Strategic Disclosure and the Pricing of Initial Public Offerings

Kathleen Weiss Hanley and Gerard Hoberg*

Current version: October 25, 2007

ABSTRACT

In this paper, we examine the word content of 2,044 initial IPO prospectuses along with their full time series of amendments. We find that the relative size of four key document sections predicts the magnitude of the partial price adjustment, first day IPO returns, and long-run post-offer performance. By assessing the word similarity between IPOs, we show that the lead underwriter is influential in the writing of the Prospectus Summary but not in the MD&A, indicating that the latter's authorship is most likely management. We find two key results that motivate a new explanation of the partial adjustment phenomenon. First, issuing firm managers perform a surprisingly integral role in the bookbuilding process as greater management disclosure generates higher offer prices and superior long-run performance. Second, litigation risk plays an important function in strategic disclosure, and only negative information learned during from bookbuilding is disclosed in amendments to the prospectus. Thus, positive information is withheld for strategic or proprietary reasons while negative information is disclosed as a hedge against litigation risk.

*Securities Exchange Commission and University of Maryland, respectively. The Securities and Exchange Commission disclaims responsibility for any private publication or statement by any of its employees. This study expresses the authors views and does not necessarily reflect those of the Commission, the Commissioners or other members of the staff. Hanley can be reached at hanleyk@sec.gov, and Hoberg can be reached at ghoberg@rhsmith.umd.edu. All errors are the authors alone.

The role of disclosure in reducing asymmetric information and whether increased disclosure is reflected in security prices is one of the central debates in finance and accounting (see Verrecchia (2001), Dye (2001) and Healy and Palepu (2001) for a review of the literature). Unlike other studies of the effect of disclosure on stock returns, which must control for prior disclosure history, firms undergoing an initial public offering are making their first large scale public disclosure via the offering prospectus. Thus, there exists, at the time of the offering, a natural experiment in which to examine the impact of differential disclosure on the bookbuilding process and the subsequent evolution of IPO pricing.

Although the average IPO prospectus is more than 50 pages long and is drafted in parts by several participants of the IPO team, an in-depth analysis of the relationship between different sections of the document has not yet been conducted. In particular, existing studies are primarily limited to examining single sections of the document in isolation.[1]

Our study employs a novel methodology that reads the entire prospectus and measures the size of the total document along with its four most important sections: the Prospectus Summary, discussion of Risk Factors, Use of Proceeds and Management's Discussion and Analysis (MD&A). By examining both the prospectus as a whole, and the likely authorship of each section, we are able to shed new light on the interaction between the legal environment and the different motives of IPO participants. We present evidence that basic relationships between these sections, which can be measured even in the initial prospectus, can predict both IPO pricing and subsequent aftermarket performance.

The legal and regulatory environment surrounding IPOs has been the focus of many papers (see for example Tinic (1988) and Lowry and Shu (2002)). These studies focus on the incentives and consequences from material omissions in the offering prospectus. Importantly, liability for these omissions is shared by issuers and underwriters alike, and damages in such cases are generally limited to the decline in the aftermarket trading price below the offer price. Classical disclosure theories would

[1]For example, Beatty and Ritter (1986), Beatty and Welch (1996), and Leone, Rock, and Willenborg (2007) examine the Use of Proceeds section while Beatty and Welch (1996)and Arnold, Fishe, and North (2006) examine the Risk Factors section.

suggest that issuers should disclose all information in order to reduce information asymmetry and thus, litigation risk. In addition, issuers that face greater legal risk will offer securities at lower prices in order to reduce the probability that the aftermarket price will fall below the IPO price.

Our results suggest that the size of the Risk Factors section is driven by a simple tradeoff. A larger Risk Factors section reduces potential legal liability and allows a higher IPO price because it reduces the probability of a material omission and subsequent litigation. At the same time, a larger Risk Factors section signals to investors that the firm is riskier, which forces the underwriter to price the IPO lower.[2] Our results suggest that the Risk Factors section is, in fact, informative regarding expected firm risk. We also find that a larger Risk Factors section leads to a higher divergence of opinion among investors, as measured by price revisions. Like Beatty and Welch (1996) and Arnold, Fishe, and North (2006), we also find that a larger Risk Factors section is associated with greater initial underpricing and inferior one-year post-IPO returns.

We recognize that IPO disclosure is influenced not only by potential legal liability but also by the relationship and incentives of IPO participants. We provide evidence that different sections of the document perform separate functions that are jointly consistent with the incentives of the author of each section and with the nature of the litigation risk.[3] For example, conversations with practitioners suggest that the Prospectus Summary is the main marketing tool used and primarily drafted by underwriters. In contrast, MD&A reflects management's assessment of the business of the firm and should be less influenced by other participants. We test these conjectures by examining the word content similarity between documents to assess the likely authorship of each of these sections and find confirming evidence that the lead underwriter is influential in the drafting of the Prospectus Summary but not in the writing of the MD&A. The authorship of these two sections has important implications for interpreting our findings and sheds new light on the different objectives and

[2] An additional reason why riskier IPOs must be priced lower is the possibility of a larger winner's curse. Uninformed investors will demand a lower issue price to compensate them for greater losses to informed investors.

[3] For example, Field, Lowry, and Shu (2005) suggest that the choice of disclosure can potentially deter certain types of litigation.

contributions of each IPO participant.

Our results on the role of disclosure by underwriters through the Prospectus Summary are consistent with classical theories that suggest that greater disclosure can reduce information asymmetry between the firm and its shareholders (e.g. Diamond and Verrecchia (1991) and Easley and O'Hara (2004)). The greater is the relative size of the Prospectus Summary, the lower is the change in the offer price during the bookbuilding process and the lower is the subsequent initial return. We interpret these findings as an indication of the potential for underwriter disclosure to increase the efficiency of IPO pricing.

Although classical theories of disclosure predict that managerial disclosure will reduce information asymmetry and will lead to smaller changes in offer prices and lower initial returns, we find the opposite to be true. Uniformly, larger MD&A sections are followed by large positive changes in the offer price during bookbuilding. This result is invariant to whether the final offer price is above or below the midpoint of the file range. We find no corresponding link to initial returns. Most surprising, larger MD&A sections are followed by superior one-year post-IPO abnormal stock returns.

Our findings suggest that the initial price range ignores information contained in the MD&A section, but that this information is incorporated later during the bookbuilding process. Kim and Ritter (1999) document that initial offer price ranges are primarily set using accounting information and comparable firm multiples, and further state that the "additional information they (underwriters) process about the market's demand results in more accurate pricing." The authors do not explain the source of this improvement in accuracy. Our results suggest that part of this improvement comes from management. While traditional theories of bookbuilding such as Benveniste and Spindt (1989) have focused on the role of regular investors in providing information to the underwriter, our results indicate that additional information provided by management can also lead to higher offer prices. The positive nature of this information is genuine, as investors who listen to management are rewarded with superior post-IPO abnormal returns.

Although other studies have found a relationship between the number and specificity of uses of proceeds and initial returns, we find little evidence that the Use of Proceeds section has an impact on IPO pricing. This may be due to the relatively small contribution this section makes to the prospectus in terms of characters, or to the fact that the relative section size might not be highly correlated with the number and specificity of proceed uses.

Finally, we examine how prospectus disclosure changes during the bookbuilding process. While traditional theories of disclosure propose that more disclosure reduces information asymmetry, other theories suggest that increased disclosure can be harmful because it reveals strategic or proprietary information to rivals (e.g. Darrough and Stoughon (1990) and Maksimovic and Pichler (2001)). Therefore, the IPO team has an incentive to not only fully reveal bad information to protect against liability, but also to withhold positive information for strategic reasons. Our results are consistent with this view. When positive information is revealed during the offering process, as indicated by a positive revision in offer prices, there is no corresponding change in the information content of amended prospectus filings. Conversely, when negative information is revealed, the content of the prospectus changes significantly. Changes in content are followed by both price reductions and lower initial returns.

Our empirical findings motivate two key extensions to classical theory. First, although classical bookbuilding theory credits only investors with information production, we find that management also plays a central role, especially in an upward direction. This suggests that issuing firm managers act on behalf of shareholders to obtain the highest offering price possible. The road show provides a likely mechanism and stage for managers to perform this duty. Second, although classical disclosure theory suggests that issuers will disclose all available information to reduce information asymmetry, we find that issuers, when amending the initial prospectus, disclose only negative information. This is consistent with an incentive to protect proprietary information, and to disclose only information that is most critical to avoiding legal damages.[4]

[4]Disclosing negative information provides better lawsuit protection than disclosing positive information because damages are generally limited to investor losses when the share price sinks below the IPO price.

Our findings provide a new explanation of the partial adjustment phenomenon. When bookbuilding reveals positive information, this new information may be withheld from investors due to its proprietary value, leaving the issuer and underwriter especially prone to litigation risk. Hence, partial adjustment arises because underwriters set the IPO price lower to mitigate this increased litigation risk, and more importantly, to preserve their reputational capital. Because reputational capital is especially valuable given its link to future business (Hoberg (2007)), this might explain why virtually all IPOs experiencing upward price revisions are especially deeply underpriced.

The remainder of the paper is organized as follows: A summary of the relevant literature is in discussed in Section I. The data, methodology and summary statistics are presented in Section II. The determinants of the size of the initial prospectus and its subsections as well as its impact on price changes and aftermarket pricing are discussed in Section III. The impact of information revealed on changes in the prospectus as well as the relation of these changes on aftermarket pricing is discussed in IV. The paper concludes in Section V.

I Literature Review and Hypotheses

There has been an extensive discussion of both mandated and discretionary disclosure and its impact on the cost of capital (see Verrecchia (2001), Dye (2001) and Healy and Palepu (2001) for a review of the literature.) The primary difficulty of determining the impact of disclosure choices on stock prices, as noted by Core (2001), is "that the US disclosure environment is already so rich that it would be difficult to find strong disclosure-related effects in broad cross-sections of US firms." Thus, he argues that disclosure represents only second-order effects, which could only be detected when there is a large change in disclosure policy.

The IPO process, therefore, provides a unique opportunity to examine the effect of disclosure on the offering process and subsequent pricing. Unique liability concerns at the time of the IPO favors disclosing as much information as possible, even though that information may be noisy and possibly, uninformative. The issuer and

its underwriter are liable for any material omissions in the prospectus and any damages are calculated as the decline in the market trading price from the offer price. Tinic (1988) and Hughes and Thakor (1992) hypothesize that IPOs require more underpricing as insurance against liability risk. Lowry and Shu (2002) argue that firms deciding to go public have incentives to insure against this risk by performing due diligence.

The central tension in the determinants of disclosure (in the absence of litigation concerns) and its impact on IPO pricing is the tradeoff between providing additional information to investors which may reveal strategic or proprietary information to competitors and maximizing the proceeds to the issuing firm. The assumption underlying many models of disclosure is that increasing the amount of information provided to investors decreases the firm's cost of capital by reducing information asymmetry. However, there may be instances in which additional disclosure may reveal valuable strategic information to rivals which, in the long run, may adversely affect shareholder welfare (see for example, Darrough and Stoughon (1990), Bhattacharya and Chiesa (1995), and Maksimovic and Pichler (2001)).

Evidence that greater disclosure reduces information uncertainty in an IPO context, is provided by Guo, Lev, and Zhou (2004) who focus on product related disclosures in the prospectus by firms in the biotech industry. The authors construct a product disclosure index and relate this index to various IPO characteristics as well as its impact on bid-ask spreads. They find a negative relation between the extent of disclosure and the bid-ask spread but do not provide an analysis as to the impact of the index on IPO underpricing. In this paper, we argue that increasing disclosure should information asymmetry and therefore, mitigate potential changes in the offer price during the bookbuilding process and reduce initial returns.

Prior research on the role of disclosure in the Use of Proceeds section and the pricing of IPOs has shown mixed results. Beatty and Ritter (1986) find a positive relation between the number of uses of proceeds and underpricing which they conclude is consistent with higher uncertainty regarding the issue. Beatty and Welch (1996) find no relation between the number of uses and subsequent initial returns. Leone, Rock, and Willenborg (2007) examine the specificity of the uses of proceeds in the

IPO prospectus. Specificity is defined as the extent of dollar specificity within the Use of Proceeds section. They find that an increase in specificity is associated with a decline in underpricing. The authors suggest that specificity reduces the information asymmetry problem faced by investors. Ljungqvist and Wilhelm (2003) find that firms citing the funding of operating expenses (less specificity) as the primary use have higher underpricing. Since our technology is unable to measure specificity, we argue that increased disclosure is beneficial to reducing the information asymmetry and therefore, hypothesize that the greater the Use of Proceeds section, the lower should be the adjustment in offer price and subsequent underpricing.

