflects the overall similarity based on the word content measured. Overall, the similarity for the entire document and three sections are all at a very high level. The mean cosine similarity score for entire document is 99.3. Risk Factors section has the highest similarity, while MD&A section has the lowest. This is expected since management's viewpoint are most likely to be revised per SEC's screening during the confidential revision process. The high cosine similarity score indicates the changes are small, relative to the large amount of word content measured. Note that having small changes does not necessarily mean the impact of the changes is little. Suppose a few sentences describing legal proceedings are added during the confidential revision process. The impact of those changes can be significant, even though a high cosine similarity score will be observed.

4 Discussion of Empirical Results

In this section, I examine the associations between textual analysis measures and several variables of interest that related to the pricing strategy of the offering, underpricing phenomenon, operating performance, IPO withdrawal decision, as well as SEC comment

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

[22] Cosine similarity score has been scaled by 100 for easy interpretation.

letters. For each dependent variable, tests are conducted focusing on the confidential revision process (R&R) using 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 DRS and S-1 are also tested separately. All regressions include an intercept, Fama-French 48-industry and calendar year dummies.[23]

4.1 Link Between Revision Process and SEC Comment Letter

As described in previous discussion about the confidential revision process, the evolution of DRS to S-1 is essentially an information production process by SEC. The regulator, SEC, serves an important role in balancing information transparency and regulatory burden in IPO market. In Lowry, Michaely, and Volkova (2020), the authors have shown that the comment letters issued by SEC during the registration process contains valuable information that are relevant to firm valuation. However, SEC comment letters are not publicly available until at least 20 days after IPO.[24] Although knowing the comment letters are informative, the investors, unfortunately, have no access to the specific concerns raised by the regulator before making their investment decisions on IPO.

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

In this section, a sample of comment letters for all IPOs in 2013-2020 is constructed. Specifically, all the comment letters regarding the IPO firm have been downloaded by

[23]The results for intercept and control variables are omitted some tables due to space limitation. See Appendix for full regression results.

[24]Under current regulation, SEC comment letters will not be released at least 20 days after IPO.

selecting "UPLOAD" as form type on the EDGAR. Since only the comment letters that related to the confidential revision process are interested, the comment letters interested are required to satisfy the following to screening rules: First, the comment letter must be issued to the IPO firm with the same CIK. Second, the issue date of comment letter (not be confused with the release date) must between the filing date of the DRS and the filing date of the S-1. Third, the comment letter must be publicly available on EDGAR.[25] The final sample includes all IPO firms described in Section 3. In total, there are 2668 comment letters issued to 770 IPOs firms during the R&R process in 2013-2020.[26]

Since the objective of this analysis is to investigate whether changes in textual analysis measures lead to different market reactions when the market sees the comment letters, an event study on the comment letter release is performed. Figure 3 plots the cumulative abnormal returns (CAR) over the window of 10 trading days before and 20 trading days after the release day of comment letters. CAR is defined as the cumulative difference between firm stock return and corresponding size and book-to-market 25 industry portfolio returns over the $[-10, +20]$ window relative to the release day of comment letters. The black line represents all IPOs in this sample and serves as the benchmark. Ryans (2020) and Lowry, Michaely, and Volkova (2020) have shown that revenue recognition or topics related to operating performance are the dominant concerns in the SEC comment letters. Therefore, changes in %Positive in Business section is focused on in this part of analysis, as the results in later section have shown the strong association between positive words and operating performance. Based on deciles of the percentage changes in %Positive in Business section during the revision process, two groups of firms are also plotted in Figure 3. The blue line represents the lowest decile whose proportion of positive words have been decreased by 11.6% during the revision process, while the red line represents the highest decile whose proportion of positive words have been increased by 8.2%.

As Figure 3 shown, there is a clear difference in CAR since the release of the comment letter for the two groups. Firms in the highest decile where more positive words added

[25] Although it is possible to obtain all comment letters via a Freedom of Information Act (FOIA) request, it is unlikely a regular investor will do so. Therefore, this study keeps the information set at the most realistic level as investors.

[26] There are 10 firms that fail the screening rules described above at the time of the data is collected.

during revision process experience an upward trend in CAR, while firms in the lowest decile underperform relative to the highest decile group as well as the benchmark. This observation supports the argument that the ex-ante measure (i.e., changes in %Positive in Business section during confidential revision process) is able to separate firms in terms of the concerns raised by SEC. Prior to the release of comment letters, firms in the highest decile slightly underperform comparing to firms in the lowest decile. However, a market adjustment has occurred within two days after the comment letter release.[27] Then, the difference between high performance firms and low performance firms has started to widen. By the 20 trading days after the comment letter release, there is a 6.5% difference in CAR for those two groups.

[Table 3 here]

To further explore the return predictability, Table 3 presents the analysis of the link between the proxy constructed during the revision process and the content of comment letter in a regression framework. The dependent variable is the same CAR in Figure 3 for various windows from (0, +3) to (0, +20). The results in Panel A show that percentage changes in %Positive in Business section are strongly correlated with CAR in 5 days, 10 days, and 20 days after comment letter release. Economically speaking, one standard deviation increase of %Positive is associated with 2.7% cumulative abnormal return in 20 days since release. Panel B reports a similar analysis by separating changes in %Positive in Business section into three different bins (i.e., top decile, middle deciles, and bottom decile). The findings support the observations made in Figure 3 that the more positive words added during the revision process, the better predictive power in terms of firm performance. Specifically, the coefficient of %Positive in top decile is strongly associated with CAR across 3-day, 5-day, 10-day, and 20-day windows. In summary, findings in this event study provide new and strong evidence that the changes in textual analysis measures during the confidential revision process is an ideal ex-ante proxy for the valuable information contained in SEC comment letters that are not publicly available prior to IPO. This means

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

although the comment letters are not observable before the IPO, investors still can make informed investment decision using the novel proxy proposed in this study.

4.2 Results with Offer Price Revision

Table 4 presents the regression results using the percentage changes in the proportion of LM word lists during the confidential revise-and-resubmit process which is defined as the period between DRS submission and S-1 submission. Results in each column show various sections of prospectus used in textual analysis (i.e., Entire Document, Risk Factors, Business, and MD&A). Offer Price Revision is the percentage change in the offer price from the mid-point of the filing range.

In untabulated results of Table 4, some known determinants that are used as control variables are found significantly correlated with IPO pricing activities as documented in previous literature. For example, IPOs with top-tier lead underwriters and less offered shares have higher offer price revision. VC-backed IPOs have a strong positive relationship with offer price revision. Since the above-mentioned 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 Appendix for full regression results.

