Strategic Disclosure and the Pricing of Initial Public Offerings

Kathleen Weiss Hanley and Gerard Hoberg *

Current version: February 2008

ABSTRACT

We examine IPO price formation and strategic disclosure by analyzing the word content of 9,818 IPO filings including the initial prospectus as well as each amendment. We find three primary results that motivate extensions to IPO theory. First, disclosure is an important component in price formation as the relative size of the document sections predicts the magnitude of the partial price adjustment, first day IPO returns, and long-run post-offer performance. Second, the writing of the prospectus is collaborative effort involving underwriters, legal counsel, auditors and the issuing firm with different authors performing separate functions in the disclosure of information. A key conclusion is that issuing firm managers, through MD&A, play a surprisingly integral role in the bookbuilding process, as greater management disclosure generates higher offer prices and superior long-run performance. Third, information generated during bookbuilding is asymmetrically disclosed. Positive information is withheld for strategic or proprietary reasons while negative information is disclosed as a hedge against litigation risk.

Although there exists a substantial body of literature on the IPO process and the determinants of IPO pricing, unresolved questions remain regarding the role of information generation and bookbuilding on price formation. In an attempt to answer these questions, extant empirical work has relied on financial or numerical information from the prospectus. This approach utilizes only a small portion, roughly 25%, of the IPO prospectus itself. The remaining 75% of the text has generally been ignored by all but a few researchers and existing studies which do examine the text in the prospectus are limited to examining single sections of the document in isolation.[1] In addition, the IPO prospectus undergoes a number of revisions or amendments during the bookbuilding process and to our knowledge, no study has yet to examine changes in the prospectus and whether new information produced during bookbuilding is publicly revealed in additional filings.

To fill this void and to study how information is generated throughout the IPO process, we examine not only the content of 2,043 initial IPO prospectuses but also how the text of each individual amendment to the prospectus changes throughout the offering period for at total of 9,818 documents. Our study employs a methodology that reads the entire prospectus and measures the size of the total document along with its four most important sections: the Prospectus Summary, the Risk Factors section, Use of Proceeds and Management's Discussion and Analysis (MD&A). 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. Even after controlling for known determinants of IPO pricing and industry fixed effects, a one standard deviation shift in our key section size variables predicts a three to seven percentage point shift in initial returns, an eight percentage point shift in one year post-IPO returns, and a two percentage point shift in the size of price adjustments.

Our method also allows us to assess the contribution of the offer participants in the drafting of the prospectus and each of its sections. By examining the word similarity between documents, we find evidence that the writing of the prospectus

---

[1]See 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 (2007) examine the Risk Factors section.

is a collaborative effort between the issuing firm, lead underwriter, legal counsel and auditor. The underwriter and the issuing firm, through its agents, are most influential in the Prospectus Summary and MD&A.[2] The underwriter is less likely to be involved in the drafting of the Risk Factors and Use of Proceeds sections.

An examination of the type of word content of each of section sheds additional light on the complementary role different parts of the prospectus play as well as their primary function. For example, conversations with practitioners suggest that the Prospectus Summary is the main tool used by underwriters to market the IPO to potential investors and we find that the word content of the Prospectus Summary is most related to product market or business descriptions. In contrast, MD&A is intended to reflect management's assessment of the business of the firm and its word content is heavily weighted toward information on accounting and corporate strategy.

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.[3] Unlike other studies of the effect of disclosure on stock returns, which must control for prior disclosure history, firms undergoing an initial public offering (IPO) 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 Use of Proceeds section has a relatively small contribution in terms of its size in the prospectus, it does have an impact on IPO pricing. We find, consistent with classical disclosure theory, that the larger the Use of Proceeds, the lower is the change in offer prices during bookbuilding and as with Leone, Rock, and Willenborg (2007), the lower is the initial return.

Our results on the size of the Risk Factors section are driven by a simple tradeoff. A larger Risk Factors section reduces potential legal liability and allows a higher

---

[2]We are unable to assess the relative contribution of management in our tests because we require the documents to have the same participants. Since it is very rare that two IPOs have the same management, we cannot directly measure management's influence. Therefore, we rely on the presence of their agents, such as issuer counsel and auditors, and on the assumption that management plays an important role in the drafting of the prospectus in interpreting management's contribution.

[3]See Verrecchia (2001), Dye (2001) and Healy and Palepu (2001) for a review of the literature.

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, this forces the underwriter to price the IPO lower. Our results suggest that the Risk Factors section is, in fact, informative regarding expected firm risk which is consistent with other studies. We find that a larger Risk Factors section leads to a higher divergence of opinion among investors, as measured by price revisions and confirm the findings of Beatty and Welch (1996) and Arnold, Fishe, and North (2007) that an increase in the Risk Factors section is associated with greater initial underpricing and inferior post-IPO performance.

Our results on the role of disclosure through the Prospectus Summary are consistent with classical theories that suggest that greater disclosure by agents 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. If this section is, indeed, used as the primary marketing tool of underwriters, 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 should reduce information asymmetry and lead to smaller changes in offer prices and lower initial returns, we find the opposite to be true. Uniformly, bigger 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 provide an

3

explanation for the source of this improvement in accuracy. Our results suggest that part of this improvement comes from the section devoted to management disclosure. 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.

Finally, we examine how prospectus disclosure is updated 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 amended prospectus increases significantly and the size of the change is negatively associated with initial returns.