To our knowledge, we are the first study to examine the role of the Prospectus Summary and MD&A and we do so in the context of the likely authorship of these two sections. The Prospectus Summary is the primary marketing tool used by underwriters, while MD&A is management's assessment of the financial condition and outlook of the firm. Thus, we argue that underwriters and managers may be able to reduce the information asymmetry between the issuing firm and potential investors by disclosing additional information in both the Prospectus Summary and MD&A. Therefore, we hypothesize that greater disclosure in both sections should mitigate any revisions in the offer price and also reduce the subsequent initial return.

Recent papers on media and company press releases have highlighted the importance of disclosure for IPO pricing (Cook, Kieschnick, and Ness (2006), Schrand and Verrecchia (2005), and Liu, Sherman, and Zhang (2007)). Thus, our work contributes to the growing body of literature on the complexity of the disclosure process surrounding IPOs.

II Data and Methodology

A Data and Initial Prospectus Variables

IPO characteristics data are from the Securities Data Company (SDC) U.S. New Issues Database. The sample initially consists of all U.S. IPOs issued between January 1, 1996 and October 31, 2005. We eliminate ADRs, unit issues, REITs, closed-end funds, financial firms, and firms with offer prices less than five dollars. A CRSP

permno must also be available for an observation to remain in the sample, and the IPO must also have a valid founding date, as identified in the Field-Ritter dataset, as used in Field and Karpoff (2002).[5] These initial exclusions reduce the sample to 2,112 IPOs.

For each IPO passing these initial screens, we use a web crawling algorithm to download its entire series of prospectus filings. This includes both the IPO's initial prospectus, and also its entire series of prospectus amendments that are filed up until the given firm's effective date. We do not include the final prospectus itself in this series (Form 424a or 424b). In order for an IPO to remain in our sample, it must have SEC Edgar filings available online, and the online documents must also be machine readable. In order to satisfy our definition of machine readable, a Table of Contents pagination algorithm must be able to detect, and accurately identify, the start and end of the four key sections of the prospectus. These sections are the "Prospectus Summary", "Risk Factors", "Use of Proceeds", and "Management's Discussion and Analysis".[6] This additional screen eliminates 68 IPOs, leaving us with 2,044 machine readable IPOs. Because these 68 IPOs are a small fraction of our sample, and because most are also small firms using the SB-2 filing method (larger firms generally file use the S-1 filing method), we do not believe that omitting these firms induces any bias into our sample.

Our algorithm to read each prospectus or amendment is written in a combination of PERL and APL, and the methodology used to construct each variable is presented in Appendix 1. We store the text of the prospectus in a character vector, which we define as $chars_{tot}$. Next, we store the text from the each of these four sections in separate character vectors, which we define as $chars_{ps}$, $chars_{rf}$, $chars_{use}$, and $chars_{mda}$, respectively and construct the following variables for use in our price and prospectus regressions:

[5] We thank Jay Ritter for generously providing the database of IPO founding dates on his website.

[6] A significant amount of work has been done to maximize the fraction of prospectuses that are deemed machine readable. This includes hand-checking each prospectus failing our machine readability condition to determine if our document pagination algorithm can be improved via exception handling. An example of an exception is that some filings have slight variations to the section names which we list. For example, the Prospectus Summary is occasionally called "Summary". The 68 IPOs failing machine readability generally lack pagination or may even lack a Table of Contents.

totchars: The number of characters in the text vector $chars_{tot}$.

ps_{pct}: The relative size of the Prospectus Summary section. This is defined as the ratio of the number of characters in the text vector $chars_{ps}$ divided by the number of characters in the text vector $chars_{tot}$.

rf_{pct}: The relative size of the Risk Factors section. This is defined in a parallel fashion as ps_{pct} using $chars_{rf}$.

use_{pct}: The relative size of the Use of Proceeds section. This is defined in a parallel fashion as ps_{pct} using $chars_{use}$.

mda_{pct}: The relative size of the MD&A. This is defined in a parallel fashion as ps_{pct} using $chars_{mda}$.

We compute a number of variables that are common to the existing IPO literature.

$$\Delta P = \frac{P_{ipo} - P_{mid}}{P_{mid}}, \qquad IR = \frac{P_{mkt} - P_{ipo}}{P_{ipo}}. \tag{1}$$

P_{mid}, P_{ipo}, and P_{mkt} are the filing date midpoint, the IPO price, and the after-market trading price, respectively, ΔP is underwriter's price adjustment from the filing date to the IPO date, and IR (initial return) is the market's price adjustment from P_{ipo} to P_{mkt}. Investors who purchase shares at the IPO price P_{ipo} can realize returns equal to IR by selling their shares at the closing price on the first day of public trading.

We also compute one-year post IPO abnormal returns as the intercept of a regression of excess daily stock returns (raw returns minus the riskless thirty-day T-bill rate) on the three Fama-French factors (MKT, HML, SMB) plus momentum (UMD):

$$r_{i,t} - r_f = \alpha + \beta_1\ MKT + \beta_2\ HML + \beta_3\ SMB + \beta_4\ UMD + \ \epsilon \tag{2}$$

We compute one such regression for each IPO, and one observation is one daily return realized on the IPO date up until the IPO's one year anniversary. We also account for the following variables identified in the existing IPO literature:

$\Delta P+$: ΔP+ is the positive component of ΔP: $\max[\Delta P, 0]$. This variable controls for the partial adjustment phenomenon documented in Hanley (1993). This form was first used in Lowry and Schwert (2002).

$\Delta P–$: Negative price adjustment $\min[\Delta P, 0]$.

Firm Age: IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002).

UWdshare: Lead underwriter's dollar market share in the past calendar year. This variable was first used in Megginson and Weiss (1991).

LAWdshare: This variable is calculated as the dollar market share in the past calendar year and a separate variable is constructed for the lead underwriter's legal counsel and the issuer firm's legal counsel.

VC: Dummy variable equal to unity if the firm is VC-backed, and zero otherwise. This was first studied in Barry, Muscarella, Peavy, and Vetsuypens (1990).

Mkt30: We construct two variables of this sort. Our first is the NASDAQ return for the 30 trading days preceding the filing date. Our second is the NASDAQ return for the 30 trading days preceding the issue date. Logue (1973) first examined whether past market returns can predict future underpricing, and this measure has been used more recently by Loughran and Ritter (2002).

LogSize: We consider two variables of this sort. Our first is the natural logarithm of the original filing amount. Our second is the natural logarithm of the offering amount.

Tech Dummy: Dummy variable equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004).

Although not reported, we also collect data on revenue and assets prior to the offer from SDC. Our results are robust to the inclusion of these variables but the size of the sample is significantly reduced.

B Summary Statistics

Table I presents summary statistics for the 2,044 IPOs in the sample. Panel A has information on the price variables and our sample is similar to other studies that include the bubble period of 1999 and 2000. On average, this sample of IPOs has an average initial return of 33% with a much lower median of 12%. The average upward price adjustment from the midpoint of the file range is almost 11% and approximately 47.1% of the companies in the sample revise their offer prices upward. 38.4% percent of the sample IPOs have a downward price movement and the corresponding average decline in the offer price from the midpoint of the file range is -7.5%. The remaining 14.5% do not experience any price adjustment from the filing midpoint. The one year post-IPO abnormal return is not significantly different from zero.

Panel B consists of statistics on IPO characteristics. There is substantial variation in offering characteristics within our sample. The mean IPO files an offer amount of approximately $187 million. At the time of the IPO, this average is much smaller at $115 million. The mean age of the firm at the time of the offering is 14 years but the median is significantly smaller at 7 years of age. Forty-four percent of the IPOs are classified as Tech firms as in Loughran and Ritter (2004) while 47% have venture capital backing. The average market share of the underwriter in the year prior to the offer is 2.9% with an affiliated law firm market share of 1.2%. The average market share of the issuer's counsel is greater than that of underwriter counsel at 2.3%.[7] Consistent with Lowry and Schwert (2002), IPOs are brought to market when prior returns are high with an average return in the thirty days prior to filing or offer of approximately 5%.

[7]This is a surprising result since one argument for underpricing is that underwriters are repeat players but issuers are not. Therefore, issuers are at a competitive disadvantage to underwriters who may prefer, for a variety of reasons, a lower offer price to a higher one. One possible interpretation of the finding that the average issuer counsel tends have greater market share than the average manager counsel is that a sophisticated issuer would recognize their disadvantage in the IPO process and would rely on the expertise of others including their counsel.

III The Initial Prospectus

Table II presents summary statistics describing the initial prospectus allocation. The average (and median) prospectus has just over 200,000 characters of which 6% is the Prospectus Summary, 18% are Risk Factors, less than 1% are Use of Proceeds and 13% consists of the MD&A. Overall, these four sections, on average, comprise 38% of the entire prospectus.

The small size of the Use of Proceeds section is somewhat surprising given the results of Leone, Rock, and Willenborg (2007) who find that an increase in the specificity of the intended use of proceeds reduces subsequent underpricing. This finding suggests that even small sections of the prospectus can convey important information to investors. If this is the case, then our tests are biased toward the null hypothesis which suggests that the size of the prospectus and the corresponding sections should have no impact on IPO pricing.

Panels B and C of Table II present the correlation coefficients of both the raw character sizes and the relative section sizes. As expected, larger prospectuses have larger individual sections as measured by raw character size. The exception is the Use of Proceeds section which is uncorrelated with any other section including the size of the prospectus as a whole.

The percent of the document devoted to each section presents a different picture due to the fact that this variable, in some sense, measures the tradeoffs the firm and its underwriter make in deciding how much of the entire document to allocate to the various sections. Larger documents tend to have a larger proportion devoted to the Prospectus Summary and MD&A. Note that this does not imply that larger Prospectus Summaries are correlated with large MD&As as the correlation between the two is insignificant. This lack of correlation is consistent with the separation of authorship we document later. In contrast, the size of the Risk Factors section is negatively correlated with total document size and the proportion of the document that is composed of either the Prospectus Summary or MD&A. The Use of Proceeds section is uncorrelated with the proportion of the document devoted to the Risk Factors section and Prospectus Summary and negatively correlated with the size of

the total document and MD&A.[8]

A Determinants of the Initial Prospectus

Table III presents the determinants of the document as a whole, each of the four subsections, and the combination of the four subsections. Larger document sizes are associated with larger offerings, more prestigious underwriters and law firms as well as venture capital backing. The percent of the prospectus that is composed of the Prospectus Summary is larger when the offering is larger and when the firm is older. VC-backed and tech firms tend to have smaller Prospectus Summaries. As support for the hypothesis that the underwriter views the Prospectus Summary as important in the marketing of the IPO, the size of the Prospectus Summary is significantly and positively related to the prestige of the underwriter but unrelated to the prestige of either the issuer or underwriter law firm.

The size of the Risk Factors section is correlated with factors that proxy for the ex ante risk of the issue such as low age of the firm, small expected proceeds and low prestige of the underwriter.[9] Firms with large Risk Factors sections are more likely to be VC-backed and tech firms as well. These findings support our broader conclusion that the Risk Factors section is indeed informative regarding actual firm risks.

Leone, Rock, and Willenborg (2007) find that the greater Use of Proceeds specificity is significantly related to higher age, larger and non-tech firms. In contrast, however, we find that the Use of Proceeds section, as a percent of the prospectus document, is unrelated to all of these factors with the exception of VC-backing. Further, the R^2 of regression is very low at only 3.2%. The lack of significance on the size of the Use of Proceeds section indicates that the size of the section may not be a good proxy for specificity.[10] Unlike Beatty and Ritter (1986) we find little evidence

[8]We obtain similar results if we normalize section sizes by the number of characters in the complementary part of the document (sections other than our four key sections). Hence, our results are not driven by correlations between the sections induced by the fact that the relative sizes are bounded in the interval (0,1).

[9]This finding is consistent with Arnold, Fishe, and North (2006).

[10]We are unable to determine the relative R^2 between our results and theirs due to the fact that the authors use a Tobit specification.

that the Use of Proceeds section is associated with ex ante risk.

Consistent with our hypothesis that a larger MD&A is associated with firms that are more mature and have more technical operations, MD&A, as a proportion of the prospectus, is larger when expected proceeds are higher and the firm is older. Firms with large MD&As are also less likely to have VC-backing and are more likely to be classified as tech firms. Unlike the Prospectus Summary, we document that neither the prestige of the underwriter nor its counsel has a significant impact on the size of MD&A, which is consistent with the notion that this section of the document is primarily management's and not the underwriter's responsibility.