[Table 4 here]

When focusing on the changes in the entire document, none of the LM word lists have significant coefficient values (column 1). It means after controlling for known determinants the changes in the proportions of LM word lists in the entire document cannot explain the changes made in the offer price. This result is not surprising given the large amount of word content in the entire document. Suppose a few words has been changed that contains critical information. Those changes might not be picked up by the textual analysis measures. In this case, as this study proposed, focusing on major section may overcome this issue and capture those subtle changes. In column 2, the results for Risk Factors section shows a negative and significant correlation between the %Negative and offer price

revision. One standard deviation increases of %Negative (+7.7%) implies 1% downward revision in offer price relative to the mid-point of suggested price range. As noted before, the revision process should be considered as an information production by SEC. Thus, adding more negative words after SEC's review is considered as an unfavorable signal to the firm. It suggests either some risks are omitted, or the severity of the identified risks is underestimated in the original draft filing (DRS). For instance, three negative words ("fail", "harm", "loss") used in Risk Factors section of We Co. (also known as WeWork) are significantly increased during the revision process.[28] In the Risk Factors section of We Co.'s filing, "fail" is used 60% more in S-1 than in DRS, while "harm" is used 22 times in S-1 comparing to 14 times in DRS. Moreover, for all the addition of "fail" and "harm" after SEC's review, 90% of "fail" added and 100% of " harm" added are located in the Risk Factors section. This supports the finding that the changes of %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 revision process result in a downward adjustment on the valuation (reflected by reducing offer price), and eventually contributes to its withdrawal.

Although the changes in %Negative is found negatively correlated, previous literature has shown that the levels of %Negative are positively associated with offer price revision. In column 1, the coefficient value on the level of %Negative is not statistically significant, while the positive sign indicates a consistent result with Loughran and McDonald (2013). Given the nature of Risk Factors section and the findings about the %Negative during the revision process, this study provides additional evidence to interpret negative words in IPO filing. According to SEC's guidance, 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. 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 using more negative words. The hypothesis is that the more negative words used by the management in Risk Factor section, the more known risks are identified, and the more prepared the firm's risk management strategies are, and

[28] See Appendix for details.

hence 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". Example of "Known-knowns" can be seasonal fluctuations in demand or in raw material cost, while "Known-unknowns" usually refers 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 will be purely speculative information that are prohibited. Similarly, "Unknown-knowns" refers to hidden facts which can be simply addressed by hiring external specialists or professionals, and hence not likely to be focused here. Normally, once a risk is identified in the prospectus, a corresponding strategy to mitigate its impact is proposed and explained. In addition, sufficient disclosure of underlying risks involved may also reduce the litigation risk the firm faced after going public.[29]

Overall, a high level of %Negative in 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 revision process suggests underestimated risk exposures. When using Risk Factors section, the observed negative relationship of the changes in %Negative with the offer price revision supports this hypothesis.

The second 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 revision. In column 1 using entire document, the results report that the coefficient on %Constraining is –15.5 (with $t-$statistic of –2.91). This implies, for example, 1% decrease in the proportion of constraining words in S-1 is associated with 16% upward price revision. Similar interpretation as in Bodnaruk, Loughran, and McDonald (2015), having a low level of constraining words in the prospectus reflects fewer constraints the firm is facing, in turn, indicates a higher degree of financial flexibility. Therefore, the offer price will be revised upward during the price discovery process reflecting a healthier firm's financial condition. The results also show that high level of strong modal words is an indicator of upward revision in offer price. This implies that a confident tone may suggest a strong believe in the firm's value,

[29]Consistent with the findings in Hanley and Hoberg (2012).

therefore, a strong tendency of increasing offer price is observed. Interestingly, too many strong modal words may indicate over-confidence in management. The analysis with IPO withdrawal decision supports this interpretation, which will be discussed in detail in later sections.

4.3 Results with First-day Returns

This section presents the analysis of underpricing phenomenon. IPO literature has documented a strong and positive relationship between offer price revision and first-day returns, such as Hanley (1993) and Lowry and Schwert (2002). To clearly identify the power of textual analysis measures, the previous dependent variable (i.e., offer price revision) is added as a control when testing underpricing. Table 5 reports the regression results using the percentage changes in the proportion of LM word lists during the confidential revise-and-resubmit process. First-day Returns is the percentage change from the offer price to the closing price on the first trading day. Similar to Table 4, results in each column show various sections of IPO prospectus used (i.e., Entire Document, Risk Factors, Business, and MD&A).

[Table 5 here]

Consistent with previous studies, the results show that offer price revision is positively correlated with first-day returns at 1% confidence level, no matter entire document is used or any of the three major sections is used. In column 1, a strong positive relationship is found between cosine similarity and first-day returns. It suggests one standard deviation increase in cosine similarity score (+4.5) leads to 1.9% increase in first-day return. 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 is produced during the revision process, and thus there is a high probability that the firm is omitting information. As a result of this possible omission, the firm is facing higher post-IPO litigation risk. To hedge against this increased liability risk, the firm may choose to underprice more in

offer price.[30] Results using major sections of the prospectus also find similar results as previously shown. In Business section, changes in %Negative has a significant coefficient of –0.1. This translates to a one standard deviation increase of %Negative in Business section (+12.5%) associating with 1.2% decreases in first-day returns. When more negative words are added after SEC's scrutiny, probably indicates the firm is holding back some unfavorable information in its previous draft (DRS). Therefore, this additional negative information revealed during the revison process, will lead to a downward adjustment on the firm valuation.

When focusing on the levels of LM word lists, the results in column 3 show that %Negative is positively correlated with first-day returns and statistically significant. Strong negative tone implies a more pessimistic view regarding future cash flows, and thus riskier for investors to hold shares of the IPO firm, which leads to a higher degree of first-day return through heavy underpricing. For example, 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 return in the sample. This finding supports several key theories of IPO pricing literature in terms of sign, magnitude, and confidence level. Rock (1986) describes the "Winner's Curse" in IPO and argues underpricing is necessary for attracting uninformed investors in an asymmetric information-based setting. Beatty and Ritter (1986) provide empirical evidence showing 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 is positively associated with first-day returns, suggesting a strong and persistent relationship across past 20 years of IPO market.[31] In addition, a negative relationship between %Constraining in entire document and first-day returns is found. It implies the demand for those who has less financial flexibility is relatively low. On the other hand, firms with less financial constraints

[30]Section 11 of the Securities Act of 1933 require investors to sue issuers and underwriters with evidence of material omissions in the prospectus AND losses below offer price.

[31]Loughran and McDonald (2013) use IPO samples in 1997-2010, while the sample period in this study is 2013-2020.

attract higher demand, which pushes up the stock price in the open market yielding a high first-day return.

4.4 Results with Operating Performance

To investigate the value of the confidential revision process, this paper also explores the extent to which changes in textual analysis measures associate with the operating performance of the IPO firm. Unlike Risk Factors where lots of concerns have been discussed, Business section describes the firm's lines of business, its principal products or services and its markets. In addition, the competitive landscape, including advantages and disadvantages to its competitors, significant suppliers and customers, among other matters can be found in this section. Regarding the nature of the information related to firm's operating performance in Business section, it is usually forward-looking with long horizon. Therefore, the hypothesis is that an overall optimistic tone in Business section should be expected for "good quality" firms. Consistent with the 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 Business section is more than two times as in 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, a univariate test is shown in Table 6.

[Table 6 here]

In Table 6 Panel 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 performance are reported: Sales, EPS, EBIT, and EBITDA with the mean and median values based on the most recent fiscal year prior to IPO. In either DRS or S-1, a monotonic rise across quintiles is observed for all four operating performance measures, except the lowest quintile.[32] There is a clear trend that implies the higher %Positive is correlated with higher pre-IPO operating performance. For example, in Panel B although the mean of EPS reported in S-1 is below zero for all quintiles, it increases from $–0.84 in the lowest

[32]One possible reason for the distinct observation in the quintile with lowest level of %Positive is firm's earning management behavior.

to $–0.11 in the highest quintile. The difference of 73 cents in EPS is considerably large in the financial analysis, given 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 found among operating performance measures in both the mean and median. Overall, 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. Intuitively, higher demand for those growth firms will lead to an upward revision in offer price, which is confirmed in previous analysis with offer price (see Table 3 column 3).