The results of this paper support the hypothesis that potential litigation risk is important in both the choice of disclosure and IPO pricing (see for example, Tinic (1988), Ritter and Welch (2002) and Lowry and Shu (2002)). Liability for any material omission in the offering prospectus 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. However, the importance participants place on hedging litigation risk is still unresolved in the literature. For example, Ritter and Welch (2002) state, "in our opinion, leaving money on the table appears to be a cost-ineffective way of avoiding subsequent lawsuits." In contrast, Lowry and Shu (2002) find that the potential costs of litigation are substantial and that underpricing is used as insurance against litigation. Our results are consistent with offering participants

managing potential litigation risk when bad information is revealed by increasing disclosure to reduce the likelihood of a material omission in the prospectus. When bad information is revealed, underwriters may be unable to reduce the offer price sufficiently to insure underpricing and will, instead, increase disclosure as insurance against litigation.

Our empirical findings also motivate two extensions to classical theory. First, although classical bookbuilding theory credits only investors with information production, we find that management disclosure, through MD&A, also plays a central role, especially in an upward direction. 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.

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.

One of the first papers to examine liability risk in the context of the Risk Factors section was Beatty and Welch (1996), who found that the greater the number of risk factors disclosed, the larger is the subsequent underpricing. More recently, Arnold, Fishe, and North (2007) use word counts, both raw and scaled, to determine the effect of the size of the Risk Factors section on IPO pricing. They find that not only does the amount of disclosure in the Risk Factors section matter for short-run price effects but that greater information in this section results in greater underperformance in the long-run.[4] Thus, we expect that the size of the Risk Factors section is related to revisions in the offer price and subsequent returns.

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, Bhattacharya and Ritter (1983), Darrough and Stoughon (1990), Bhattacharya and Chiesa (1995), and Maksimovic and Pichler (2001)).

---

[4]Our definition of disclosure size is similar to theirs except we use characters rather than words.

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 reduce information asymmetry and therefore, mitigate potential changes in the offer price during the bookbuilding process and decrease 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 use 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 use 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) present evidence 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. Conversations with practitioners indicate that 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. Schrand and Verrecchia (2005) find that greater pre-IPO disclosure frequency reduces underpricing, while Cook, Kieschnick, and Ness (2006) present evidence that the greater the number of news articles, prior to going public, the larger is the price revision and underpricing. Finally, Liu, Sherman, and Zhang (2007) argue that the effect of pre-IPO media coverage differs when positive and negative information is revealed during bookbuilding. Thus, our work contributes to the body of literature on the complexity of the disclosure process surrounding IPOs.

Finally, our paper reflects a growing interest in the use of word content analysis to analyze the informativeness of written disclosure and media coverage. In the context of managing litigation risk, Nelson and Pritchard (2007) and Mohan (2007) find that certain word usage is related to the probability of being sued. In other contexts, papers such as Tetlock (2007), Tetlock, Saar-Tsechanksy, and Macskassy (2008), Li (2006) and Boukus and Rosenberg (2006) find word content to be informative in predicting stock price movements.

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.

---

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

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 69 IPOs, leaving us with 2,043 machine readable IPOs. Because these 69 IPOs are a small fraction of our sample, and because most are also small firms that file using an SB-2 (larger firms generally file an S-1), we do not believe that omitting these firms induces any bias into our sample.

Our prospectus reading algorithm 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:

$totchars$: 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 $chars_{rf}$.

---

[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 in the section names which we list. For example, the Prospectus Summary is occasionally called "Summary". The 69 IPOs failing machine readability generally lack pagination or may even lack a Table of Contents.

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 $chars_{mda}$.

Our measure of disclosure, like many in the literature, assumes that greater disclosure is consistent with higher quality disclosure and greater effort on the part of the participants drafting the individual sections. It is not clear, *a priori*, that this should be the case. Because the degree of information asymmetry between the issuing firm and investors is greatest around an IPO, there exists an incentive for the issuing firm (and possibly the underwriter) to obfuscate the information in the prospectus in order to mislead investors into paying a higher price for the shares. We argue that a number of factors constrain such behavior in a way that makes disclosure in the IPO prospectus meaningful.[7]

Including false or misleading information in the prospectus creates three potential risks for the underwriter and the issuing firm. First, when the true value of the information is revealed in the aftermarket, share prices will decline and IPO participants might be sued. Second, obfuscation or false statements in the prospectus can damage reputational capital, particularly underwriters and lawyers, over and above the value of legal damages. Finally, the SEC, who reviews IPO filings, may also scrutinize and comment on the inclusion of useless or misleading statements. In particular, the Plain English rule finalized by the SEC in 1998, states that prospectus writers must avoid the following:[8]

1. Legalistic or overly complex presentations that make the substance of the disclosure difficult to understand;

2. Vague "boilerplate" explanations that are imprecise and readily subject to different interpretations;

---

[7]Note we are not implying that the information in the prospectus is fully revealing but simply that when information is disclosed, it is meaningful. As will be shown later, some information may be withheld from revisions to the document for proprietary or strategic reasons.

[8]This rule was proposed early in 1997 which means that only a small portion our IPO sample occurs before the Commission provided guidance regarding the use of Plain English. For additional information on the rule, see Plain English Disclosure Release Nos. 33-7497; 34-39593, IC-23011 International Series No. 1114, File No. S7-3-97. http://www.sec.gov/rules/final/33-7497.txt

3. Complex information copied directly from legal documents without any clear and concise explanation of the provision(s); and

4. Disclosure repeated in different sections of the document that increases the size of the document but does not enhance the quality of the information.