B Evidence of Authorship

The preceding section presented preliminary evidence that authorship may differ between sections of the prospectus. We are therefore interested in whether we can ascertain the potential authorship of each section. In order to do so, we first construct a variable that measures the degree of similarity between documents, a measure we call "document similarity". We then examine whether IPOs brought to market by the same underwriter and/or issuer or manager counsel exhibit greater similarity. This test allows us to explore whether there is a "signature" associated with each of the participants and how this "signature" is manifested in each section of the document.[11]

The dependent variable we use to measure authorship of a section is the document similarity between two initial IPO prospectuses. In Appendix 2, we explain in detail how we compute document similarity. This is a numerical variable bounded in the interval [0,1]. A value of zero indicates that the two documents have exactly the same distribution of word roots being used. A value of one indicates that the documents are entirely different and have no word roots in common. One observation is one pair of IPOs i and j, and we include all unique IPO pairs as observations (we exclude pairs in which $i = j$). For our sample of 2044 IPOs, a maximum of $\frac{2044^2-2044}{2}$ unique

[11] Ideally, we would like to use this test to ascertain an issuer signature. Unfortunately, this is impossible since the vast majority of issuers only go public once. Therefore, we can only determine a noisy indicator of issuer authorship which is proxied by the issuer's counsel.

pairs exist, and hence a maximum of 2,087,946 observations appear in any regression (fewer appear in some specifications as some sections are missing for a small number of IPOs). To ensure T-statistics remain unbiased given the repeated use of each document, we report T-statistics that are adjusted for clustering by IPO.

The first three explanatory variables we consider are dummy variables identifying whether IPOs i and j had the same lead underwriter, the same manager's counsel, and the same issuer's counsel. When more than one underwriter serves as lead, and i and j share at least one lead underwriter, we assign the "same lead underwriter" variable a value equal to the number of common underwriters divided by the maximum number of underwriters associated with either IPO. The next four dummy variables are one if IPO i and j reside in the same one digit to four digit SIC code, respectively.[12] We also include a dummy variable identifying whether IPO i and j are issued in the same year, and a dummy indicating whether both are Tech oriented as identified in Loughran and Ritter (2004). Finally, we include four variables that capture how different IPO i and j's characteristics are using the log of firm age, the IPO year, the log of filing size, and the underwriting spread. We then calculate the absolute value of the difference in characteristics for IPO i and j. Larger values of each characteristic indicate that i and j differ more with regards to a given characteristic.

Table IV presents a series of regressions based on the document similarities of the prospectus as a whole and of the individual sections. The underwriter's total signature is the sum of the Same Lead UW and Same UW Counsel coefficients. In Panel A, the influence of the underwriter on the content of the entire document is very high. Once the document is parsed into the relevant sections, however, the influence of the underwriter and its counsel on the individual sections is most pronounced for the Prospectus Summary (Panel B). The magnitude of the underwriter's impact on the document similarity of this section even exceeds the sum of all the industry variables in some specifications.

The influence of the underwriter in the remaining sections is far lower. There is still a positive relationship between the same underwriter and the degree of similarity in the Risk Factors section (Panel C). This is consistent with both the underwriter

[12]Thus, the total impact of being in the same industry is the sum of the four coefficients.

and the issuer bearing the risk of a material omission in the prospectus and using the Risk Factors section as a hedge against future liability.

More importantly, for our purposes, is the drop in the magnitude of the underwriter's signature from the Prospectus Summary to the MD&A in Panel E. The combined underwriter coefficients are significantly less in MD&A than in any of the other three sections, particularly the Prospectus Summary. Further, the reduced predictability of the remaining variables in the MD&A indicates that this section is more idiosyncratic and thus, likely issuer driven.[13]

Overall, the findings on authorship indicate that the underwriter is influential in the drafting of the entire document but its influence is most pronounced in the Prospectus Summary. In contrast, the imprint of the underwriter in the MD&A is significantly reduced consistent with our conjecture that this section of the prospectus most likely reflects the views of management and thus the issuer.

C Effect on Changes in Offer Price

In this section, we examine whether the amount of information in the initial prospectus has predictive power for price changes during the bookbuilding process despite the fact that, frequently, the initial prospectus does not include any information regarding the expected offer price. We hypothesize that changes in offer prices are related to the dispersion of opinions of investors regarding the IPO's true value. In order to reduce the potential for dispersion of beliefs, the issuing firm and/or the underwriter could conceivably convey more information to investors through the prospectus. Therefore, we expect that larger prospectuses with larger Prospectus Summaries, Use of Proceeds sections and MD&A should result in a lower change in the offer price during the bookbuilding period and lower subsequent underpricing.

[13]When identifying the marginal impact of the independent variables, for example, same underwriter versus same industry, it is important to note that while the total coefficient impact of having the same underwriter is equal to or even greater than that of being in the same exact SIC4 industry, the latter generates a significantly larger marginal improvement to R^2. This is because far fewer IPOs have the same lead underwriter as those that have the same industry. Even though the underwriter's signature exists for *every IPO*, we can only observe it when the IPOs have the same lead underwriter. Thus, if we could measure the underwriter signature for every IPO, it most likely would have a greater impact on R^2 than industry alone.

Consistent with our view of the incentives created by the legal environment in IPOs, we hypothesize that the Risk Factors section contains information on both the overall uncertainty surrounding the firm as well noise due to incentives to aggressively enlarge this section because of its role as a hedge against liability. Therefore, we expect that the greater the Risk Factors section, the greater should be the dispersion of beliefs which should increase both changes in offer prices and underpricing.

Table V presents OLS regressions on ΔP, and Tobit regressions on ΔP+ and ΔP-. Panel A of Table V presents an OLS regression where the dependent variable is the percentage change in the offer price from the midpoint of the file range (ΔP). The sign and significance of the control variables in the regression echoes the findings in the IPO literature. Greater ex ante uncertainty as measured by lower firm age, smaller expected proceeds and tech companies, are all associated with greater price adjustments. Higher prestige underwriters and VC-backed IPOs also have greater price adjustments, consistent with possible evidence of access to better informed investors. Finally, the return in the Nasdaq index over the 30 days prior to filing has a positive and significant impact on the change in offer price.

Although traditional disclosure theories suggest that greater disclosure should reduce information asymmetry and thus the potential for a change in the offer price, we find no evidence that the size of the total prospectus has any effect on the size of the price adjustment.

An examination of the individual sections, however, suggests a different story. As expected, a larger Prospectus Summary reduces the overall size of the offer price adjustment. We interpret this findings as support for the conjecture that greater information contained in the Summary conveys more precise information by the underwriter to investors ex-ante that, in turn, lessens the magnitude of the ex-post price adjustment. In contrast, the Risk Factors section increases the magnitude of the offer price adjustment. The greater is the size of the Risk Factors section in the prospectus, the greater is the likelihood that investors will have a dispersion of beliefs regarding the value of the firm and hence, a larger change in the expected offer price. This finding is consistent with our argument that the Risk Factors section contains real information about firm uncertainty that increases the likelihood of an

adjustment in offer prices and suggests that underwriters and/or issuers may be more conservative when setting the initial filing range for offers that are potentially more risky. Unlike the findings of some of the previous literature, there is no evidence that the size of the Use of Proceeds section has any effect on the change in offer price.

Although we hypothesized that the larger the MD&A, the lower should be the offer price adjustment, the results in Panel A indicate otherwise. We find that the larger the MD&A, the greater is the price adjustment. Thus, it appears as if the underwriter discounts this information when setting the initial offer price and it is not until bookbuilding is complete that this information becomes incorporated into offer prices.

The result on MD&A points to the potential value of the roadshow in conveying more precise information to investors.[14] The roadshow allows the management of the firm the opportunity to explain the information contained in the MD&A as well as mitigate the impact of the information in the Risk Factors section. Firms with larger MD&A are more likely to have good information revealed during the road show which translates into higher offer prices. Pava and Epstein (1993) examine the eventual realization of disclosures in the MD&A and find that "management is much more likely to correctly anticipate and disclose good news relative to bad news."[15] Note, however, that the potential for legal liability constrains management from being overly-optimistic. Suppose management falsely reveals good information in order to increase the offer price. Once the market learns that the good information provided by management was false, the market price will fall below the offer price. This decline will lead to subsequent shareholder lawsuits.

The asymmetry of disclosure and its impact on price adjustment is examined in last two panels of Table V, which split the sample of IPOs to those that have a positive price adjustment (Panel B) and those that have a negative price adjustment (Panel C). There is a substantial difference in the magnitude of the price adjustments associated with the prospectus as a whole as well as the subsections based on whether good information or bad was revealed during the offering process. Similar to the

[14]By regulation, information conveyed to the investors during the road show is to be limited to the information in the prospectus.

[15]This quote is included in Bryan (1997).

results for the price adjustment as a whole, smaller Prospectus Summaries are related to larger positive price adjustments but this section is not related to the size of the negative price adjustment.

Unlike Panel A, we find a slightly significant and negative relationship between the Use of Proceeds section and $\Delta P+$ in Panel B. The greater is the amount of information conveyed in the Use of Proceeds section, the smaller is the positive partial adjustment. Thus, one interpretation of the results of Leone, Rock, and Willenborg (2007) is that the specificity in the use of proceeds reduces underpricing because it also reduces the magnitude of the offer price adjustment. This finding is also consistent with the conjecture that greater specificity reduces information asymmetry among investors and hence, lowers the potential dispersion of valuations. There is no relation, however, between the relative size of the Use of Proceeds section and $\Delta P-$

The findings on the size of the Risk Factors section indicate that its impact is only for positive changes in the offer price. The relative size of this section is unrelated to the magnitude of negative changes in offer prices.

Like the Risk Factors section, larger relative sizes of MD&A increase the change in the offer price from the initial filing date to the offer date. Unlike other sections, however, only greater disclosure in the MD&A affects and reduces the magnitude of $\Delta P-$. This means that management's disclosure results in higher offer prices regardless of the type of information that is revealed during bookbuilding. This result is consistent with management efforts to maximize the final offer price. While the IPO literature has primarily focused on the role of the underwriter, the ability of management to influence offer prices has not been studied. These results highlight the potentially important role management may play in the offering process and their ability to significantly influence the setting of the final offer price.

D Predictability of Initial Returns and Post-Offering Returns

This section examines whether the information content of the initial prospectus can predict subsequent short- and long-run returns. Table VI, Panel A, presents the results of an OLS regression using first day returns as the dependent variable. In addition, we control for the same factors as in the previous section that are known to affect initial returns.

While the size of the entire prospectus has no effect on underpricing all but the Use of Proceeds section is significantly related to initial returns. Consistent with the hypothesis above, larger Prospectus Summaries significantly reduce the amount of underpricing on the first trading day holding other characteristics of the offer constant. This finding suggests that additional disclosure in this section of the document is able to reduce the amount of money left on the table. If underwriters face heterogeneous incentives to disclose information, perhaps due to a heterogeneous tradeoff between reputation building (long-term profitability) and current profitability, this finding might also explain a fraction of the underwriter persistence reported in Hoberg (2007).

Consistent with prior literature, we find that the Risk Factors section of the prospectus leads to higher underpricing. Thus, we confirm the findings of Beatty and Welch (1996) and Arnold, Fishe, and North (2006) that greater risk factors may be associated with greater ex ante uncertainty regarding the valuation of the firm. The relationship of the size of the Risk Factors section is compatible with the conjecture of Lowry and Shu (2002) in which, the issuing firm and its underwriter, to hedge against liability, discount the offer price as insurance against the higher probability that a bad outcome will occur.

Unlike Beatty and Ritter (1986) and Leone, Rock, and Willenborg (2007), we find no evidence that the relative size of the Use of Proceeds section impacts underpricing even though an increase in this section creates greater positive changes in offer prices. As noted previously, the size of this section, relative to both the entire document and the other subsections, is very small. Therefore, we acknowledge that the correlation

between size and specificity may also be small and hence, our results do not confirm the hypothesis that more information, as measured by relative section size, has an impact on aftermarket pricing.

Although, we hypothesized that more information in MD&A should reduce information asymmetry, the size of MD&A has no marginal relation to subsequent underpricing. This is somewhat surprising given the fact that the relative size of MD&A results in greater adjustment in offer prices.

We examine the post-offering returns of IPOs in Panel B of Table VI and find that both the Risk Factors section and MD&A matter. The coefficient on the Risk Factors section is negative indicating that firms listing more risk factors underperform. Arnold, Fishe, and North (2006) find a similar result and conclude that this negative association is due to "the realization of some of the named prospectus risk factors." An alternative interpretation is that investors underestimate the true risk of the firm. This leads to overvaluation at the time of the offering and subsequent price declines.