Table 7 further examine the predictability of %Positive in a regression framework as additional supports for above-memtioned findings. Using the percentage changes in the proportion of LM word lists in Business section, each column reports the regression results for common measures of post-IPO operating performance for various time periods after 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 clustered at industry level.

[Table 7 here]

Consistent with the hypothesis that %Positive in Business section is a valid predictor of real difference in operating performance for the IPO firm, coefficient values of percentage changes in %Positive during the revision process are statistically significant and economically meaningful. For example, an one standard deviation increase of %Positive is associated with 0.7% increase and 1.3% increase in asset-adjusted net income in 6 months and 12 months since IPO, respectively. These findings highlight the fact that the changes in proportion of positive words after SEC's reviewing process is a valid proxy for fundamental risks of IPO firms. More importantly, these findings with operating performance document the channel where textual analysis measures capture the unobservable firm risks at the time of the IPO.

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 firm's withdrawal decision focusing on DRS and the revise-and-revision (R&R) process. Unlike SEC comment letter, DRS is released at the time S-1 is publicly filed, making the revison process a perfect ex-ante measure of firm risk. In addition, DRS and its amendments are available regardless of the final decision of the IPO process (i.e., withdraw or complete). Therefore, DRS along with the revision process provide an unique opportunity to study the firm's withdrawal decision because the counterfactual can be identified as those who filed DRS initially and went through the revision process, but eventually decided to withdraw, such as the case of We Co. In this section, this paper examines the link between textual analysis measures and the withdrawal decision.

The sample of withdrawn IPOs are collected by downloading all Form RW (i.e., registration withdrawn) from EDGAR during 2013-2020.[33] In order to avoid selection biases, withdrawn IPOs are required having gone through the confidential revisoin process. Additionally, firms who withdrew their IPO but successfully went public at a later date is dropped in this analysis. These requirements provide an unbiased comparison to those completed IPOs. The final sample of this analysis contains 872 IPOs of which 719 were completed while 153 were withdrawn during 2013-2020. The summary statistics for this sample is reported in Table 8.

[Table 8 here]

First, Panel A shows that withdrawn IPOs take more time to response to each round of revision. Although both completed and withdrawn IPOs go through the same rounds of revision process (about 2.5 rounds), it takes on average 9 days more for withdrawn IPOs to response to SEC's comment in each round. Intuitively, if the feedback from 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

[33]There are firms who filed DRS but never files S-1 or RW. Since the real intention of those firms are unclear, they are not included in the sample.

is possible that the information production process contributes directly to the substantial difference in final decision of the IPO. Second, the comparison shows withdrawn IPOs spend much more time before making a final decision on the offering and submit less S-1 amendments during the public road-show period. One thing to note here is that not all firms filed Form RW immediately after the withdrawal decision are made inside the company. Some firms wait for up to nine months (or pick a more favorable date) before publicly given up on IPO. Thus, the total number of days spend on IPO registration process may be an inflated measure for the analysis in withdrawal decision. Therefore, this study focuses on focusing on the confidential revision process that is prior to S-1 submission to avoid this bias.

Panel B presents summary statistics for LM word list levels in entire S-1 filing. There are clear distinctions between completed and withdrawn IPOs in LM word list levels. First, withdrawn IPOs contains lower level of %Positive and higher level of %Litigious in their prospectus. As shown in previous analysis, this implies the withdrawn IPOs are less profitable in terms of historical operating performance and expose to high litigation risk. Second, the findings show that withdrawn IPOs includes more strong modal words, and less negative and uncertain words in their disclosures. For example, on average, there are 16 basis points less %Risky for withdrawn IPOs. Low level of %Negative and high level of %Strong-modal suggest an over-confidence in management of withdrawn firms. As documented in 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 over-optimistic tone in the prospectus. Panel C presents summary statistics for the percentage changes in the proportions of LM word lists during the confidential revision process. Consistent with findings in previous sections of paper, SEC's scrutiny has forced the withdrawn firm to disclose the negative information holding back in their initial draft filings. During the revision 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 R&R process, measured by total number of LM words. This implies that for IPO firms who submitted a

less informative filing, SEC has successfully discovered more valuable information in their review, especially for withdrawn IPOs. To further explore the decision-making process of withdrawn IPOs, a series of logit regression are tested focusing on DRS, S-1 and the confidential revision process.

[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 where 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 summary statistics, longer days spent and less rounds of reviews in the revision process indicate a higher probability of withdrawal. More time spent on the revision process means more issues need to be addressed before qualifying for IPO. Second, High level of %Litigious leads to high chance of withdrawal. Intuitively, having more legal words reflects higher litigation risk, thus making a subsequent withdrawal more likely. Third, high level of %Strong-modal is strongly contributes to the withdrawal decision. The coefficient value on %Strong-modal is +4.1 and significant at 1% confidence level. As an ex-ante measure of managerial hubris, the interpretation for the effect of strong modal words is that using large amount of overconfident language (e.g., always, definitely, never) may be considered as over-selling the firm potential from the investor's point of view. This result also supports the findings in Loughran and McDonald (2013). Fourth, high level of %Negative words reduces the likelihood to withdraw an IPO. More negative words used in the prospectus may followed by a downward revision in offer price therefore make selling shares offered much easier. This is in line with the idea of "Winner's Curse" described in Rock (1986) and found in the previous analysis with underpricing where %Negative is positively correlated with underpricing. Moreover, more negative information can be seen as a more practical and conservative view of firm management, and thus reducing the post-IPO litigation risk.

So far, the analysis has shown high litigation risk and managerial hubris contributes to the withdrawal decision. Since the decision is made by the firm insiders (i.e., the management), it is interesting to see whether the word content in the specific 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 entire document as well as MD&A section only. One of the reasons MD&A section attracts so much attention from investors and regulators is that MD&A section not only contains management's perspectives regarding firm's financial condition and results of operations, but also discussed in narrative. This narrative section should be the easiest way for outsiders to understand the risks incorporated with the shares issued.[34] However, prior literature focusing on 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. Recently, using MD&A disclosures in 10-K, Hoberg and Lewis (2017) find 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 MD&A section compared to entire document. High similarity in MD&A section indicates consistent statements before and after SEC's review process. An intuitive explanation is that if the story management told is inconsistent during SEC's review, it probably should be considered as a flaw in the statement or changing in the risk environment over the revision process. Either way, from 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 word list level of 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

[34]In 2003, SEC has issued interpretive guidance for MD&A which states the principal objectives of MD&A section as "To provide a narrative explanation of a company's financial statements that enables investors to see the company through the eyes of management... To 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."

department of the firm. The assumption is that all the statements provided in MD&A are well prepared to avoid subsequent persecutions. This means what to say and how to say it in MD&A section has been strategically designed to reduce the post-IPO litigation risk. Following this assumption, one should expect no significant results for text-based analysis in MD&A section. For example, Loughran and McDonald (2011) use firm 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 to Risk Factors, Business section and Entire Document. Among the seven LM word lists examined, six of them have the lowest level for MD&A section, with the exception of %Uncertain. Specifically, the level of %Litigious in MD&A section of S-1 is 0.51%, which is 46% less than the level in Business section, or 70% less than the level in Risk Factors section. It seems firm management avoid mentioning legal matters in this section. Indeed, Hoberg and Lewis (2017) find personal names are mentioned less frequently in MD&A section to avoid subsequent prosecutions. Chen et al. (2020) document that firms may alter their wordings in ways 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, in turn, reduces the probability to be used as evidence against the managers in subsequent prosecutions. Therefore, it is reasonable to assume that statements in MD&A section have been catered to reduce litigation risk against the management. In summary, although the levels of LM word lists in MD&A section fail to explain withdrawal decision, the changes during the confidential revision process shows the consistency in management's statement, measured by cosine similarity, is associated with the decision-making process of withdrawal.