Given the existence of enhanced scrutiny of disclosure language during the offering, not only by the SEC but also by investors, analysts and the media, the risk of a delayed IPO and the possibility of reputational damage, we argue that prospectus obfuscation is unlikely, and that section sizes are valid measures of quality.

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

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

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 the monthly alpha over the first year after the IPO as the intercept of a regression of excess monthly stock returns (raw returns minus the riskless thirty-day T-bill rate) on the three Fama-French factors (MKT, HML, SMB) plus momentum (UMD):

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

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

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

$\Delta P-$: The negative component of ΔP equal to $\min[\Delta P, 0]$.

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).

Lead UW $ Market Share: Lead underwriter's dollar market share in the past calendar year as calculated by Megginson and Weiss (1991).

Law $ Market Share: 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: Dummy variable equal to one if the firm is VC-backed, and zero otherwise as in Barry, Muscarella, Peavy, and Vetsuypens (1990).

Nasdaq Return: We construct two measures of this variable. 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).

IPO Size: We construct two measures of this variable. 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 one 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 2,043 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 change in the offer price from the first offer price range to the final offer price is 3.5%. 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 $116 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 2.3%. The average market share of the issuer's counsel is lower than that of underwriter counsel at 1.1%. The average market share of the auditing firm is 16.1% which reflects the higher concentration in the audit firm industry. 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%.

III   The Initial Prospectus

Table II presents summary statistics describing the initial prospectus allocation.[9] 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

---

[9]All results throughout the paper are robust to normalizing the section sizes by the number of words in the orthogonal part of the document (i.e., the part of the document other than the four sections we examine). Hence, our results are not driven by correlations induced by the fact that relative sizes are bounded in the interval (0,1), as one section being larger requires that other sections be smaller. Scaling by the orthogonal part of the document removes this correlation, but does not alter our results. We discuss additional robustness tests (for example, in which all section sizes are included in the same regression) later in the paper.

authorship that we conjecture 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.

A   Determinants of the Initial Prospectus

Table III presents the determinants of the document as a whole and each 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. 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 (as measured by dollar detail) is significantly related to higher age, larger offers and non-tech firms. In contrast, we find that the Use of Proceeds section, as a percent of the prospectus document, is greater the smaller the offering, the less likely the offer is VC-backed, and the lower the reputation of the underwriter, lawyers and auditors. We find no relation as to whether the offering is classified as a tech firm or the age of the firm. Our results are more in line with Beatty and Ritter (1986) in which the Use of Proceeds section may be associated with higher 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. 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 but the market share of the auditor is slightly significant. These relations are in line with the conjecture that this section of the document is primarily management's and not the underwriter's responsibility.

B   Evidence of Authorship and Section Word Content

Results using summary statistics suggest that individual prospectus sections might have different authors. In this section, we explore this possibility further. Before doing so, however, we examine whether there is consistency among the underwriting team, e.g. underwriters, lawyers and auditors. Table IV presents both unconditional and conditional probabilities that two IPOs, for example, with the same underwriter, will choose the same lawyer and auditor. We measure the extent to which the two IPOs have similar lead underwriters in three ways: 1) the proportion of the total number of underwriters the IPOs have in common (the commonality of the lead underwriter), 2) whether the IPOs have identical lead underwriters and 3) whether there is at least one common lead underwriter. 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 2,043 IPOs, a maximum of $\frac{2043^2 - 2043}{2}$ unique pairs exist, and hence a maximum of 2,085,903 observations

The unconditional probabilities in Panel A indicate that the probability of any two IPOs having a common underwriter or the same lawyer is quite low at around 2%. The probability of having the same auditor is higher at 14.5% because of the concentration in the accounting industry. Once the probabilities are conditional on the IPO pair having common or the same participants, as in Panel B, the probabilities increase substantially but are never higher than 19%. These results indicate that there is substantial variation in the underwriting team and therefore, our results on authorship are not related to a high degree of correlation among offering participants.

16

In order to determine the authorship of each section, we first construct a variable that measures the degree of similarity between documents, a measure we call "document similarity", which is explained in more detail in Appendix 2.[10] We then examine whether IPOs brought to market by the same underwriter, issuer or manager counsel, or auditor 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]

Table V presents a series of regressions based on the document similarities of the prospectus as a whole and of the individual sections. The dependent variable we use to measure authorship of a section is the document similarity between two initial IPO prospectuses. This is a numerical variable bounded in the interval [0,1] in which a value of zero indicates that the two documents have exactly the same distribution of words while a value of one indicates that the documents are entirely different and have no words in common. As in the previous table, there are 2,085,903 observations for each 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 identify the commonality of IPO i and j's lead underwriting syndicate, whether they have the same underwriter's counsel, the same issuer's counsel or the same auditor. As in Table IV, to account for the situation when more than one underwriter serves as lead, and i and j share at least one lead underwriter, the commonality of the lead underwriter variable is set to the proportion of common lead underwriters (number of common underwriters divided by the total number of lead underwriters). This measure has the nice property of being zero when no lead underwriters are common and one when all lead underwriters are common.

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

---

[10]A similar method of determining authorship was proposed by Markov (1913) as one of the first examples of a Markov chain.

[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 and its auditor.

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

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.

The underwriter's total signature is the sum of the Common Lead UW and Same UW Counsel coefficients.[13] 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) and MD&A (Panel E). While we find support for the conjecture that the Prospectus Summary is heavily authored by the underwriter, our evidence also indicates that the underwriter is providing a template to help management navigate its first public disclosure.