More importantly, we find that the amount of information provided by management through MD&A is positively related to the long-run performance of the firm. This finding supports the credibility of managerial disclosures and suggests that investors who listen to management are rewarded with superior long-run performance.

IV Changes in the Prospectus

This section examines whether new information gathered through the road show becomes incorporated into amendments to the initial prospectus.[16] Although the SEC requires the issuing firm to disclose all material information in the prospectus, the cost of withholding valuable, positive information may be small. Because the underwriter and issuing firm are only liable for declines in market prices below the offer price, the impact of withholding good information is simply to reduce the expected offer price. However, the omission of bad information from the prospectus results in higher than expected offer prices that subsequently fall in the aftermarket once the

[16]Information may also be provided from SEC review of disclosure documents.

bad information is revealed thereby increasing the potential for a lawsuit. Thus, the tension between disclosing valuable positive information to potential rivals and liability concerns should lead to an asymmetry in the disclosure of information obtained during the offering process. We hypothesize that good information revealed during the bookbuilding process will not be incorporated into changes in the prospectus, while bad information that is revealed will lead to revisions in the prospectus.

Because revisions or amendments filed following initial prospectuses are also machine readable, and because they generally follow the same format as initial prospectuses, we are also able to paginate and separately process each amendment (revision), and compare it to the initial prospectus. In particular, we examine how the relative size of each key section changes over time, and we also measure the severity of revisions relative to the original document over the offering period.

To examine the severity of revisions, we first compute the "document distance" (defined in Appendix 2) between neighboring documents in each IPO's time series of amendments. For an IPOs entire time series of amendments, we then compute the "normalized document distance" as the normalized sum of these distances. Appendix 3 explains this calculation in detail. The normalization mitigates the impact of extreme observations, and also mitigates the impact of the first revision following the initial prospectus. This first revision is often quite substantial, and hence raw distances tend to have extreme outliers.[17] For later amendments, we generally find varied, but less extreme heterogeneity in distances. Some firms experience very little document distance and converge quickly to a final revision. Other firms experience second or even third waves of substantial revisions. These later waves can impose document drifts similar in magnitude to that of the first revision. Hence, this normalized measure can be viewed as a rough count of the number of substantive revisions experienced during the filing process. We consider the following variables:

Δ *totchars*: The percentage change (from initial prospects to final amendment) in the number of characters in the text vector $chars_{tot}$ from the initial prospectus to the final revision.

[17]The results of our study are robust to using the first amendment as the initial prospectus.

Δps_{pct}: The change (from initial prospects to final amendment) in the relative size of the Prospectus Summary section. This is defined in a similar fashion as Δ totchars, except it is based on $chars_{ps}$.

Δrf_{pct}: The change in the relative size of the Risk Factors section. This is defined in a similar fashion as Δ totchars, except it is based on $chars_{rf}$.

Δuse_{pct}: The change in the relative size of the Use of Proceeds section. This is defined in a similar fashion as Δ totchars, except it is based on $chars_{use}$.

Δmda_{pct}: The change in the relative size of the MD&A. This is defined in a similar fashion as Δ totchars, except it is based on $chars_{mda}$.

Δ*normdist tot*: The total normalized distance from the initial prospectus to the final revision for the entire prospectus document.

Δ*normdist ps*: The total normalized distance for the Prospectus Summary section.

Δ*normdist rf*: The total normalized distance for the Risk Factors section.

Δ*normdist use*: The total normalized distance for the Use of Proceeds section.

Δ*normdist mda*: The total normalized distance for MD&A.

Table VII shows the summary of prospectus filing patterns. The vast majority of IPOs have at least three amendments to the initial prospectus. This number begins to rapidly decline with only a few of the remaining IPOs having seven or more amendments. As can be seen in the columns denoted by the number of characters in Table VII, the number of characters in each subsequent prospectus tends to increase in order to incorporate both SEC comments and information acquired during the road show. In terms of each subsection, the Prospectus Summary and Use of Proceed section remain relatively constant with an average of around 14,000 and 2,000 characters, respectively. In contrast, additional information is added to both the Risk Factors section and MD&A. From the filing of the initial prospectus to the fifth amendment, the Risk Factors section increases from around 38,000 characters to almost 43,000 characters. The increase in MD&A is even greater, as the size

over the same number of amendments goes from 29,000 characters to almost 37,000 characters.

The change in content, as measured by the document distance from the previous amendment, is greatest for the first revision (for both the full prospectus and individual subsections). The amount of new information that is incorporated into subsequent revisions by the fifth amendment declines fairly rapidly for all but the Use of Proceeds section. These findings are not surprising given that only roughly half of the sample has five or more amendments. Overall, we conclude from Table VII that new information is indeed disclosed during the offer process. The remainder of this section examines both the determinants of this change in disclosure and whether the new information impacts underpricing and long-run returns.

A The Determinants of Changes in Prospectus Variables

Table VIII presents the summary statistics associated with the number of amendments, days in registration and overall changes in the prospectus that will be examined in further detail. Panel A summarizes the amendments and days in registration. The average IPO spends almost 94 days in registration, files four amendments of which one is late in the filing process. Panel B documents changes in the size or allocation of the prospectus and corresponding subsections. Only the prospectus as a whole has a substantial average change. Each of the mean or median individual subsections show little or no change although there is substantial variation among IPOs.

Although the average size of the sections appears to remain fairly close to the initial prospectus, the change in the content of the subsections, as measured by the normalized distances in Panel C, shows substantial variation.

Table IX presents OLS regressions on the amendments and registration and the changes in both the document allocation and normalized distances for the document as a whole as well as the four subsections. In order to capture the potential incentives for asymmetric disclosure, both ΔP+ and ΔP- are included as independent variables along with other control variables used previously.

In Panel A, positive price adjustments and negative price adjustments are both associated with the filing of more amendments. Further, more of these amendments tend to be late amendments when there is large positive or negative price adjustment. These findings are consistent with agents updating disclosed information in response to information revealed during the bookbuilding process. The number of days in registration is negatively related to positive price adjustment as well as negative price adjustment. This means that the number of days in registration increases as the price adjustment declines and IPOs with large price adjustments have fewer days in registration than those with negative price adjustments. Although not displayed, IPOs below the file range spend an average of 108 days in registration compared to only 86 days for IPOs above the file range.

Panel B presents the determinants of changes in the document allocation for the prospectus as a whole and the four subsections. The only section whose size is significantly affected by information revealed in the offering process is the Prospectus Summary, and a change in its relative size only occurs when negative information is revealed. Otherwise, there is no significant increase in the size of either the document or the remaining three subsections. Panel C presents evidence of the impact of ΔP on the normalized distances. We find little evidence that positive price adjustments affect the information content in the amendments to the initial prospectus. Only for the Uses of Proceeds section do we see a marginally significant relationship between ΔP+ and the normalized distance. In contrast, there is a highly significant and negative relationship between ΔP- and the normalized distances of the prospectus as a whole and all four subsections. Thus, revisions to the offering document are significant only when bad information is revealed during the bookbuilding process.

These findings are consistent with the incentive to reveal negative information as a hedge against liability but to withhold positive information from public disclosure for proprietary or strategic reasons. Further, the issuing firm and the underwriter are unlikely to face enhanced liability for withholding potentially valuable good information because, upon the market learning the information, the price of the shares will increase. Since shareholder damages are limited to declines in value, there is little cost, from a liability perspective, from not disclosing good information learned

during the offering process. In contrast, withholding bad information learned during the offering process exposes the issuing firm to a greater risk that a bad outcome will occur and the firm's share price in the aftermarket will decline below the offer price. In order to ensure that the issuing firm protects itself from liability for material omissions, negative information learned from bookbuilding is revealed in amendments to the prospectus.[18]

B The Effect of Changes in the Prospectus on Initial and Post-IPO Returns

This section examines whether changes to the prospectus during the offering period affect initial and post-IPO returns. Table X presents a regression analysis with both initial returns and one year post-offering returns as the dependent variables. Control variables are similar to those used in the previous analyses of initial returns and post-offering returns.

We find that changes in the size of the total document reduces initial returns but changes in the individual sections do not affect the size of the first day return. The finding for the total document may indicate that potentially valuable information is being revealed that is not captured in the individual sections. The increase in the size of the total prospectus appears to reduce information uncertainty and the associated underpricing.

Although the change in the size of the sections is uninformative, the normalized distances provide valuable information that is relevant to the market pricing of IPOs. For all subsections except Use of Proceeds, the change in the content of the document has a statistically significant and negative impact on initial returns.[19] Therefore, the results of this section indicate that there are meaningful disclosures associated with revisions in the offering prospectus that translates into lower initial returns. This is consistent with the hypothesis that greater disclosure lowers ex ante uncertainty and

[18]Note that the SEC may also require the issuing firm to provide additional information regarding the effect of a lower than expected offer amount.

[19]One might infer, however, that the reduction in initial return is limited to IPOs that have negative price adjustments. However, further analysis (not presented) indicates that the decline in initial return associated with greater change in the document content occurs even in offers that have upward price adjustments.

attendant information asymmetry.

Panel B of Table X examines whether changes in the prospectus are related to post-IPO abnormal returns. We find that neither the change in the relative section sizes, nor the normalized distances, has any effect on long-term abnormal returns.

V Conclusion

We employ a new methodology to examine the information contained in the initial offering prospectus text, and find that simple measures predict changes in offer prices, initial returns, and subsequent aftermarket pricing. Our results indicate that the prospectus as a whole, and the size of four key sections play important, and sometimes distinct, roles. We show that key differences in how document sections interact can be explained by the incentives of the likely author of each section. To explain these differing incentives, we conduct tests of the likely authorship of certain sections.

Our findings suggest that that the underwriter is the most dominant author of the Prospectus Summary and underwriter driven disclosure is consistent with traditional theories of disclosure. Greater disclosure in this section of the document reduces the degree of information asymmetry between the issuing firm and potential investors and results in smaller partial adjustment and lower initial returns. These findings suggest that underwriter disclosure improves the efficiency of IPO prices.

We find that managers are the most likely author of the MD&A and that greater management-driven disclosure is associated with a higher the final offer price and superior post-IPO abnormal returns. Because information in MD&A is priced in the final offer price, but not in the initial filing estimate, our results suggest that management actively participates in the bookbuilding process. The superior subsequent post-IPO performance indicates that management's contributions are credible. These findings motivate extensions to classical book building theory which posits that investors are the only information providers that account for management's participation.

Consistent with prior studies, we find that the Risk Factors section is positively

related to offer price changes during bookbuilding and subsequent initial returns. Our results indicate that a larger Risk Factors section reveals that the IPO firm is, in fact, riskier, and that firms may over-disclose bad information as a hedge against potential future liability. We do not find that the size of the Use of Proceeds section plays an important role.

Finally, we document that significant changes in the content of the prospectus arise only when bookbuilding reveals negative information (as measured by a decline in the offer price relative to the initial estimate). This is consistent with issuers and underwriters having incentives to conceal positive information when it has proprietary value, and disclose only information that has the greatest impact on reducing legal liability (i.e. negative information). These findings motivate extensions to classical disclosure theory that account for incentives to under-disclose information which has proprietary value.

Overall, our findings provide a new explanation of the partial adjustment phenomenon. Positive information may be withheld from investors to preserve proprietary advantages, which leaves the issuer and underwriter especially prone to litigation risk. Partial adjustment then arises because underwriters set the IPO price low in order to mitigate legal damages and hence, insure against this increased litigation risk and protect valuable reputational capital. Thus, our argument suggests that all offers with positive information revealed during bookbuilding will experience both large partial adjustment and underpricing due to the increased legal liability that comes with withholding proprietary information. Further research along these dimensions might also explain why some underwriters persistently underprice more than others.

Appendix 1

Our algorithm to read each prospectus and prospectus amendment is written in a combination of PERL and APL, and a flow chart is displayed in Figure 1. Once a document is downloaded and paginated, our algorithm's next step is to purge the document of attachments, headers, and exhibits so that we can focus on the prospectus itself. This achieved using a three prong approach that ensures a very high degree of accuracy: (1) we use the pagination implied by the Table of Contents to identify the beginning and end of the document, (2) we examine the placement of the "additional information" statement and the placement of accounting statements (exhibits) to confirm accuracy,[20] and (3) we hand check the algorithm's accuracy for most documents and include exception handling where necessary. We store the text of the prospectus in a character vector, which we define as $chars_{tot}$.