5 Impact of JOBS Act on IPO Market

In the last part of the study, I look into the impact of the JOBS Act on IPO market from two aspects: the efficiency of capital formation and transparency of corporate disclosures.

[Table 11 here]

Table 11 Panel A and B show the impacts of the JOBS Act on the U.S. IPO market in terms of the total time spend on 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.[35] Post-JOBS Act period includes 872 IPOs, of which 719 were completed while 153 were withdrawn during 2013-2020.

Prior to JOBS Act, it takes on average 117 calendar days to complete an IPO, while the number of days spent increases to 157 in the post-JOBS Act period. The increased time on the IPO process is primarily due to the confidential revise-and-resubmit (R&R) process. When excluding the period of confidential revision process, the results suggest that the registration period after filing 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 DRS confidentially, is lengthened by 40 days (or 34%) on average. Considering the expensive consulting and legal fee associated with on average 1.6 amendments filed on DRS, the firm's cost is expected to increase significantly. However, it seems the additional time and money spent on the confidential revision process is not justified by a higher IPO proceeds for the firm, as the mean of post-JOBS Act first-day returns is 24%, comparing to 11% during 2001-2011. On the other hand, the post-JOBS Act IPO completion rate, defined as firms who filed DRS and S-1 and successfully went public, is notably increased from 70.4% to 82.5%. It seems the lengthened registration time contributes to the improvement in the completion rate. Future research should consider looking into the market timing behavior and the overall quality of IPO firms in the post-JOBS Act era.

[Figure 4 here]

[Figure 5 here]

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

Regarding the transparency of IPO disclosures, this study majorly focuses on the readability of the prospectus. Word Complexity is commonly used in textual analysis literature as one of the measures for the readability of the financial disclosures. It is defined as the average syllables per word as in Loughran and McDonald (2011). Figure 4 plots the median level of word complexity in DRS and S-1 for the sample used in this study, respectively. The comparison between DRS and S-1 shows that the word complexity of IPO disclosure is consistently reduced after SEC's review since the enactment of JOBS Act and across major industries. According to SEC's Plain Writing Initiative, one of the objectives in the review process is to improve the readability of company disclosures. Clearly, since JOBS Act of 2012, there is a notable decrease in word complexity after the confidential review. Although the word complexity increases over the past 8 years, the complexity level has been consistently reduced. Figure 5 plots the median of percentage changes in word complexity during the confidential revision process by Fame-French 12-industry classification. Across all industries, it shows a consistent reduction in word complexity, with the exception of in utility and chemistry industry. Intuitively, the exceptions observed are probably due to the characteristics of those two industries, as more complex words are needed to clearly explain the complicated business operation evolved in the utility and chemistry industry. Overall, it shows the confidential revision process with SEC improves readability of the corporate disclosure, which may have indirect impact on reducing the uncertainty and asymmetric information issue 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.[36]

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

6 Conclusion

One of the most critical stages in a firm's life cycle is going public. Since the enactment of 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 (R&R) process in the post-JOBS Act era and strives to answer the vital question: is there any valuable information produced during the confidential revision process between the firm and SEC? Specifically, using seven categories of Loughran-McDonald (LM) word lists, this study documents the link between the changes in content during the evolution of DRS to S-1 and several critical matters related to IPO, such as offer price revision, first-day returns, post-IPO performance, IPO withdrawal decision, as well as the link to SEC comment letter and the overall impacts of JOBS Act on U.S. IPO market. In addition to using entire content of prospectus, 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 revision process on the pricing strategy of the offering, underpricing phenomenon, firm's information environment, rationale of withdrawal decision, market efficiency, as well as the role of regulator. First, the event study on the release of SEC comment letters provides new evidence that the change in textual analysis measures during confidential revision process is an ideal ex-ante proxy for the content of comment letters prior to IPO. Second, the evidence suggests that the additional risks revealed by SEC's review process are mostly 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 revision process, measured by cosine similarity, is found strongly and positively correlated with first-day returns. Fourth, the analysis with operating performance documents that it is the real effects that drives the return predictability of textual analysis measures. Fifth, the study shows high level of litigation risk and managerial hubris makes a subsequent withdrawal more likely. Lastly, JOBS Act substantially improves disclosure readability and IPO completion rate at a cost of lengthened registration process.

A natural extension of this study is to utilize the DRS filings and the confidential revision process further. Since DRS and its amendments are available regardless of the final decision of the IPO, it provides an opportunity to study the firm's choice of raising capital because the counterfactual can be identified as those who filed DRS initially but later decided to pursue other alternatives (e.g., merge and acquisition, or SPAC-IPO). Moreover, examining the impact of the confidential revision process on the IPO market has policy implications regarding the role of regulator on balancing information transparency and regulatory burden. For example, the cost-benefit of regulator's attempts on increasing 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.

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.

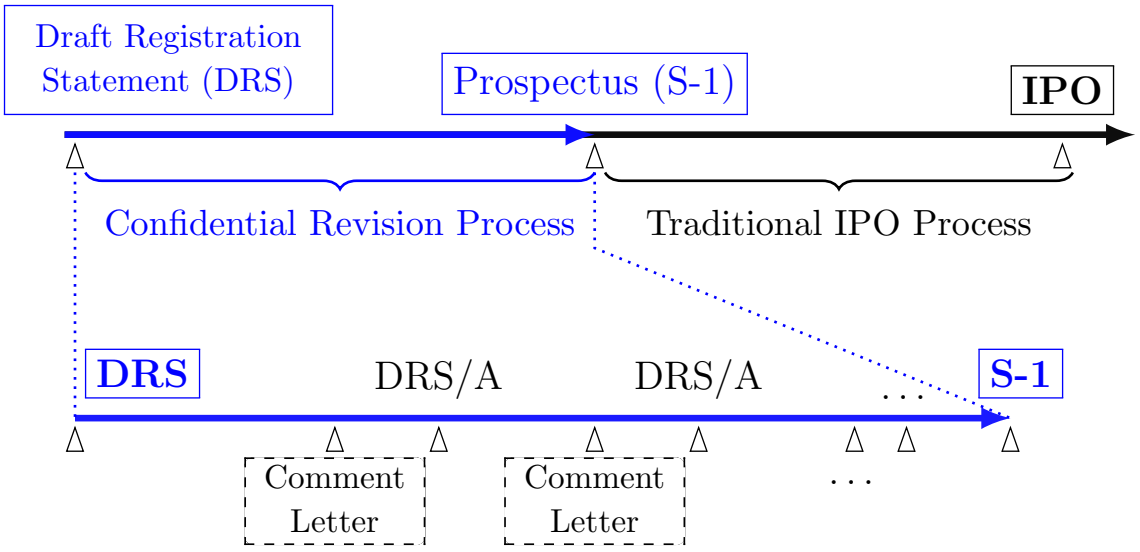

Figure 1: The IPO Registration Process in Post-JOBS Act Era

The entire timeline on the top describes a typical IPO registration process in post-JOBS Act era. Before the JOBS Act, the traditional IPO process starts with filing prospectus (S-1) publicly (the black part of the timeline). The JOBS Act of 2012 allows a confidential revise-and-resubmit process that is defined as the period between Draft Registration Statement (DRS) submission and S-1 submission (the blue part of the timeline). The bottom timeline describes detailed revision process during the confidential process. Form DRS/A is amendment to DRS. Comment Letter is the feedback provided by SEC regarding concerns in prospectus, which is not observable before IPO.