The influence of the underwriter in the remaining two sections is far lower and points to the influence of legal counsel in mitigating potential liability. Beatty and Ritter (1986) contend that more speculative firms are required to provide greater information on the Use of Proceeds (Panel D) and more speculative firms are likely to have larger Risk Factors sections (Panel C). The similarity of these two sections is highest when the lawyers are the same regardless of whether they represent the issuer or the underwriter.

Additional evidence on the differing roles each section plays in information dissemination as well as the potential author is provided by Table VI which examines the document word content. We construct word vectors that relate to specific content as follows: Panel A) the *Product Market* list of 1,180 words is from the four digit SIC

---

[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 common lead underwriters than have the same industry. Even though the underwriter's signature exists for *every IPO*, we can only observe it when the IPOs have common lead underwriters. 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.

code industry definitions excluding financial, accounting and legal terms, Panel B) the *Accounting* list of 329 words are those that appear in the list of COMPUSTAT data items, Panel C) the *Underwriting Strategy* list of 191 words are from the IPO glossary at www.investorwords.com, Panel D) the *Corporate Legal Jargon* list of 132 words is from the corporate legal glossary in www.learnaboutlaw.com and Panel E) the *Corporate Strategy* list of 835 words is from the glossary of the *Strategic Management* textbook of Hill and Jones.[14] We compute the document similarity for each of the four key sections of an IPO to each of the five word content vectors and use this as the dependent variables in a series of regressions.

The Risk Factors word content is most related to discussion of the product markets, underwriting and corporate strategy. Since this section is primarily authored by lawyers, it also has a strong relationship to the Corporate Legal Jargon word list.[15] These relationships are compatible with the role of the Risk Factors section as an analysis of the factors that may adversely affect the business or the offering. The Use of Proceeds section is less likely to incorporate any of these word lists. This may be due to the use of "general corporate purposes" as the most frequently mentioned use of proceeds or the types of words used in this section are not captured by our choice of word lists.

Consistent with our understanding of the roles and uses of the various sections, the Prospectus Summary word content is most likely to include a discussion of the firm's business and the underwriter's offering strategy. In contrast, MD&A includes word content related the financial and operating performance of the firm and is significantly less likely to have word content related to underwriting strategy. The Prospectus Summary and MD&A have the opposite relations to all but one word vector which suggests that these two sections perform separate functions with respect to information dissemination.

These findings supports our conjecture that the underwriter uses the Prospectus Summary as a marketing tool to investors by selling both the business prospects of the firm and its underwriting strategy. While the underwriter helps management by

---

[14] Available at http://college.hmco.com/business/hill/strategic/6e/students/glossary/index.html.
[15] See de R. Barondes, Nyce, and Sanger (2007) for an examination of the role of underwriter counsel in IPOs.

providing a template for disclosure in MD&A, our word content analysis shows that MD&A incorporates words related to the financial and operating performance of the firm and it is management who would provide the best analysis of this information. Indeed, the SEC indicates that "MD&A should be a discussion and analysis of a company's business as seen through the eyes of those who manage that business. Management has a unique perspective on its business that only it can present."[16] The potential author as well as the content of each of the four key sections has important implications for the interpretation of the relationship between section size and subsequent pricing.

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.

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 as 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 VII presents OLS regressions on ΔP, and Tobit regressions on $\Delta P+$ and $\Delta P-$. Panel A of Table VII presents an OLS regression where the dependent variable

---

[16]For additional information see Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations, Release Nos. 33-8350; 34-48960; FR-72.

is the percentage change in the offer price from the midpoint of the file range (ΔP). The independent variable "% of Document" is different for each row in each panel depending on the section considered. For these and all remaining regressions, inclusion of all sections at the same time does not materially affect the results or change the significance.

The sign and significance of the control variables in the regression echoes the findings in the IPO literature on partial adjustment. 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.

Traditional disclosure theories suggest that greater disclosure should reduce information asymmetry and thus the potential for a change in the offer price. There is weak evidence against this hypothesis for the prospectus as whole as larger prospectuses have greater revisions in the offer price. Once the revision is broken into positive and negative, however, this relation disappears.

An examination of the individual sections 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 conveyed by the underwriter in the Prospectus Summary communicates more precise information to investors, ex-ante, which 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. The size of the Use of Proceeds section is negatively related to the change in the offer price which is consistent with a reduction in information asymmetry during the offer period.

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. This result points to one of the key findings of this paper: managers are more integral in the bookbuilding process than the literature would suggest. We argue that the value of management's disclosure is one of the central aspects of information gathering, or in this case, dissemination during bookbuilding.

The asymmetry of disclosure and its impact on price adjustment is examined in last two panels of Table VII, which uses as the dependent variable in Panel B, the positive (ΔP+) and in Panel C, the negative (ΔP-) truncated components of ΔP. The results in Panel B, with the exception of the prospectus as a whole, are similar to those in Panel A for the overall change in offer price. In Panel C, only greater disclosure in the MD&A affects and reduces the magnitude of ΔP-. There is no impact of either the prospectus as a whole or any other section on negative revisions in the offer price.

The findings in Table VII for the Prospectus Summary and the Use of Proceeds are consist classical disclosure theory in which larger section sizes reduce the information asymmetry between the issuing firm and investors. This, in turn, leads to a lower change in the offer price, particularly in a positive direction.

The results on MD&A and the Risk Factors section point to the potential value of the road show in conveying more precise information to investors.[17] 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

---

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

revealed during the road show which translates into higher offer prices. Even when bad information is revealed, as evidenced by ΔP-, greater management disclosure mitigates the size of the change in offer price.