Our next step is to use the pagination implied by the Table of Contents to identify the beginning and end of each of the four key sections we seek to examine: the Prospectus Summary, the Risk Factors section, the Use of Proceeds section, and Management's Discussion and Analysis (MD&A). We store the text from the each of these four sections in separate character vectors, which we define as $chars_{ps}$, $chars_{rf}$, $chars_{use}$, and $chars_{mda}$, respectively.

[20]The overwhelming majority of prospectuses filed in our sample have a statement indicating where investors can find additional information toward the end of the prospectus document.

Appendix 2

This Appendix explains how we compute the "document similarity" and the "document distance" between two documents i and j. This same procedure can be applied to document sections, in which case the result would be the "section similarity" or "section distance". We first take the text in each document (or document section) and construct a numerical vector summarizing the counts of its English Language word roots. This vector has a number of elements equal to the number of word roots, and one element is the number of times the given word root appears in the document. Word roots are identified by Webster.com, and we use a web crawling algorithm to build a database of the unique word roots that correspond to all English Language words that appear in the universe of all IPO prospectuses. For example, the words display, displayed, and display all have the common word root "display".[21] We exclude articles and conjunctions from these counts because they are not informative regarding content. For a given section, whose character vectors we denoted as $chars_{x,i}$ (x can be either *tot, ps, rf, use, or mda*), we now have a corresponding numerical vector $P_{x,i}$, with the corresponding frequency of each possible word root in the given section of the given document. Because it is a vector of relative frequencies, each element of $P_{x,i}$ is a non-negative integer.

We next define the normalized frequency vector $V_{x,i}$, which normalizes the vector $P_{x,i}$ to have unit length.

$$V_{x,i} \;=\; \frac{P_{x,i}}{\sqrt{P_{x,i} \;\cdot\; P_{x,i}}} \tag{3}$$

To measure the degree of similarity of documents i and j, we simply take the dot product of the two normalized frequencies, a quantity we define as "document similarity". We utilize this measure in section III.B to identify the likely authorship of document sections.

$$Document\ Similarity_{x,i,j} \;=\; (V_{x,i} \;\cdot\; V_{x,j}) \tag{4}$$

To measure the severity of revision from document i to document j, we simply take one minus the dot product of the two normalized frequencies, a quantity we

[21] Methodologically, we first create a vector of all word counts in the given section of the document, and we then replace each word with its word root. We then tabulate the frequency vector for the given document section based on the total counts of each word root.

define as "document distance".

$$Document\ Distance_{x,i,j} \ = \ 1 \ - \ (V_{x,i} \ \cdot \ V_{x,j}) \tag{5}$$

We utilize this measure in section IV to identify the severity of revisions to the initial prospectus for each IPO.

Because all normalized vectors $V_{x,i}$ have length one, document distance and document similarity both have the nice property of being bounded in the interval (0,1). Intuitively, the distance between two documents is zero if they are the same, and can never exceed one if they are entirely different.

Appendix 3

This appendix explains how we assess the severity each IPO's prospectus revisions from the initial prospectus to the final amendment filed with the SEC. This calculation can be done over the entire document, or for individual document sections. Consider an IPO with a total of I documents filed (one initial prospectus and $I-1$ amendments). Let i denote a given document form 1,...,I. We first compute the document distance between each pair of documents in a series $Distance_{i,i+1}, \forall i = 1, ..., I-1$. This procedure is described in Appendix 2. For a sequence of I revisions, we then define the total "normalized document distance" from the initial prospectus to the final revision as the normalized sum of distances:

$$Total\ Normalized\ Distance = \frac{\sum_{i=1}^{I-1} Distance_{i,i+1}}{Max[Distance_{1,2}, ..., Distance_{I-1,I}]} \tag{6}$$

When there is only one document in a series (211 IPOs, see Table VII), we assign the Total Normalized Distance a value of zero to reflect the fact that no revisions were made. The normalization mitigates the impact of extreme observations, and also mitigates the impact of the first revision following the initial prospectus. We now summarize the normalized distance variables we use in our regressions.

Δ*normdist tot*: The total normalized distance from the initial prospectus to the final revision for the entire prospectus document. This is computed in three steps: (1) compute $\{V_{tot,1}, ..., V_{tot,1}\}$ from the text vectors $\{chars_{tot,1}, ..., chars_{tot,1}\}$. (2) Compute a time series of $I-1$ distances from these vectors using equation (5). (3) The variable Δnormdist tot is then the resulting expression from equation (6).

Δ*normdist ps*: The total normalized distance from the initial prospectus to the final revision for the Prospectus Summary section. This is computed in a parallel fashion as Δnormdist tot based on the starting character vectors $\{chars_{ps,1}, ..., chars_{ps,1}\}$.

Δ*normdist rf*: The total normalized distance from the initial prospectus to the final revision for the Risk Factors section. This is computed in a parallel fashion as Δnormdist tot based on the starting character vectors $\{chars_{rf,1}, ..., chars_{rf,1}\}$.

Δ*normdist use*: The total normalized distance from the initial prospectus to the final revision for the Use of Proceeds section. This is computed in a parallel fashion as Δnormdist tot based on the starting character vectors $\{chars_{use,1}$, ..., $chars_{use,1}\}$.

Δ*normdist mda*: The total normalized distance from the initial prospectus to the final revision for MD&A. This is computed in a parallel fashion as Δnormdist tot based on the starting character vectors $\{chars_{mda,1}$, ..., $chars_{mda,1}\}$.

References

Arnold, Tom, Raymond P.H. Fishe, and David North, 2006, The effects of "Risk-Factor" disclosure on the pricing of IPOs and long run returns, University of Richmond Working Paper.

Barry, C., C. Muscarella, J. Peavy, and M. Vetsuypens, 1990, The role of venture capital in the creation of public companies, *Journal of Financial Economics* 27, 447–471.

Beatty, Randolf, and Jay Ritter, 1986, Investment banking, reputation and the underpricing of initial public offerings, *Journal of Financial Economics* 15, 213–232.

Beatty, Randolf, and Ivo Welch, 1996, Issuer expenses and legal liability in initial public offerings, *Journal of Law and Economics* 39, 545–602.

Benveniste, Lawrence, and Paul Spindt, 1989, How investment bankers determine the offer price and allocation of new issues, *Journal of Financial Economics* 24, 343–362.

Bhattacharya, Sudipto, and Gabriella Chiesa, 1995, Proprietary information, financial intermediation, and research incentives, *Journal of Financial Intermediation* 4, 328–357.

Bryan, Stephen H., 1997, Incremental information content of required disclosures contained in management discussion and analysis, *Accounting Review* 72, 285–301.

Cook, Douglas, Robert Kieschnick, and Robert Van Ness, 2006, On the marketing of ipos, *Journal of Financial Economics* 82, 35–61.

Core, John E., 2001, A review of the empirical disclosure literature: A discussion, *Journal of Accounting and Economics* 31, 441–456.

Darrough, Masako N., and Neal M. Stoughon, 1990, Financial disclosure policy in an entry game, *Journal of Accounting and Economics* 12, 219–243.

Diamond, Douglas, and Robert E. Verrecchia, 1991, Disclosure, liquidity and the cost of capital, *The Journal of Finance* 66, 1325–1355.

Dye, Ronald A., 2001, An evaluation of "essays on disclosure" and the disclosure literature in accounting, *Journal of Accounting and Economics* 32, 181–235.

Easley, David, and Maureen O'Hara, 2004, Information and the cost of capital, *The Journal of Finance* 59, 97–180.

Field, Laura, Michelle Lowry, and Susan Shu, 2005, Does disclosure deter or trigger litigation?, *Journal of Financial Economics* 39, 487–507.

Field, Laura Casares, and Jonathan Karpoff, 2002, Takeover defenses of IPO firms, *The Journal of Finance* 57, 1857–89.

Guo, Re-Jin, Baruch Lev, and Nan Zhou, 2004, Competitive costs of disclosure by Biotech IPOs, *Journal of Accounting Research* 42, 319–364.

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

Healy, Paul M., and Krishna G. Palepu, 2001, Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature, *Journal of Accounting and Economics* 31, 405–440.

Hoberg, Gerard, 2007, The underwriter persistence phenomenon, *The Journal of Finance* 62, 1169–1206.

Hughes, Patricia, and Anjan Thakor, 1992, Litigation risk, intermediation, and the underpricing of initial public offerings, *Review of Financial Studies* 5, 709–42.

Kim, M., and Jay Ritter, 1999, Valuing IPOs, *Journal of Financial Economics* 53, 409–437.

Leone, Andrew J., Steve Rock, and Micheal Willenborg, 2007, Disclosure of intended use of proceeds and underpricing of initial public offerngs, *Journal of Accounting Research* forthcoming.

Liu, Laura Xiaolei, Ann E. Sherman, and Yong Zhang, 2007, Media coverage and ipo pricing, Hong Kong University and University of Notre Dame Working Paper.

Ljungqvist, Alexander, and William Wilhelm, 2003, IPO pricing in the Dot-com bubble, *The Journal of Finance* 58, 723–752.

Logue, D, 1973, On the pricing of unseasoned equity issues 1965-69, *Journal of Financial and Quantitative Analysis* 8, 91–103.

Loughran, Tim, and Jay Ritter, 2002, Why don't issuers get upset about leaving money on the table in IPOs, *Review of Financial Studies* 15, 413–433.

Loughran, Tim, and Jay Ritter, 2004, Why has IPO underpricing changed over time?, *Financial Management* 33, 5–37.

Lowry, Michelle, and William Schwert, 2002, IPO market cycles: Bubbles or sequential learning, *The Journal of Finance* 57, 1171–1200.

Lowry, Michelle, and Susan Shu, 2002, Litigation risk and IPO underpricing, *Journal of Financial Economics* 65, 309–335.

Maksimovic, Vojislav, and Pegaret Pichler, 2001, Technological innovation and initial public offerings, *Review of Financial Studies* 14, 459–494.

Megginson, William, and Kathleen Weiss, 1991, Venture capitalist certification in initial public offerings, *The Journal of Finance* 46, 879–903.

Pava, M.L., and M. Epstein, 1993, Md&a as an investment tool: User beware!, *Journal of Accountancy* March, 51–53.

Schrand, Catherine, and Robert E. Verrecchia, 2005, Information disclosure and adverse selection explanations for IPO underpricing, Wharton Working Paper.

Tinic, Seha, 1988, Anatomy of initial public offerings of common stock, *The Journal of Finance* 43, 789–822.

Verrecchia, Robert E., 2001, Essays on disclosure, *Journal of Accounting and Economics* 32, 97–180.

Table I: Summary Statistics: IPO Pricing and Characteristics

Summary statistics are reported for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. **Initial Return** (IR), is the actual return from the IPO offer price to the first CRSP reported closing price. **ΔP** is the actual return from the filing date midpoint to the IPO offer price, and **ΔP+** and **ΔP-** are its positive and negative truncated components. The **1-year post IPO return** is the abnormal return from the closing price on the first day of public trading until the IPO's one year anniversary date. This is the intercept of a regression of twelve monthly stock returns on the three Fama-French factors. The **IPO size at filing** is the original filing amount. The **IPO size at offering** is the final Issue Proceeds. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UWdshare** is the lead underwriter's dollar market share in the past calendar year. We construct two measures of **LAWdshare**. Our first is the underwriting firm's legal counsel's dollar market share in the past calendar year. Our second is the issuer firm's legal counsel's dollar market share in the past calendar year. We construct two measures of **Mkt30**. Our first is the NASDAQ return for the 30 trading days preceding the filing date. Our second is the NASDAQ return for the 30 trading days preceding the issue date.