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

- 2018/09/11: Filed DRS confidentially
- 2018/10: First DRS amendment filed
- 2018/11: Second DRS amendment filed
- 2018/11/16: Filed S-1 publicly (DRS and its amendments are released)
- 2019/05/02: IPO at $25 on NASDAQ (closing at $65.75)

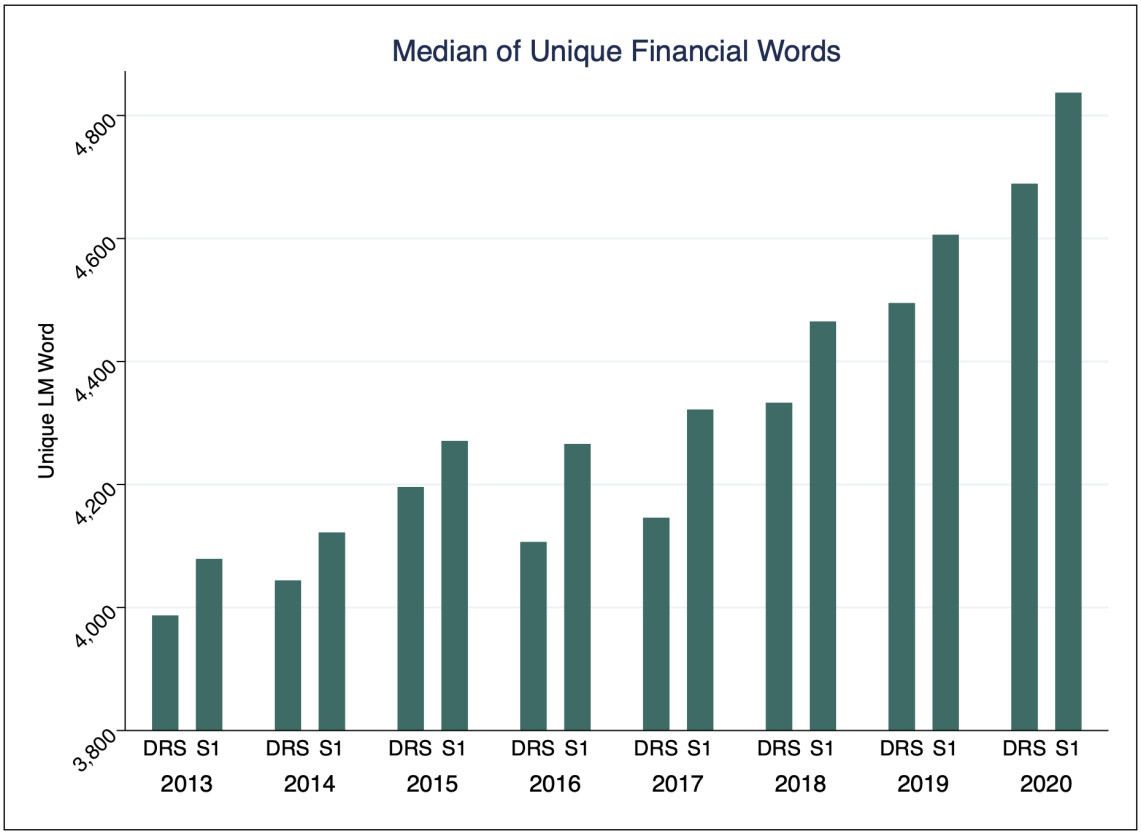

Figure 2: Unique Financial Words in DRS and S-1 Filings

This figure plots the median number of unique financial words in the Draft Registration Statement (DRS) and the prospectus (S-1) since the JOBS Act of 2012, respectively. The sample includes 780 IPOs who 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).

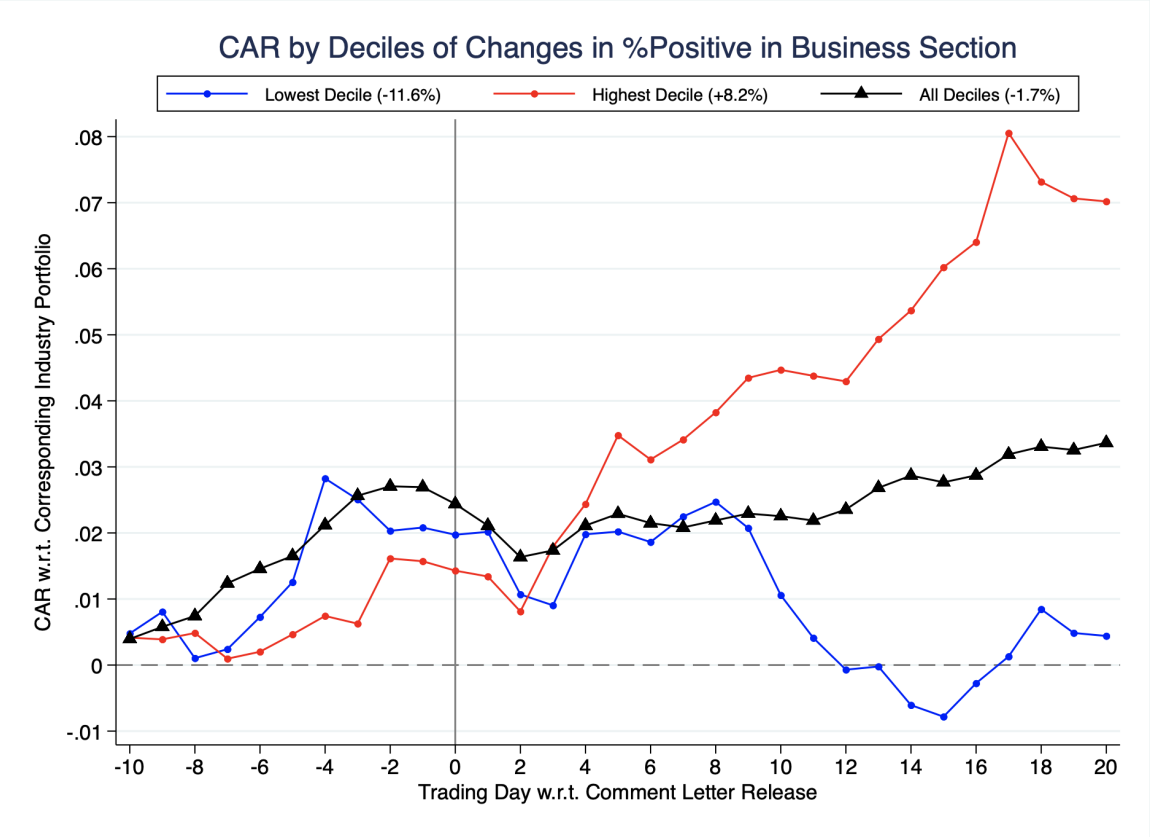

Figure 3: CAR by Decile of Changes in Positive Words During R&R

This figure plots the cumulative abnormal returns (CAR) for two portfolios of IPOs, as well the benchmark. The portfolio is constructed based on decile of the percentage changes in %Positive in Business section during the confidential revise-and-resubmit process (from DRS to S-1). The blue line represents the lowest decile whose proportion of positive words have been decreased by 11.6% during the confidential revision process, while the red line represents the highest decile whose proportion of positive words have been increased by 8.2% on average. The black line represents all IPOs. CAR is defined as the cumulative difference between firm stock return and corresponding size and book-to-market 25 industry portfolio returns over the $[-10, +20]$ days window relative to the release day of comment letters.