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."[18] Note, however, that the potential for legal liability constrains management from being overly-optimistic in the IPO prospectus. 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 provide grounds for subsequent shareholder lawsuits.

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 and the word content of MD&A might play in the offering process and management's ability to significantly influence the setting of the final offer price. The section devoted to management's disclosure, and only this section, results in higher offer prices regardless of the type of information that is revealed during bookbuilding. This finding is consistent with management efforts to maximize 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 VIII, Panel A, presents the results of an OLS specification, controlling for year and industry fixed effects, that uses 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.

All sections but MD&A and the prospectus as a whole are significantly related to initial returns. Consistent with the hypothesis above, larger Prospectus Summaries

---

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

significantly reduce the amount of underpricing on the first trading day holding other characteristics of the offer constant. A one standard deviation increase in the size of the Prospectus Summary reduces initial returns by 7%. This finding suggests that additional disclosure in this section of the document is able to reduce the amount of money left on the table. Because better disclosure and more accurate pricing might be an effective reputation building tool, this finding might also explain some of the underwriter persistence reported in Hoberg (2007). In particular, underwriters with the strongest need for reputation improvement will both disclose more and underprice less over long periods of time until their reputation is at an optimal level. In contrast, underwriters with surplus reputation capital might disclose less and underprice more aggressively over an extended period of time in order to convert their excess reputational capital into higher medium-term profits.

Consistent with prior literature, we find that the Risk Factors section of the prospectus leads to higher underpricing. Economically, a one standard deviation change in the size of the Risk Factors section increases initial returns by over 6%. Thus, we confirm the findings of the prior literature that greater risk factors may be associated with greater ex ante uncertainty regarding the valuation of the firm. However, it is important to note that this underpricing is associated with a significant positive revision in offer prices during bookbuilding and to our knowledge, there has not been a theory proposed as to why a large Risk Factors section should lead to positive revisions in offer price. One interpretation of these findings is that the Risk Factors section, while listing possible risks, makes no assessment regarding the likelihood of any of the risks occurring. Thus, the pricing of the Risk Factors section may be left, by the underwriter, to investors during the road show. This interpretation is consistent with Benveniste and Spindt (1989) where investors are compensated through greater underpricing by providing valuable information to the issuing firm. Since IPOs with larger Risk Factors sections are also those most exposed to litigation risk, the positive relation between the size of the section and underpricing is also consistent with Lowry and Shu (2002) who contend that underpricing is a hedge against liability. Thus, underpricing in this context may reflect both compensation to investors as well as ex ante liability risk associated with the offering.

Although the determinants of the Use of Proceeds implied that this section may be a proxy for ex ante uncertainty, we find evidence compatible with Leone, Rock, and Willenborg (2007), that the relative size of the Use of Proceeds section reduces underpricing. This finding is consistent with the hypothesis that this section of the prospectus reduces information asymmetry between the issuing firm and investors.

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. However, if underpricing is compensation for information *from* investors, then information *to* investors by management during the road show should not result in underpricing. Thus, the findings on both MD&A and the Risk Factors sections may be interpreted as initial support for Benveniste and Spindt (1989).[19]

We examine the post-offering returns of IPOs in Panel B of Table VIII and find that both the Risk Factors section and MD&A are related to returns one year after the offering. The coefficient on the Risk Factors section is negative indicating that firms with larger Risk Factors section underperform, thus, confirming the finding of Arnold, Fishe, and North (2007). One interpretation is that investors in the aftermarket underestimate the true risk of the firm. This leads to overvaluation at the time the offer begins trading and results in subsequent price declines.

Unique to our paper and consistent with our view on the importance of management to bookbuilding, we find that the amount of information provided by management through MD&A is positively related to the long-run performance of the firm. Not only is this finding statistically significant, it is economically significant as well. A one standard deviation increase in the size of the MD&A increases the monthly alpha by 69 basis points or over 8% per year. This finding points to the credibility of managerial disclosures and suggests that investors who listen to management are rewarded with superior long-run performance.[20]

---

[19]In the Conclusion to the paper we posit an alternative, but not mutually exclusive reason, why firms may only partially adjust in response to positive information revealed during the road show.

[20]To examine robustness, we reproduce the long-run returns using twelve month buy-and-hold, style matched abnormal returns as suggested by Barber and Lyon (1997). We confirm that the

IV   Changes in the Prospectus

This section examines whether new information gathered through the road show becomes incorporated into amendments to the initial prospectus.[21] Although the SEC requires the issuing firm to disclose all material information in the prospectus, the benefits of withholding valuable, positive information for proprietary reasons may outweigh the potential cost of a possible material omission in the prospectus. However, the omission of bad information from the prospectus results in higher than expected offer prices that subsequently fall in the aftermarket once the 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 book-building 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 IPO's 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

---

MD&A section size still positively predicts ex-post one year performance using this method. We report results based on the Fama-French plus momentum intercepts to maintain consistency with Loughran and Ritter (2000), whose unique focus is on IPOs.

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

raw distances tend to have extreme outliers.[22] 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.

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 $chars_{mda}$.

$\Delta normdist\ tot$: The total normalized distance from the initial prospectus to the final revision for the entire prospectus document.

$\Delta normdist\ ps$: The total normalized distance for the Prospectus Summary section.

$\Delta normdist\ rf$: The total normalized distance for the Risk Factors section.

$\Delta normdist\ use$: The total normalized distance for the Use of Proceeds section.