Variable	Description	Mean	Std. Dev.	Minimum	Median	Maximum
		Panel A: Price Variables				
IR	Initial Return	0.332	0.649	-0.399	0.122	6.267
ΔP	Price Adjustment	0.035	0.271	-0.984	0.000	2.200
$\Delta P+$	$Max[0, \Delta P]$	0.109	0.205	0.000	0.000	2.200
$\Delta P-$	$Min[0, \Delta P]$	-0.075	0.123	-0.984	0.000	0.000
Abret12	1-year post IPO return	-0.004	0.116	-1.084	-0.006	1.088
		Panel B: IPO Characteristics				
Ipsiz	IPO size at filing	186.976	1157.50	2.750	56.175	46926.1
Offsiz	IPO size at offering	115.643	333.430	2.320	55.500	8680.00
firmage	Age of firm at offering	13.752	20.107	0.000	7.000	165.000
TECH	Tech dummy variable	0.435	0.496	0.000	0.000	1.000
VC	VC dummy variable	0.470	0.499	0.000	0.000	1.000
UWdshare	Lead UW $ Market Share	0.029	0.026	0.000	0.023	0.147
mLAWdshare	UW Counsel $ Market Share	0.012	0.022	0.000	0.004	0.177
iLAWdshare	Iss. Counsel $ Market Share	0.023	0.033	0.000	0.010	0.216
naswin30f	Nasdaq return 30 days before filing	0.052	0.090	-0.260	0.058	0.350
naswin30	Nasdaq return 30 days before offer	0.050	0.084	-0.265	0.053	0.359

Table II: Summary Statistics: Initial Prospectus Content

Summary statistics are reported for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. We report the **Total characters in document**, which is the total number of characters in the prospectus document (after removing headers and footers from the filing submitted to SEC Edgar), and the percentage of characters allocated to each of four key sections. ps_{pct} is the percentage of prospectus characters that appear in the prospectus summary. rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages for the risk factors section, the use of proceeds section, and the management's discussion and analysis section. In Panels B and C we report Pearson Correlation Coefficients for each of the document sections. In Panel B, we present correlations based on each section's raw total number of characters. In Panel C, we present correlations for each document section's size as a percentage of the total document. * indicates statistical significance at the 10% level or better.

Variable	Description	Mean	Std. Dev.	Minimum	Median	Maximum
		Panel A: Summary Statistics				
totchars	Total characters in document	215633	71715.6	68085.0	201500	820082
ps_{pct}	Prospectus Summary % of document	0.061	0.026	0.007	0.055	0.430
rf_{pct}	Risk Factors % of document	0.181	0.052	0.036	0.182	0.594
use_{pct}	Use of Proceeds % of document	0.009	0.017	0.000	0.008	0.754
mda_{pct}	Mgmt Discussion % of document	0.132	0.050	0.001	0.129	0.699

	Variable	Total Document Characters	Prospectus Summary	Risk Factors	Use of Proceeds
		Panel B: Correlation Coefficients (Raw Character Sizes)			
(1)	Prospectus Summary Total Characters	0.650*			
(2)	Risk Factors Total Characters	0.590*	0.147*		
(3)	Use of Proceeds Total Characters	0.007	0.018	-0.021	
(4)	Mgmt Discussion Total Characters	0.616*	0.412*	0.325*	-0.031
		Panel C: Correlation Coefficients (Relative Section Sizes)			
(1)	Prospectus Summary % of document	0.057*			
(2)	Risk Factors % of document	-0.185*	-0.401*		
(3)	Use of Proceeds % of document	-0.130*	0.033	-0.012	
(4)	Mgmt Discussion % of document	0.051*	0.032	-0.145*	-0.060*

Table III: Determinants of Initial Prospectus

The dependent variable is the **Log total characters in document** (Row one), which is the total number of characters in the prospectus document after removing headers and footers from the filing submitted to SEC Edgar. The dependent variable in Rows two to five is the percentage of characters allocated to each of the following four key sections, respectively. ps_{pct} is the percentage of prospectus characters that appear in the prospectus summary. rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages for the risk factors section, the use of proceeds section, and the management's discussion and analysis section. We now list our explanatory variables. **Log dollars filed** is the natural logarithm of the initial filing amount. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UW market share** is the lead underwriter's dollar market share in the past calendar year. We construct two measures of **LAW market share**. Our first is the underwriting firm's legal counsel's dollar market share in the past calendar year. Our second is the issuer firm's legal counsel's dollar market share in the past calendar year. **Mkt30** is the NASDAQ return for the 30 trading days preceding the filing date. **Year and industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries.

Row	Dependent Variable	Log Dollars Filed	Log Firm Age	VC Dummy	UW $ Market Share	UW Law $ Market Share	Iss Law $ Market Share	Tech Dummy	Pre-file Nasdaq Return	Year+Ind Fixed Effects	R^2
(1)	Log Document Characters	0.078	0.014	0.026	0.786	0.521	0.519	0.034	0.098	Yes	0.477
		(13.72)	(2.62)	(2.38)	(3.32)	(2.30)	(3.33)	(2.27)	(1.81)		
(2)	% Prospectus Summary	0.002	0.001	-0.010	0.069	-0.011	-0.001	-0.009	-0.008	Yes	0.354
		(3.48)	(1.77)	(-9.01)	(2.91)	(-0.48)	(-0.08)	(-6.23)	(-1.47)		
(3)	% Risk Factors	-0.007	-0.011	0.008	-0.208	0.115	-0.064	0.021	-0.011	Yes	0.355
		(-6.58)	(-10.22)	(3.89)	(-4.39)	(2.52)	(-2.06)	(7.14)	(-0.98)		
(4)	% Use of Proceeds	-0.001	-0.001	-0.002	-0.028	-0.012	-0.018	-0.000	-0.005	Yes	0.118
		(-1.58)	(-2.48)	(-2.12)	(-1.48)	(-0.65)	(-1.41)	(-0.06)	(-1.23)		
(5)	% Management's Discussion	0.003	0.007	0.001	-0.017	-0.038	-0.019	0.009	0.003	Yes	0.254
		(2.77)	(6.58)	(0.44)	(-0.35)	(-0.80)	(-0.58)	(2.76)	(0.23)		
(6)	% Above Four Sections	-0.003	-0.004	-0.002	-0.185	0.054	-0.102	0.021	-0.021	Yes	0.194
		(-1.80)	(-2.40)	(-0.69)	(-2.73)	(0.83)	(-2.29)	(4.84)	(-1.37)		

Table IV: Determinants of Document Similarity and Authorship

The dependent variable is the **Document Distance** of two initial IPO prospectuses. One observation is one pair of IPOs i and j, and we include all unique possible IPO pairs as observations (we exclude pairs in which $i = j$). For our sample of 2044 IPOs, a maximum of $\frac{2044^2-2044}{2}$ unique pairs exist, and hence a maximum of 2,087,946 observations appear in any regression (fewer appear in some specifications as some sections are missing for a small number of IPOs). To ensure T-statistics remain unbiased given the repeated use of each document, we report T-statistics that are adjusted for clustering by IPO. To compute document distance, we first construct a vector with a length equal to the number of words in the union of both documents. For documents i and j, we populate this vector with the document's frequency of each word, and we then normalize each vector to have length one. Document distance is the dot product of these two vectors for document i and j. The independent variables measure how similar the characteristics of IPO i and j are. The first three variables are dummy variables identifying whether IPOs i and j had the same lead underwriter, the same manager's counsel, and the same issuer's counsel. The next four dummy variables are one if IPO i and j reside in the same one digit to four digit SIC code, respectively. We also include a dummy variable identifying whether IPO i and j were issued in the same year, and a dummy indicating whether both are in Tech oriented as identified in Loughran and Ritter (2004). Finally, we have four variables measuring how different IPO i and j characteristics are. Each is equal to the absolute value of difference in characteristics for IPO i and j, and we examine log firm age, IPO year, log filing size, and the underwriting spread. Larger quantities of each indicates that i and j differ regarding the given characteristic.

Row	Same Lead UW	Same UW Counsel	Same Issuer Counsel	Same SIC-1 Code	Same SIC-2 Code	Same SIC-3 Code	Same SIC-4 Code	Same IPO Year	Both Tech IPOs	Absol. Age Diff.	Absol. Year Diff.	Absol Log Size Diff.	Absol Spread Diff.	R^2	Obs
							Panel A: Entire Document								
(1)	0.063	0.064	0.032											0.010	2,087,946
	(11.51)	(27.80)	(17.55)												
(2)	0.045	0.053	0.025	0.005	0.046	0.044	0.032		0.018					0.106	2,087,946
	(8.82)	(23.85)	(15.79)	(2.16)	(12.63)	(13.42)	(14.06)		(10.79)						
(3)	0.036	0.041	0.019	0.006	0.045	-0.004	0.034	0.055	0.008	-0.012	-0.003	-0.000	-0.006	0.162	2,087,946
	(7.14)	(20.61)	(13.12)	(2.99)	(12.74)	(-1.13)	(15.23)	(23.10)	(4.50)	(-11.23)	(-4.39)	(-1.22)	(-5.34)		
							Panel B: Prospectus Summary Section								
(4)	0.085	0.040	0.024											0.008	2,087,946
	(18.05)	(18.41)	(15.96)												
(5)	0.069	0.032	0.019	-0.001	0.037	0.034	0.023		0.020					0.071	2,087,946
	(15.41)	(15.30)	(13.64)	(-0.76)	(11.89)	(10.77)	(9.68)		(12.25)						
(6)	0.059	0.024	0.015	-0.000	0.036	0.004	0.024	0.034	0.008	-0.007	-0.003	0.000	-0.005	0.100	2,087,946
	(12.86)	(12.45)	(11.21)	(-0.11)	(11.82)	(1.16)	(10.28)	(14.96)	(4.42)	(-7.38)	(-6.27)	(4.00)	(-5.67)		
							Panel C: Risk Factors Section								
(7)	0.063	0.041	0.027											0.005	2,081,820
	(9.95)	(17.52)	(13.66)												
(8)	0.034	0.031	0.021	0.007	0.042	0.030	0.029		0.049					0.079	2,081,820
	(5.98)	(13.69)	(11.59)	(3.43)	(12.07)	(9.26)	(11.17)		(24.45)						
(9)	0.020	0.020	0.015	0.008	0.038	-0.015	0.031	0.051	0.020	-0.008	-0.008	-0.000	-0.004	0.136	2,081,820
	(3.43)	(9.86)	(9.36)	(4.45)	(11.08)	(-4.12)	(12.20)	(18.93)	(9.06)	(-6.21)	(-11.58)	(-3.47)	(-3.47)		
							Panel D: Use of Proceeds Section								
(10)	0.055	0.027	0.009											0.001	2,081,820
	(6.97)	(8.43)	(4.19)												
(11)	0.042	0.024	0.007	-0.010	0.031	-0.004	0.006		0.027					0.008	2,081,820
	(5.31)	(7.65)	(3.25)	(-4.59)	(6.35)	(-0.76)	(1.73)		(9.70)						
(12)	0.026	0.019	0.004	-0.009	0.028	-0.021	0.006	0.018	0.001	-0.008	-0.008	0.000	-0.000	0.026	2,081,820
	(3.30)	(5.99)	(1.92)	(-4.13)	(5.80)	(-4.18)	(1.98)	(4.94)	(0.51)	(-5.23)	(-8.35)	(3.48)	(-0.29)		
							Panel E: Management's Discussion and Analysis Section								
(13)	0.037	0.010	0.003											0.001	2,041,210
	(6.41)	(3.64)	(1.89)												
(14)	0.026	0.008	0.002	-0.001	0.011	0.005	-0.007		0.026					0.007	2,041,210
	(4.49)	(3.03)	(1.25)	(-0.67)	(3.08)	(1.31)	(-3.08)		(11.85)						
(15)	0.017	0.005	0.000	-0.000	0.009	-0.003	-0.007	0.008	0.014	-0.003	-0.003	0.000	-0.002	0.014	2,041,210
	(2.87)	(1.95)	(0.26)	(-0.25)	(2.60)	(-0.84)	(-2.93)	(2.79)	(6.45)	(-2.57)	(-4.59)	(8.41)	(-2.41)		

Table V: Predicting Price Adjustments Using Initial Prospectus (ΔP)

OLS regressions with yearly fixed effects are presented for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. The dependent variable is the change in price **ΔP** (Panel A) from the filing date midpoint to the IPO offer price, and **ΔP+** and **ΔP-** (Panels B and C) are its positive and negative truncated components. The **IPO size at filing** is the original filing amount. We now list the explanatory variables. In the first row in each panel, the "% of document field" refers to the explanatory variable **Total characters in document**, which is the natural logarithm of the total number of characters in the prospectus document (after removing headers and footers from the filing submitted to SEC Edgar). In the remaining four rows in each panel, this column refers to the percentage of characters allocated to each of four key sections. The corresponding section is listed in the "Prospectus Section" column. The **Log dollars filed** is the natural logarithm of the initial filing amount. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UW market share** is the lead underwriter's dollar market share in the past calendar year. We construct two measures of **LAW market share**. Our first is the underwriting firm's legal counsel's dollar market share in the past calendar year. Our second is the issuer firm's legal counsel's dollar market share in the past calendar year. **Mkt30** is the NASDAQ return for the 30 trading days preceding the filing date. **Year and industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries.