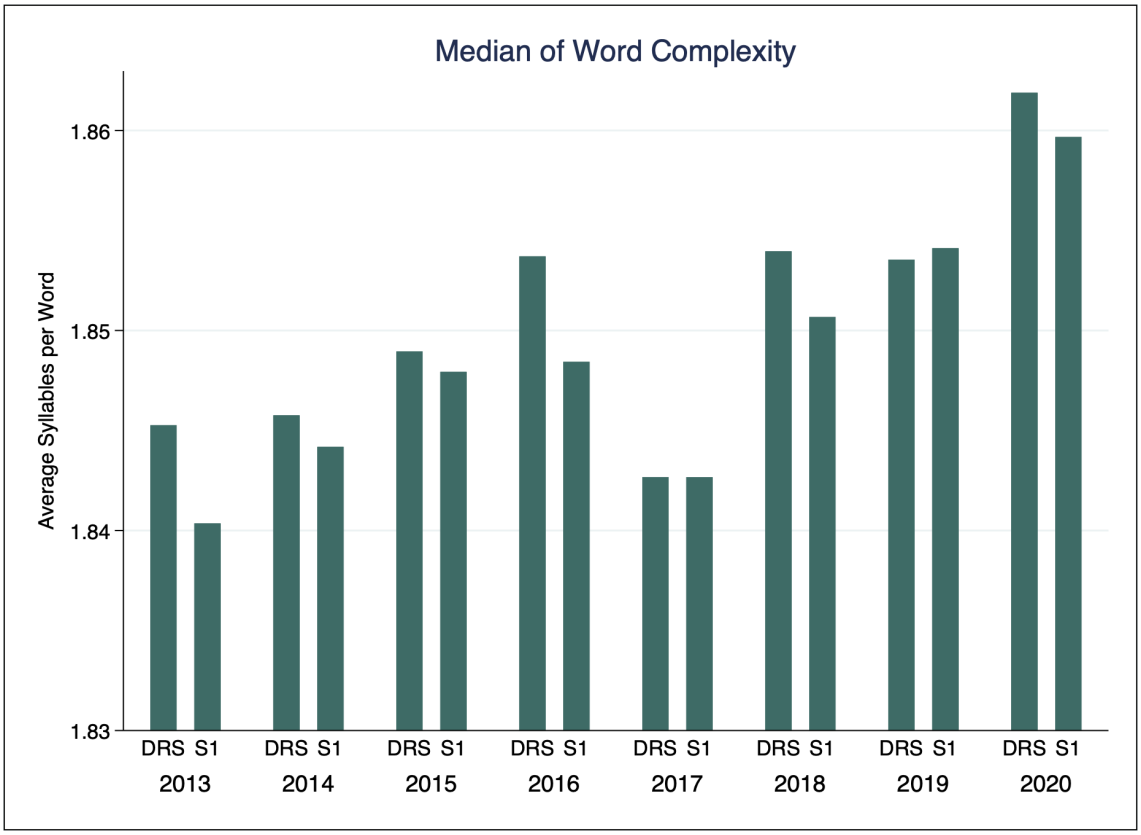

Figure 4: Word Complexity in DRS and S-1 Filings

This figure plots the median level of word complexity in the Draft Registration Statement (DRS) and the prospectus (Form S-1) for 780 IPOs who went through the confidential revise-and-resubmit process (defined as from DRS to S-1), respectively. Word Complexity is defined as the average syllables per word as in Loughran and McDonald (2011).

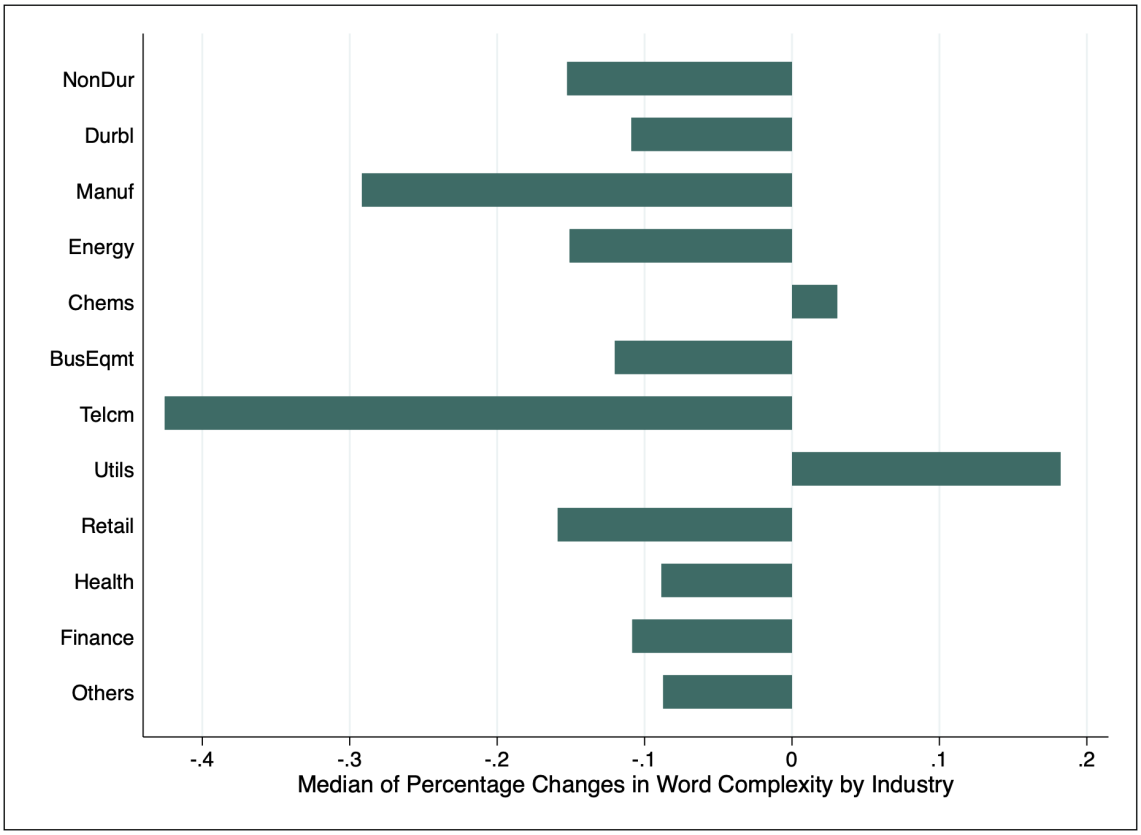

Figure 5: Changes in Word Complexity by Industry

This figure plots the median of percentage changes in word complexity during the confidential revise-and-resubmit process by industry. The sample includes 780 IPOs who went through the confidential revise-and-resubmit process (defined as from DRS to S-1). Industry is defined using Fama and 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 who 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 Revision 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 Revision (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 of Public Review (in days)	780	51.95	70.32	21	26	31	44	155
IPO Process Period (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