$\Delta normdist\ mda$: The total normalized distance for MD&A.

---

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

Table IX 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 IX, the average 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 Proceeds 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 IX 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 X presents OLS regressions on 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.

Panel A presents the determinants of changes in the document allocation for the

prospectus as a whole and the four subsections. The only sections whose size is significantly affected by information revealed in the offering process (as proxied by ΔP+ and ΔP-) are the Prospectus Summary and MD&A. The Prospectus Summary gets larger when either good or bad information is revealed while MD&A only gets slightly larger when bad information is revealed. Otherwise, there is no significant increase in the size of either the entire document or the remaining two subsections.

Panel B presents evidence of the impact of ΔP on the normalized distances or change in content. We find little evidence that positive price adjustments affect the information content in the amendments to the initial prospectus. Only for the MD&A section do we see a marginally significant relationship between ΔP+ and the normalized distance. Interestingly, as ΔP+ becomes larger, there is less content change in MD&A indicating that management's disclosure remains fairly constant. This is consistent with management's positive assessment of the firm initial prospectus being validated (and priced) during the road show.

The change in the content of the sections is most dramatic when negative information is revealed and the offer price is lower than the midpoint of the filing range. 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 and each of the sections 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.[23]

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 XI 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

---

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

post-offering returns.

We find that changes in the size of the total document reduces initial returns but only the Use of Proceeds section has any effect on 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 most 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.[24] 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 attendant information asymmetry.

Panel B of Table XI 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 methodology that examines 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

---

[24]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.

and find that the writing of the prospectus is a collaborative effort among the IPO participants: underwriters, lawyers, auditors and presumably, management.

Our findings on both content and authorship support the conjecture that the underwriter is the most likely author of the Prospectus Summary and the effect of the size of this section of the prospectus is consistent with traditional theories of disclosure. Greater information in the Prospectus Summary reduces the degree of information asymmetry between the issuing firm and potential investors which results in smaller partial adjustments and lower initial returns. These findings suggest that underwriter disclosure through the Prospectus Summary improves the efficiency of IPO prices.

Although the underwriter plays a part in the authorship of MD&A, both regulation and content point to management as the most likely author of the MD&A. Greater management-driven disclosure is associated with a higher final offer price and superior post-IPO abnormal returns. There is no relation between the size of MD&A and underpricing which is consistent with information dissemination rather than information gathering. In this case, management is providing information to investors which should not result in any compensation through underpricing. 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 and the superior subsequent post-IPO performance indicates that management's contributions are credible. These findings motivate extensions to classical book building theory, which currently focuses only on investors as only information providers, to account for management's participation.

The Risk Factors section is authored by attorneys and consistent with prior studies, we find that the size of this section is positively related to offer price changes during bookbuilding and subsequent initial returns. Our results indicate that a larger Risk Factors section results in additional information being revealed during the offer process that is then incorporated into subsequent underpricing. This underpricing may be compensation to investors for revealing information and/or bearing greater ex ante risk.

31

While the size of the Use of Proceeds section is quite small in relation to the other sections of the prospectus, it plays an important role in reducing information asymmetry. Like the Prospectus Summary, the larger is the Use of Proceeds section, the smaller is the change in the offer price from the midpoint of the filing range and underpricing. These results point to the important role that even small amount of text can play in IPO pricing.

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 that has proprietary value.

Overall, our findings provide a possible new explanation of the partial adjustment phenomenon that incorporates the need to balance the two forms of liability risk associated with IPOs: the probability of a material omission in the prospectus or the probability that a lawsuit will be triggered by a fall in the aftermarket trading price below the offer price. When positive information is revealed during bookbuilding, it may be withheld from investors to preserve proprietary advantages, which leaves the issuer and underwriter especially prone to litigation risk associated with a material omission in the prospectus. Partial adjustment then arises because underwriters set the final IPO price low in order to mitigate the probability of a lawsuit being filed and protect valuable reputational capital. Thus, our argument suggests that 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.

When bad information is revealed, however, the ability of the underwriter to withhold negative information and hedge against price declines by underpricing becomes more difficult. Therefore, our results indicate that when bad information is revealed, liability risk is managed through reducing the probability of a material omission by

increasing the amount of disclosure. Price declines, at least in the short-run, are mitigated by price stabilization. Further research along these dimensions might also explain why some underwriters both persistently underprice more than others and engage in costly price stabilization.

Appendix 1

Our algorithm to read each prospectus and prospectus amendment is written in a combination of PERL and APL. 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,[25] 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.

---

[25] 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 same word root "display".[26] In order to conserve computing space, we exclude articles, conjunctions, personal pronouns, abbreviations, compound words and any words that appear fewer than a total of five times from these counts because they are not informative regarding content. This leaves a vector of 5,803 possible words.[27] 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.

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

---

[26]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.

[27]Our results are unchanged if we use all words or characters in the document.

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 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:

$$Total\ Normalized\ Distance = \frac{\sum\limits_{i=1}^{I-1} Distance_{i,i+1}}{Max[Distance_{1,2}, ..., Distance_{I-1,I}]} \qquad (6)$$

When there is only one document in a series (211 IPOs, see Table IX), 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.

$\Delta normdist\ tot$: 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).

$\Delta normdist\ ps$: 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 $\{chars_{ps,1}, ..., chars_{ps,1}\}$.

$\Delta normdist\ rf$: 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 $\{chars_{rf,1}, ..., chars_{rf,1}\}$.