Row	Prospectus Section	% of Document	Log Firm Age	UW $ Market Share	UW Law $ Market Share	Iss Law $ Market Share	TECH Dummy	VC Dummy	Pre File Nasdaq Return	Log Dollars Filed	Year+Ind Fixed Effects	R^2
					Panel A: ΔP							
(1)	Whole Document	0.036	-0.013	2.262	0.466	-0.256	0.032	0.027	0.117	-0.020	Yes	0.191
		(1.34)	(-2.14)	(8.06)	(1.74)	(-1.38)	(1.81)	(2.10)	(1.84)	(-2.89)		
(2)	Prospectus Summmary	-0.575	-0.012	2.329	0.479	-0.238	0.028	0.022	0.116	-0.016	Yes	0.192
		(-2.17)	(-1.98)	(8.31)	(1.79)	(-1.29)	(1.56)	(1.71)	(1.82)	(-2.44)		
(3)	Risk Factors	0.344	-0.009	2.362	0.445	-0.215	0.026	0.025	0.125	-0.015	Yes	0.193
		(2.60)	(-1.43)	(8.41)	(1.66)	(-1.17)	(1.45)	(1.95)	(1.95)	(-2.21)		
(4)	Use of Proceeds	-0.382	-0.013	2.279	0.480	-0.244	0.033	0.027	0.119	-0.018	Yes	0.191
		(-1.14)	(-2.13)	(8.14)	(1.79)	(-1.32)	(1.88)	(2.12)	(1.86)	(-2.65)		
(5)	Management's Discussion	0.390	-0.016	2.297	0.500	-0.230	0.030	0.028	0.120	-0.019	Yes	0.194
		(3.09)	(-2.50)	(8.22)	(1.87)	(-1.25)	(1.69)	(2.15)	(1.88)	(-2.81)		
					Panel B: ΔP+							
(6)	Whole Document	0.044	-0.016	3.054	0.547	-0.356	0.035	0.069	0.210	-0.002	Yes	0.204
		(1.15)	(-1.79)	(7.87)	(1.52)	(-1.40)	(1.44)	(3.84)	(2.39)	(-0.21)		
(7)	Prospectus Summmary	-1.186	-0.014	3.148	0.554	-0.327	0.025	0.059	0.207	0.003	Yes	0.210
		(-3.16)	(-1.52)	(8.16)	(1.54)	(-1.30)	(1.03)	(3.26)	(2.36)	(0.35)		
(8)	Risk Factors	0.456	-0.010	3.173	0.520	-0.305	0.027	0.066	0.220	0.004	Yes	0.210
		(2.43)	(-1.10)	(8.20)	(1.44)	(-1.21)	(1.11)	(3.67)	(2.51)	(0.45)		
(9)	Use of Proceeds	-5.005	-0.016	2.948	0.525	-0.387	0.035	0.063	0.204	-0.005	Yes	0.204
		(-2.66)	(-1.83)	(7.58)	(1.46)	(-1.53)	(1.45)	(3.48)	(2.33)	(-0.52)		
(10)	Management's Discussion	0.565	-0.019	3.108	0.599	-0.319	0.031	0.069	0.217	-0.000	Yes	0.205
		(3.22)	(-2.16)	(8.05)	(1.66)	(-1.26)	(1.30)	(3.82)	(2.47)	(-0.05)		
					Panel C: ΔP-							
(11)	Whole Document	0.031	-0.002	2.562	0.439	-0.129	0.012	-0.009	-0.017	-0.040	Yes	0.106
		(1.01)	(-0.28)	(7.51)	(1.32)	(-0.59)	(0.59)	(-0.62)	(-0.23)	(-4.79)		
(12)	Prospectus Summmary	0.364	-0.002	2.550	0.463	-0.112	0.017	-0.005	-0.011	-0.038	Yes	0.107
		(1.17)	(-0.27)	(7.47)	(1.39)	(-0.51)	(0.79)	(-0.32)	(-0.15)	(-4.78)		
(13)	Risk Factors	0.059	-0.001	2.598	0.445	-0.110	0.012	-0.009	-0.014	-0.037	Yes	0.106
		(0.39)	(-0.15)	(7.58)	(1.34)	(-0.50)	(0.57)	(-0.60)	(-0.19)	(-4.62)		
(14)	Use of Proceeds	-0.390	-0.002	2.571	0.447	-0.121	0.013	-0.009	-0.018	-0.038	Yes	0.106
		(-1.16)	(-0.30)	(7.55)	(1.34)	(-0.55)	(0.64)	(-0.61)	(-0.24)	(-4.72)		
(15)	Management's Discussion	0.543	-0.005	2.598	0.482	-0.108	0.008	-0.008	-0.019	-0.040	Yes	0.111
		(3.50)	(-0.76)	(7.64)	(1.45)	(-0.50)	(0.40)	(-0.53)	(-0.26)	(-4.96)		

Table VI: Predicting Initial Returns and Post-IPO Returns Using Initial Prospectus

OLS regressions with yearly fixed effects are presented for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. The dependent variable is the **Initial Return** (Panel A), which is the actual return from the IPO offer price to the first CRSP reported closing price. In Panel B, the dependent variable is the **1-year post IPO return**. This is the abnormal return from the closing price on the first day of public trading until the IPO's one year anniversary date, and is the intercept of a regression of twelve monthly stock returns on the three Fama-French factors. We now list the explanatory variables. In the first row in each panel, the "% of document field" refers to the explanatory variable **Total characters in document**, which is the natural logarithm of the total number of characters in the prospectus document (after removing headers and footers from the filing submitted to SEC Edgar). In the remaining four rows in each panel, this column refers to the percentage of characters allocated to each of four key sections. The corresponding section is listed in the "Prospectus Section" column. The **Log dollars filed** is the natural logarithm of the initial filing amount. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UW market share** is the lead underwriter's dollar market share in the past calendar year. **Mkt30** is the NASDAQ return for the 30 trading days preceding the filing date. **Year and industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries.

Row	Prospectus Section	% of Document	Log Firm Age	UW $ Market Share	TECH Dummy	VC Dummy	Pre File Nasdaq Return	Log Dollars Filed	Year+Ind Fixed Effects	R^2	Obs
					Panel A: Initial Return						
(1)	Whole Document	0.081	-0.053	4.691	0.065	0.145	0.355	-0.049	Yes	0.239	2,044
		(1.31)	(-3.65)	(7.30)	(1.58)	(4.86)	(2.39)	(-3.04)			
(2)	Prospectus Summmary	-2.326	-0.049	4.925	0.046	0.124	0.345	-0.038	Yes	0.244	2,044
		(-3.78)	(-3.44)	(7.71)	(1.12)	(4.11)	(2.33)	(-2.48)			
(3)	Risk Factors	1.263	-0.038	5.041	0.041	0.136	0.376	-0.033	Yes	0.245	2,044
		(4.14)	(-2.58)	(7.87)	(0.99)	(4.57)	(2.54)	(-2.13)			
(4)	Use of Proceeds	-0.841	-0.052	4.740	0.068	0.145	0.359	-0.043	Yes	0.239	2,044
		(-1.09)	(-3.63)	(7.40)	(1.65)	(4.89)	(2.42)	(-2.82)			
(5)	Management's Discussion	0.276	-0.054	4.775	0.066	0.147	0.362	-0.043	Yes	0.239	2,044
		(0.94)	(-3.68)	(7.46)	(1.59)	(4.94)	(2.44)	(-2.82)			
					Panel B: One-Year Post-IPO Return						
(6)	Whole Document	-0.002	0.005	0.199	0.023	-0.014	0.026	-0.001	Yes	0.038	2,029
		(-0.13)	(1.56)	(1.53)	(2.80)	(-2.33)	(0.85)	(-0.42)			
(7)	Prospectus Summmary	0.019	0.005	0.196	0.023	-0.014	0.026	-0.002	Yes	0.038	2,029
		(0.15)	(1.54)	(1.51)	(2.79)	(-2.26)	(0.85)	(-0.49)			
(8)	Risk Factors	-0.141	0.003	0.168	0.026	-0.013	0.024	-0.003	Yes	0.040	2,029
		(-2.28)	(1.01)	(1.30)	(3.13)	(-2.12)	(0.80)	(-0.83)			
(9)	Use of Proceeds	0.021	0.005	0.198	0.023	-0.014	0.025	-0.001	Yes	0.038	2,029
		(0.14)	(1.56)	(1.53)	(2.80)	(-2.33)	(0.85)	(-0.47)			
(10)	Management's Discussion	0.119	0.004	0.202	0.022	-0.014	0.025	-0.002	Yes	0.040	2,029
		(2.01)	(1.24)	(1.56)	(2.67)	(-2.36)	(0.85)	(-0.61)			

Table VII: Summary of Prospectus Filing Patterns

The abbreviations PS, RF, USE, and MD&A represent Prospectus Summary, Risk Factors, Use of Proceeds, and Management's Discussion and Analysis, respectively. The table reports the average number of characters in each section of the prospectus, and each amendment to the initial prospectus, sorted in the order the amendments are received. The corresponding prospectus section is noted in the column name, where "PS" refers to "Prospectus Summary", "RF" refers to "Risk Factors","USE" refers to "Use of Proceeds", and "MDA" refers to "Management's Discussion and Analysis". For each section, we also report the total distance of the given amendment from the previous filing. Smaller numbers indicate that the given revision is relatively less extreme. Document distance is the dot product of the two normalized vectors, one for each document being compared. Each vector corresponds to the frequency of word roots, and each vector is normalized to have a length of unity. Document distance is bounded to be in the interval (0,1). We also report the total number of IPOs for which the given number of prospectuses are filed in the last column. For example, most IPOs do not experience more than four to five filings total.

Amend-mend	Total Number Char-acters	Total Dist from prev	PS Number Char-acters	PS Dist from prev	RF Number Char-acters	RF Dist from prev	USE Number Char-acters	USE Dist from prev	MD&A Number Char-acters	MD&A Dist from prev	Obs
Initial	215,633	0.000	13,208	0.000	38,336	0.000	1,825	0.000	28,688	0.000	2,044
2	226,158	0.036	14,028	0.284	38,805	0.095	2,114	1.069	31,878	0.255	1,833
3	232,648	0.019	13,926	0.157	40,126	0.051	2,133	0.582	33,241	0.106	1,863
4	239,717	0.011	13,952	0.095	41,318	0.028	2,128	0.336	34,687	0.061	1,589
5	248,078	0.009	13,932	0.064	42,741	0.020	2,017	0.235	36,768	0.039	1,111
6	260,961	0.006	14,495	0.043	45,697	0.014	2,066	0.202	39,070	0.033	673
7	265,221	0.006	14,075	0.079	47,358	0.018	2,033	0.287	40,568	0.036	369
8	274,725	0.003	14,307	0.041	49,513	0.009	1,930	0.175	43,739	0.022	171
9	289,225	0.004	14,392	0.037	52,936	0.008	2,226	0.094	47,130	0.025	89
10	299,972	0.002	15,074	0.027	55,212	0.005	1,875	0.066	52,194	0.014	40
11	311,185	0.002	15,759	0.031	56,605	0.005	3,105	0.013	52,452	0.007	22
12	329,594	0.011	19,987	0.144	55,481	0.021	2,716	0.727	67,823	0.063	9
13	279,352	0.073	16,382	0.367	45,270	0.196	4,696	0.101	39,133	0.109	3
14	252,851	0.006	14,568	0.072	46,090	0.006	2,737	0.217	33,588	0.027	4
15	186,581	0.000	12,183	0.000	30,943	0.000	5,213	-0.000	24,838	-0.000	1
16	307,599	0.003	22,452	0.013	51,430	0.017	1,007	0.192	74,549	0.012	1
17	307,789	0.000	22,643	0.008	51,433	0.001	1,005	0.020	74,544	0.001	1

Table VIII: Summary Statistics: Changes in Prospectus

Summary statistics are reported for IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. The number of days in registration is the number of days from the first Edgar filing until the IPO date. The number of amendments is the total number of filings, including the initial filing, but excluding the final prospectus. The number of late amendments is the number of these amendments occurring in the seven day window prior to the IPO date. We also report changes in the natural logarithm of **Total characters in document**, which is the total number of characters in the prospectus document (after removing headers and footers from the filing submitted to SEC Edgar). All reported differences are cumulative, and are based on changes from the initial filing to the final amendment. We also report changes in the percentage of characters allocated to each of four key sections. ps_{pct} is the percentage of prospectus characters that appear in the prospectus summary. rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages for the risk factors section, the use of proceeds section, and the management's discussion and analysis section. For each section, we also report the total distance of the given amendment from the previous filing. Smaller numbers indicate that the given revision is relatively less extreme. Document distance is the dot product of the two normalized vectors, one for each document being compared. Each vector corresponds to the frequency of word roots, and each vector is normalized to have a length of unity. Document distance is bounded to be in the interval (0,1).