Table 1 (Cont.): Summary Statistics for IPO Sample, 2013-2020

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

	Entire Document			Risk Factors			Business			MD&A		
	DRS (%)	S-1 (%)	Diff. (bps)	DRS (%)	S-1 (%)	Diff. (bps)	DRS (%)	S-1 (%)	Diff. (bps)	DRS (%)	S-1 (%)	Diff. (bps)
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
Total Number of LM Words	94920	104197	9277	22445	23061	616	13762	14912	1150	9081	10542	1460
%Positive LM Words	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 LM Words	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 LM Words	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 LM Words	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 LM Words	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 LM Words	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 LM Words	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 textual analysis measures 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 (2011) 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 Lists 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 Lists 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 Lists 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 Lists 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: Event Study on Comment Letter Release

Table 3 Panel A presents the regression results using the percentage changes in the proportion of LM word lists 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 decile of the percentage changes in %Positive in 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 25 industry portfolio returns. Due to space limitation, results for control variables are omitted. The t-statistics are in parentheses with the standard errors clustered by Fama-French 48-industry and calendar year.

Panel A: Regression with CAR on Comment Letter Release

	CAR(0,+5)		CAR(0,+10)		CAR(0,+20)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cosine Similarity	0.0355	-0.0228	-0.0724	-0.0915	0.187*	0.105
	(0.39)	(-0.42)	(-0.73)	(-1.27)	(1.90)	(0.95)
PctChng: %Positive LM Words	0.130***	0.111*	0.156*	0.126	0.355***	0.315**
	(4.02)	(2.28)	(2.14)	(1.45)	(3.62)	(2.57)
PctChng: %Litigious LM Words	0.0130	0.00630	0.00636	-0.00417	-0.00319	-0.0140
	(1.58)	(0.70)	(0.82)	(-0.24)	(-0.38)	(-0.85)
PctChng: %Strong Modal LM Words	-0.0184	-0.0134	-0.0226	-0.0122	-0.0302	-0.0265
	(-1.15)	(-0.47)	(-0.83)	(-0.41)	(-0.71)	(-0.50)
PctChng: %Constraining LM Words	-0.04***	-0.0324	-0.0154	-0.00396	0.0565***	0.0685**
	(-3.74)	(-1.63)	(-0.96)	(-0.11)	(7.13)	(2.49)
PctChng: %Negative LM Words	0.0220	0.00170	0.0986**	0.0750	0.129	0.0905
	(0.58)	(0.04)	(2.55)	(1.21)	(1.28)	(0.85)
S1: %Positive LM Words		0.574		1.892		2.037
		(0.42)		(0.87)		(0.99)
S1: %Litigious LM Words		2.434		2.263		4.141
		(0.88)		(0.80)		(1.16)
S1: %Strong Modal LM Words		3.457		-0.246		-4.027
		(0.77)		(-0.04)		(-0.40)
S1: %Constraining LM Words		-8.109		-4.440		-9.112
		(-1.73)		(-0.78)		(-1.73)
S1: %Negative LM Words		-0.191		0.700		-1.519
		(-0.14)		(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	597	597	594	594	594	594
R^2	0.048	0.086	0.038	0.070	0.068	0.098

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

Table 3 (Cont.): Event Study on Comment Letter Release

Panel B: Subgroup Regression by Decile of Changes in %Positive in Business Section

	(1) CAR(0,+3)	(2) CAR(0,+5)	(3) CAR(0,+10)	(4) CAR(0,+20)
Cosine Similarity	0.0767	0.0230	-0.0190	0.104
	(0.74)	(0.29)	(-0.18)	(0.73)
PctChng:%Positive × Top Decile	0.159***	0.181*	0.241**	0.314**
	(4.41)	(2.21)	(3.49)	(2.57)
PctChng:%Positive × Bottom Decile	-0.0228	-0.0698	0.0002	0.326
	(-0.17)	(-0.30)	(0.00)	(1.56)
PctChng:%Positive × Middle Deciles	-0.0166	0.234	-0.126	0.294
	(-0.09)	(0.84)	(-0.30)	(0.52)
PctChng: %Litigious LM Words	0.00447	0.00601	-0.00216	-0.0139
	(0.84)	(0.54)	(-0.16)	(-0.93)
PctChng: %Strong Modal LM Words	-0.0307*	-0.0232	-0.0221	-0.0260
	(-2.19)	(-0.84)	(-0.91)	(-0.57)
PctChng: %Constraining LM Words	-0.00899	-0.0386	-0.0113	0.0688**
	(-0.41)	(-1.11)	(-0.27)	(2.71)
PctChng: %Negative LM Words	-0.00646	-0.00665	0.0724	0.0912
	(-0.19)	(-0.16)	(1.53)	(0.73)
S1: %Positive LM Words	1.390	0.827	2.220	2.031
	(1.37)	(0.52)	(0.98)	(0.91)
S1: %Litigious LM Words	1.329	2.432	2.208	4.139
	(0.67)	(0.90)	(0.80)	(1.31)
S1: %Strong Modal LM Words	3.925	2.845	-0.542	-3.976
	(1.21)	(0.50)	(-0.08)	(-0.35)
S1: %Constraining LM Words	-6.401	-7.855	-4.206	-9.122
	(-1.75)	(-1.87)	(-0.70)	(-1.66)
S1: %Negative LM Words	-0.249	-0.284	0.726	-1.509
	(-0.20)	(-0.21)	(0.28)	(-0.38)
IPO Controls	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	597	597	594	594
R^2	0.101	0.090	0.071	0.098

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

Table 4: Regressions with Offer Price Revision

Table 4 presents the regression results using the percentage changes in the proportion of LM word lists 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 textual analysis (i.e., Entire Document, Risk Factors, Business, and MD&A). Offer Price Revision is the percentage change in the offer price from the mid-point of the filing range. Due to space limitation, 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 Revision (in %)			
	(1)	(2)	(3)	(4)
	Entire Document	Risk Factors	Business	MD&A
Cosine Similarity	0.0451	-0.263	-0.191	0.0970
	(0.39)	(-1.56)	(-1.22)	(0.42)
PctChng: %Positive LM Words	-0.101	-0.0310	-0.0566	-0.0405
	(-1.14)	(-0.30)	(-0.62)	(-0.90)
PctChng: %Litigious LM Words	-0.197	-0.007	-0.0254**	-0.0529
	(-1.13)	(-0.06)	(-2.48)	(-1.02)
PctChng: %Strong Modal LM Words	-0.0497	0.0806	-0.0571	-0.0014
	(-0.40)	(0.60)	(-1.70)	(-0.06)
PctChng: %Constraining LM Words	0.184	-0.0802*	0.101***	0.0733
	(1.19)	(-1.99)	(3.67)	(0.98)
PctChng: %Negative LM Words	0.0711	-0.128*	0.0001	-0.0394
	(0.40)	(-2.33)	(0.00)	(-0.48)
S1: %Positive LM Words	10.12	4.402	5.647**	2.524
	(1.81)	(1.11)	(2.82)	(1.28)
S1: %Litigious LM Words	2.270	-0.550	1.395	-1.589
	(0.71)	(-0.34)	(1.02)	(-0.62)
S1: %Strong Modal LM Words	18.23**	1.590	10.27*	8.182
	(3.04)	(0.21)	(2.10)	(0.92)
S1: %Constraining LM Words	-15.51**	-0.705	-4.601	-5.603
	(-2.91)	(-0.17)	(-1.31)	(-1.20)
S1: %Negative LM Words	1.563	2.425	1.255	-2.516
	(0.52)	(1.28)	(1.11)	(-0.94)
IPO Controls	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	773	673	617	494
Adjusted R^2	0.170	0.159	0.185	0.183