$\Delta normdist\ use$: 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 $\{chars_{use,1},$ ..., $chars_{use,1}\}$.

$\Delta normdist\ mda$: 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, 2007, The effects of "Risk-Factor" disclosure on the pricing of IPOs and long run returns, University of Richmond Working Paper.

Barber, Brad M., and John D. Lyon, 1997, Detecting long-run abnormal stock returns: The empirical power and specification of test statistics, *Journal of Financial Economics* 43, 341–372.

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.

Bhattacharya, Sudipto, and Jay Ritter, 1983, Innovation and communication: Signalling with partial disclosure, *Review of Economic Studies* pp. 331–346.

Boukus, Ellyn, and Joshua Rosenberg, 2006, The information content of fomc minutes, .

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.

de R. Barondes, Royce, Charles Nyce, and Gary C. Sanger, 2007, Underwriters' counsel as gatekeeper or turnstile: An empirical analysis of law firm prestige and performance in ipos, *Capital Markets Law Journal* 2, 164–190.

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 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.

Li, Feng, 2006, Do stock market investors understand the risk sentiment of corporate annual reports?, University of Michigan Working Paper.

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, 2000, Uniformly least powerful tests of market efficiency, *Journal of Financial Economics* 55, 361–389.

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.

Markov, A.A., 1913, Classical text in translation: An example of statistical investigation of the text *eugene onegin* concerning the connection of samples in chains, in *Science in Context (2006)* (Cambridge University Press: New York).

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

Mohan, Saumya, 2007, Disclosure quality and its effect on litigation risk, University of Texas working paper.

Nelson, Karen, and A.C. Pritchard, 2007, Litigation risk and voluntary disclosure: The use of meaningful cautionary language, Rice University working paper.

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

Ritter, Jay, and Ivo Welch, 2002, A review of ipo activity, pricing, and allocations, *The Journal of Finance* 57, 1795–1828.

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

Tetlock, Paul, 2007, Giving content to investor sentiment: The role of media in the stock market, *The Journal of Finance* 62, 1139–1168.

Tetlock, Paul, Maytal Saar-Tsechanksy, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *The Journal of Finance* p. forthcoming.

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 2,043 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** is the actual return from the IPO offer price to the first CRSP reported closing price. ΔP is the return from the filing date midpoint to the IPO offer price, and $\Delta P+$ and ΔP- are its positive and negative truncated components. The **1-year Post IPO Return** is the monthly alpha over the first year after the IPO computed as the intercept of a regression of excess monthly stock returns (raw returns minus the riskless thirty-day T-bill rate) on the three Fama-French factors (MKT, HML, SMB) plus momentum (UMD). 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 one if a firm resides in a technology industry as identified in Loughran and Ritter (2004). The **VC Dummy** is equal to one if a firm is VC financed. **Lead UW $ Market Share** is the lead underwriter's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Issuer Law $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. **Auditor $ Market Share** is the auditor's dollar market share in the past calendar year. **Pre-File Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the filing date and **Pre-Offer Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the issue date.

Variable	Mean	Std. Dev.	Minimum	Median	Maximum
Panel A: Price Variables					
Initial Return	0.332	0.649	-0.399	0.122	6.267
Price Adjustment(ΔP)	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
1-Year Post-IPO Return	-0.048	0.116	-1.084	-0.006	1.088
Panel B: IPO Characteristics					
IPO Size at Filing	187.031	1157.78	2.750	56.100	46926.1
IPO Size at Offering	115.640	333.512	2.320	55.500	8680.0
Firm Age	13.757	20.110	0.000	7.000	165.000
Tech Dummy	0.435	0.496	0.000	0.000	1.000
VC Dummy	0.471	0.499	0.000	0.000	1.000
Lead UW $ Market Share	0.029	0.026	0.000	0.023	0.147
UW Law $ Market Share	0.023	0.033	0.000	0.010	0.216
Issuer Law $ Market Share	0.011	0.022	0.000	0.004	0.177
Auditor $ Market Share	0.161	0.083	0.000	0.165	0.557
Pre-File Nasdaq Return	0.052	0.090	-0.260	0.058	0.350
Pre-Offer Return 30 days Before Offer	0.050	0.084	-0.265	0.053	0.359

Table II: Summary Statistics: Prospectus and Individual Sections

Summary statistics are reported for 2,043 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. **totchars** is the total number of characters in the prospectus document after removing headers and footers. ps_{pct}, rf_{pct}, use_{pct}, and mda_{pct} are the corresponding percentages of the total characters in the prospectus allocated to each of four key sections: Prospectus Summary, Risk Factors, Use of Proceeds, and MD&A. In Panel B, Pearson correlations for each of the document sections are presented based on each section's raw total number of characters. In Panel C, Pearson correlations are presented 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	215222	70348.4	91537.0	201289	775291
ps_{pct}	Prospectus Summary % of document	0.061	0.024	0.016	0.055	0.182
rf_{pct}	Risk Factors % of document	0.181	0.050	0.044	0.182	0.447
use_{pct}	Use of Proceeds % of document	0.009	0.005	0.000	0.008	0.045
mda_{pct}	Mgmt Discussion % of document	0.131	0.045	0.025	0.129	0.373
	All Four Sections % of document	0.382	0.058	0.131	0.382	0.790