Variable	Description	Mean	Std. Dev.	Minimum	Median	Maximum
		Panel A: Amendments and Registration				
daysreg	Days in registration	93.989	79.410	1	76	1016
numamend	Number of amendments	3.989	1.081	0	4	15
numlate7	Number of late amendments	1.035	0.797	0	1	5
		Panel B: Changes in Document Allocation				
Δtotchars	ΔTotal characters in document	0.079	0.101	-0.752	0.066	1.024
Δps_pct	ΔProspectus Summary % of document	-0.001	0.012	-0.134	-0.000	0.279
Δrf_pct	ΔRisk Factors % of document	-0.007	0.020	-0.143	-0.004	0.177
Δuse_pct	ΔUse of Proceeds % of document	0.001	0.005	-0.066	0.000	0.190
Δmda_pct	ΔMgmt Discussion % of document	0.009	0.035	-0.543	0.006	0.372
		Panel C: Changes in Character Distribution				
Δnormdist tot	Normalized change in Total Document distribution	0.111	0.086	0.000	0.099	0.577
Δnormdist ps	Normalized Δ in Prospectus Summary distribution	0.329	0.286	0.000	0.278	1.784
Δnormdist rf	Normalized Δ in Risk Factors distribution	0.294	0.269	0.000	0.222	1.489
Δnormdist use	Normalized Δ in Use of Proceeds distribution	0.250	0.323	0.000	0.099	1.946
Δnormdist mda	Normalized Δ in Management's Discussion distribution	0.264	0.269	0.000	0.176	1.397

Table IX: Determinants of Changes in the Prospectus

OLS regressions with yearly fixed effects are presented for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. The dependent variable is different in each row, and we now explain how each is constructed. The number of days in registration is the number of days from the first Edgar filing until the IPO date. The number of amendments is the total number of filings, including the initial filing, but excluding the final prospectus. The number of late amendments is the number of these amendments occurring in the seven day window prior to the IPO date. **Δ Document**, is the change in the natural logarithm of the total number of characters in the prospectus document. All change variables are cumulative, and are based on changes from the initial filing to the final amendment. We also include changes in the percentage of characters allocated to each of four key sections. ps_{pct} is the percentage of prospectus characters that appear in the prospectus summary. rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages for the risk factors section, the use of proceeds section, and the management's discussion and analysis section. For each section, we also report the total distance of the given amendment from the previous filing. Document distance is the dot product of the two normalized vectors, one for each document being compared. Each vector corresponds to the frequency of word roots normalized to have a length of unity. We now explain how the independent variables are constructed. The **Log dollars offered** is the natural logarithm of the filing amount. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UW Market Share** is the lead underwriter's dollar market share in the past calendar year. We construct two measures of **Law Market Share**. Our first is the underwriting firm's legal counsel's dollar market share in the past calendar year. Our second is the issuer firm's legal counsel's dollar market share in the past calendar year. **Pre Offer Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the offer date. **Year and industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries.

Row	Dependent Variable	ΔP+	ΔP−	Log Dollars Filed	Log Firm Age	VC Dummy	UW $ Market Share	UW Law Market Share	Iss Law Market Share	TECH Dummy	Pre Offer Nasdaq Return	Year+Ind Fixed Effects	R^2
						Panel A: Amendments and Registration							
(1)	Log Number of Amendments	0.207	-0.493	0.040	-0.016	0.014	0.163	0.279	0.649	-0.039	0.201	Yes	0.205
		(4.74)	(-7.14)	(4.38)	(-1.89)	(0.81)	(0.42)	(0.76)	(2.58)	(-1.58)	(2.01)		
(2)	Log Late Amendments	0.821	-0.359	0.076	-0.007	0.051	-1.572	-0.417	0.388	0.088	-0.306	Yes	0.134
		(8.36)	(-2.31)	(3.68)	(-0.35)	(1.30)	(-1.80)	(-0.51)	(0.69)	(1.59)	(-1.36)		
(3)	Log Days in Registration	-0.310	-0.931	-0.067	-0.013	-0.057	-1.980	0.126	-1.126	-0.010	0.674	Yes	0.108
		(-2.95)	(-5.60)	(-3.01)	(-0.63)	(-1.34)	(-2.12)	(0.14)	(-1.86)	(-0.17)	(2.79)		
						Panel B: Changes in Document Allocation							
(4)	Δ Document Characters	-0.007	-0.010	0.003	-0.004	-0.012	-0.015	-0.029	0.125	-0.026	-0.035	Yes	0.103
		(-0.58)	(-0.50)	(1.14)	(-1.79)	(-2.37)	(-0.13)	(-0.27)	(1.72)	(-3.68)	(-1.21)		
(5)	Δ Prospectus Summary	0.002	-0.006	-0.000	-0.001	0.001	-0.018	-0.010	-0.018	0.002	-0.001	Yes	0.083
		(1.49)	(-2.56)	(-0.32)	(-3.56)	(2.12)	(-1.26)	(-0.72)	(-1.94)	(2.27)	(-0.31)		
(6)	Δ Risk Factors	0.001	-0.002	0.001	0.001	0.001	0.034	0.025	0.013	-0.001	0.001	Yes	0.184
		(0.42)	(-0.64)	(1.29)	(1.24)	(0.67)	(1.57)	(1.24)	(0.95)	(-0.54)	(0.25)		
(7)	Δ Use of Proceeds	0.000	0.001	0.000	-0.000	-0.000	-0.009	-0.003	-0.005	-0.000	-0.002	Yes	0.097
		(0.73)	(1.45)	(0.28)	(-0.94)	(-0.29)	(-1.63)	(-0.62)	(-1.32)	(-0.40)	(-1.58)		
(8)	Δ Management Discussion	-0.003	-0.009	-0.002	-0.001	-0.003	-0.006	-0.015	-0.000	0.001	-0.008	Yes	0.059
		(-0.69)	(-1.21)	(-2.05)	(-0.98)	(-1.67)	(-0.14)	(-0.39)	(-0.00)	(0.39)	(-0.76)		
						Panel C: Changes in Document Distance							
(9)	Δ Document Distance	-0.005	-0.073	0.001	-0.006	-0.001	0.005	0.089	-0.007	0.003	0.033	Yes	0.149
		(-0.47)	(-4.42)	(0.36)	(-2.95)	(-0.13)	(0.05)	(1.03)	(-0.12)	(0.51)	(1.40)		
(10)	Δ Prospectus Summary Distance	-0.023	-0.220	0.011	-0.003	0.005	0.214	0.323	0.212	0.001	0.100	Yes	0.128
		(-0.66)	(-3.96)	(1.47)	(-0.45)	(0.38)	(0.68)	(1.10)	(1.05)	(0.04)	(1.25)		
(11)	Δ Risk Factors Distance	0.000	-0.196	0.002	-0.016	-0.002	-0.228	0.378	0.021	-0.028	0.141	Yes	0.083
		(0.00)	(-3.66)	(0.29)	(-2.43)	(-0.12)	(-0.76)	(1.33)	(0.11)	(-1.50)	(1.82)		
(12)	Δ Use of Proceeds Distance	0.072	-0.196	0.001	-0.001	-0.019	0.153	0.036	0.258	-0.007	-0.131	Yes	0.080
		(1.77)	(-3.04)	(0.07)	(-0.13)	(-1.18)	(0.42)	(0.11)	(1.10)	(-0.30)	(-1.41)		
(13)	Δ Mgmt Discussion Distance	-0.054	-0.181	0.004	-0.011	0.001	0.431	0.034	0.041	0.012	-0.017	Yes	0.095
		(-1.61)	(-3.41)	(0.52)	(-1.68)	(0.08)	(1.44)	(0.12)	(0.21)	(0.64)	(-0.22)		

Table X: Effect of Prospectus Changes on Initial Returns and Post-IPO Returns

OLS regressions with yearly fixed effects are presented for 2,044 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, unit IPOs, dual class IPOs, and REITs. The dependent variable is the **Initial Return** (Panel A), which is the actual return from the IPO offer price to the first CRSP reported closing price. In Panel B, the dependent variable is the **1-year post IPO return**. This is the abnormal return from the closing price on the first day of public trading until the IPO's one year anniversary date, and is the intercept of a regression of twelve monthly stock returns on the three Fama-French factors. **Δ Whole Document**, is the change in the natural logarithm of the total number of characters in the prospectus document. All change variables are cumulative, and are based on changes from the initial filing to the final amendment. We also include changes in the percentage of characters allocated to each of four key sections. ps_{pct} is the percentage of prospectus characters that appear in the prospectus summary. rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages for the risk factors section, the use of proceeds section, and the management's discussion and analysis section. For each section, we also report the total distance of the given amendment from the previous filing. Document distance is the dot product of the two normalized vectors, one for each document being compared. Each vector corresponds to the frequency of word roots normalized to have a length of unity. The **Log dollars offered** is the natural logarithm of the offering amount. **Firm age** is the IPO year minus the firm's founding date, where founding dates are obtained from the Field-Ritter dataset, as used in Field and Karpoff (2002). The **Tech Dummy** is equal to unity if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to unity if a firm is VC financed. **UW Market Share** is the lead underwriter's dollar market share in the past calendar year. We construct two measures of **Law Market Share**. Our first is the underwriting firm's legal counsel's dollar market share in the past calendar year. Our second is the issuer firm's legal counsel's dollar market share in the past calendar year. **Pre Offer Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the offer date. **Year and industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries.

Row	Prospectus Section	Δ Percent of Document	Normal. Δ Distribution	Log Firm Age	UW $ Market Share	TECH Dummy	VC Dummy	Pre Offer Nasdaq Return	Log Dollars Offered	Year+Ind Fixed Effects	R^2	Obs
					Panel A: Initial Return							
(1)	Whole Document	-0.313	-0.439	-0.066	3.233	0.069	0.147	1.185	0.031	Yes	0.263	2,044
		(-2.37)	(-2.74)	(-4.61)	(4.85)	(1.70)	(5.02)	(7.59)	(1.70)			
(2)	Prospectus Summmary	1.204	-0.100	-0.061	3.277	0.074	0.150	1.200	0.031	Yes	0.260	2,044
		(1.16)	(-2.13)	(-4.27)	(4.89)	(1.81)	(5.12)	(7.67)	(1.73)			
(3)	Risk Factors	-0.476	-0.101	-0.063	3.203	0.071	0.151	1.206	0.031	Yes	0.259	2,044
		(-0.69)	(-2.08)	(-4.43)	(4.79)	(1.74)	(5.17)	(7.70)	(1.72)			
(4)	Use of Proceeds	2.223	-0.037	-0.062	3.238	0.075	0.151	1.197	0.031	Yes	0.258	2,044
		(0.84)	(-0.91)	(-4.33)	(4.83)	(1.84)	(5.14)	(7.63)	(1.70)			
(5)	Management's Discussion	0.394	-0.200	-0.063	3.331	0.076	0.152	1.180	0.029	Yes	0.264	2,044
		(1.08)	(-4.07)	(-4.46)	(4.99)	(1.89)	(5.21)	(7.55)	(1.63)			
					Panel B: One-Year Post-IPO Return							
(6)	Whole Document	-0.023	0.012	0.004	0.094	0.023	-0.014	-0.097	0.003	Yes	0.042	2,044
		(-0.85)	(0.38)	(1.37)	(0.69)	(2.83)	(-2.34)	(-3.03)	(0.72)			
(7)	Prospectus Summmary	-0.019	-0.008	0.004	0.096	0.024	-0.014	-0.096	0.003	Yes	0.042	2,044
		(-0.09)	(-0.81)	(1.37)	(0.70)	(2.89)	(-2.28)	(-3.01)	(0.69)			
(8)	Risk Factors	0.196	0.006	0.004	0.088	0.025	-0.014	-0.096	0.002	Yes	0.043	2,044
		(1.41)	(0.60)	(1.36)	(0.65)	(2.97)	(-2.33)	(-3.02)	(0.65)			
(9)	Use of Proceeds	-0.131	0.009	0.004	0.090	0.024	-0.014	-0.095	0.003	Yes	0.042	2,044
		(-0.14)	(1.06)	(1.36)	(0.66)	(2.90)	(-2.28)	(-2.96)	(0.70)			
(10)	Management's Discussion	-0.061	0.009	0.004	0.087	0.024	-0.014	-0.096	0.003	Yes	0.042	2,044
		(-0.80)	(0.86)	(1.38)	(0.64)	(2.88)	(-2.33)	(-2.99)	(0.69)			