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

Table 5: Regressions with First-day Returns

Table 5 presents the regression results using the percentage changes in the proportion of LM word lists during the confidential revise-and-resubmit (R&R) process (from DRS to S-1) and the levels of the proportions in S-1 filings and offer price revision. Results in each column show various sections of IPO prospectus used in textual analysis (i.e., Entire Document, Risk Factors, Business, and MD&A). First-day Returns is the percentage change from the offer price to the closing price on the first trading day. Due to space limitation, 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.

	First-day Returns (in %)			
	(1)	(2)	(3)	(4)
	Entire Document	Risk Factors	Business	MD&A
Offer Price Revision (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
Adjusted R^2	0.263	0.288	0.279	0.271

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

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. In both panels, the mean and median of Sales, EPS, EBIT, and EBITDA from the most recent fiscal year prior to IPO is reported. %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 Appendix for other variable definitions.

Panel A: Pre-IPO Operating Performance by %Positive Quintiles in 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 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 Business Section

Table 7 presents the regression results using the percentage changes in the proportion of LM word lists during the confidential revise-and-resubmit (R&R) process (from DRS to S-1) in Business section. Results in each column show common measures of post-IPO operating performance for various time periods. All operating performance measures have been scaled by lagged total assets, and winsorized at the 1% level. Due to space limitation, results for control variables and the 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.

	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	-0.0018	-0.0005	-0.0048	0.0061
	(-0.16)	(-0.05)	(-0.63)	(0.77)
PctChng: %Strong Modal LM Words	-0.0314	-0.006	-0.0313	-0.0166
	(-1.21)	(-0.45)	(-1.23)	(-1.26)
PctChng: %Constraining LM Words	0.0236	0.0033	0.0348***	0.0011
	(1.20)	(0.17)	(3.04)	(0.06)
PctChng: %Negative LM Words	0.0322	0.0155	0.0096	0.0027
	(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)
IPO Controls	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Observations	559	512	559	512
Adjusted R^2	0.437	0.420	0.370	0.340

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

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 who have filed Form RW (registration withdrawn) and have not subsequently received an Effective Notice from SEC. Seven categories of Loughran-McDonald (LM) word lists are used: Positive, Negative, Uncertain, Litigious, Strong Modal, Constraining, and Risky. Panel A reports summary statistics for the overall IPO registration process. Panel B presents summary statistics of LM word list levels (%) in entire S-1, while Panel C presents summary statistics of the percentage changes in the proportion of LM word lists during the confidential revise-and-resubmit (R&R) process (from DRS to S-1). Results in columns show the comparison of mean differences between withdrawn and completed IPOs. ***,** and * denote significance at the 1%, 5%, and 10% level respectively.

Panel A: Overall 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 amendment	872	1.5	1.4	-0.1
Number of S1 amendment	872	2.8	2.2	-0.6***

Panel B: 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: 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 who went public after previous withdrawn are excluded. Panel A reports logit regression results using the level of word list proportions in 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 Appendix for other variable definitions. All regressions include an intercept, 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 Levels of the Word List Proportions

	Withdrawal Dummy	
	(1)	(2)
	Form DRS	Form S-1
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 Lists Level in MD&A Section

Table 10 compares the levels of word list proportions in MD&A to those of in Risk Factors, Business and the Entire Document. Each category of LM word lists is defined as the percentage of words in DRS or S-1, or in specific section that are classified in that category based on Loughran and McDonald (2011) word lists.

	MD&A	Risk Factors	Business	Entire Document
	Mean	Mean	Mean	Mean
DRS Filings				
%Positive LM Words	0.68	1.17	1.49	0.85
%Negative LM Words	1.05	3.89	1.54	1.84
%Uncertain LM Words	1.42	3.17	1.32	1.63
%Litigious LM Words	0.52	1.69	0.96	1.14
%Strong Modal LM Words	0.34	0.61	0.41	0.54
%Constraining LM Words	0.66	1.25	0.68	0.79
%Risky LM Words	2.41	7.00	2.85	3.44
S-1 Filings				
%Positive LM Words	0.67	1.16	1.46	0.84
%Negative LM Words	1.05	3.88	1.53	1.81
%Uncertain LM Words	1.36	3.16	1.33	1.59
%Litigious LM Words	0.51	1.70	0.94	1.11
%Strong Modal LM Words	0.32	0.61	0.41	0.54
%Constraining LM Words	0.62	1.25	0.67	0.77
%Risky LM Words	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 JOBS Act on U.S. 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. 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; IPO Process Period is the sum of the confidential and public review period.

Panel A: Impact on IPO Process Time

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

Appendix A: Variable Definitions

Variable Name	Definition
Offer Price Revision	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.
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.
Period of IPO 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.

Appendix A (Cont.): Variable Definitions

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 Revision 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 the proportion of LM word lists 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 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
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: %StrongModal 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: %StrongModal 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)
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)
Offer Price Revision (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 Returns (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
Adjusted. R^2	0.170	0.159	0.185	0.183	0.263	0.288	0.279	0.271

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

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

This table presents the regression results using the proportion of LM word lists 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 Revision 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 LM Words	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 LM Words	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 LM Words	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 LM Words	-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 LM Words	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 Revision (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
Adjusted R^2	0.174	0.165	0.179	0.189	0.264	0.293	0.282	0.276

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

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

This table presents the regression results using the proportion of LM word lists 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 Revision 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 LM Words	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 LM Words	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 LM Words	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 LM Words	-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 LM Words	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 Revision (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
Adjusted R^2	0.171	0.164	0.178	0.186	0.259	0.292	0.281	0.272

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

Appendix E: Examples of Changes in Negative Words Used

The following table reports the frequency of three negative words ("fail", "harm", "loss") used in DRS and S-1 filings of We Co. The frequency 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 frequency of each word used are reported in the last two columns.

Frequency of Three Negative Words Used for We Co.

	DRS		S-1		S-1 - DRS	
Three Negative Words Used	Entire Document	Risk Factors	Entire Document	Risk Factors	Entire Document	Risk Factors
fail	17	15 (88%)	27	24 (89%)	+10	+9 (90%)
harm	16	14 (88%)	24	22 (92%)	+8	+8 (100%)
loss	196	11 (6%)	337	15 (4%)	+141	+4 (3%)

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.

- **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.**