	Variable	Total Document Characters	Prospectus Summary	Risk Factors	Use of Proceeds
		*Panel B: Correlation Coefficients (**Raw Character Sizes**)*			
(1)	Prospectus Summary Total Characters	0.685*			
(2)	Risk Factors Total Characters	0.609*	0.165*		
(3)	Use of Proceeds Total Characters	0.039	0.081	0.036	
(4)	Mgmt Discussion Total Characters	0.675*	0.461*	0.358*	-0.048
		*Panel C: Correlation Coefficients (**Relative Section Sizes**)*			
(1)	Prospectus Summary % of document	0.080*			
(2)	Risk Factors % of document	-0.169*	-0.467*		
(3)	Use of Proceeds % of document	-0.408*	0.111	0.066	
(4)	Mgmt Discussion % of document	0.081*	0.042	-0.172*	-0.141*

Table III: Determinants of Initial Prospectus

OLS regressions in which the dependent variables are the **Log Document Characters** in Row 1, which is the total number of characters in the prospectus document after removing headers and footers and in Rows 2 to 5, the percentage of characters in the prospectus allocated to each of the following four key sections: Prospectus Summary, Risk Factors, Use of Proceeds, and MD&A. **Log Dollars Filed** is the natural logarithm of the initial filing amount. **Log Firm Age** is the log of 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 **VC Dummy** is equal to one if a firm is VC financed. **UW $ Market Share** is the lead underwriter's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Issuer Counsel $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. **Auditor $ Market Share** is the auditor's dollar market share in the past calendar year. The **Tech Dummy** is equal to one if a firm resides in a technology industry as identified in Loughran and Ritter (2004). **Pre-File Nasdaq Return** 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. T-statistics are in parentheses and are adjusted for clustering within industries and years.

Dependent Row Variable	Log Dollars Filed	Log Firm Age	VC Dummy	UW $ Market Share	UW Law $ Market Share	Iss Law $ Market Share	Auditor $ Market Share	Tech Dummy	Pre-File Nasdaq Return	Year+Ind Fixed Effects	R^2
(1) Log Document Characters	0.077 (6.69)	0.014 (2.29)	0.025 (2.46)	0.791 (2.99)	0.504 (2.73)	0.532 (2.81)	0.101 (1.71)	0.040 (2.95)	0.078 (1.75)	0.489	2,043
(2) % Prospectus Summary	0.002 (3.58)	0.001 (0.86)	-0.010 (-10.77)	0.075 (2.60)	-0.003 (-0.22)	-0.024 (-1.26)	-0.003 (-0.44)	-0.009 (-4.98)	-0.007 (-1.64)	0.412	2,043
(3) % Risk Factors	-0.007 (-3.89)	-0.010 (-7.75)	0.009 (4.30)	-0.200 (-4.90)	-0.063 (-2.51)	0.124 (2.78)	-0.028 (-2.17)	0.021 (3.30)	-0.008 (-0.48)	0.359	2,043
(4) % Use of Proceeds	-0.001 (-6.70)	-0.000 (-1.22)	-0.001 (-5.17)	-0.029 (-6.39)	-0.009 (-3.18)	-0.011 (-2.35)	-0.008 (-5.42)	-0.000 (-1.18)	-0.002 (-1.94)	0.353	2,043
(5) % Management's Discussion	0.002 (2.36)	0.007 (5.73)	-0.000 (-0.16)	-0.021 (-0.49)	0.002 (0.06)	-0.057 (-1.19)	0.024 (1.93)	0.006 (1.59)	0.005 (0.51)	0.312	2,043

Table IV: Common Pairings of Underwriters, Legal Counsel, and Auditors

Estimated probabilities summarizing usage of the same agents for various IPO pairs. One observation is one pair of IPOs i and j, including all unique possible IPO pairs as observations (excluding pairs in which $i = j$). For the sample of 2,043 IPOs, $\frac{2043^2 - 2043}{2}$ unique pairs and hence, 2,085,903 observations exist. The four variables identify whether IPOs i and j employ the services of common lead underwriters, the same auditor, the same manager's counsel, and the same issuer's counsel. To accommodate the case when more than one underwriter serves as lead, and i and j share at least one lead underwriter, the **Commonality of Lead UW** is equal to the proportion of common lead underwriters (number of common underwriters divided by the total number of lead underwriters).

	Variable	Commonality of Lead UW	Same UW Counsel	Same Issuer Counsel	Same Auditor
Panel A: Unconditional Probabilities					
(1)	Unconditional (all IPO pairs)	0.021	0.022	0.015	0.145
Panel B: Conditional Probabilities					
(2)	IPOs i and j have identical lead UWs	1.000	0.158	0.044	0.178
(3)	IPOs i and j have at least one common lead UW	0.145	0.046	0.024	0.173
(4)	IPOs i and j have same UW counsel	0.048	1.000	0.050	0.169
(5)	IPOs i and j have same issuer counsel	0.034	0.074	1.000	0.190
(6)	IPOs i and j have same auditor	0.025	0.026	0.020	1.000

Table V: Determinants of Document Similarity and Authorship

OLS regressions in which the dependent variable is the **Document Distance** of two initial IPO prospectuses. One observation is one pair of IPOs i and j, and included are all unique possible IPO pairs as observations (excluding pairs in which $i = j$). For the sample of 2,043 IPOs, $\frac{2043^2 - 2043}{2}$ unique pairs exist, and hence 2,085,903 observations appear in any regression. To ensure T-statistics remain unbiased given the repeated use of each document, T-statistics are adjusted for clustering by IPO. To compute document distance, first a vector is constructed with a length equal to the number of words in the union of both documents. For documents i and j, this vector is populated with the document's frequency of each word, and then normalized 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 four variables identify whether the document sections of IPO i and j are likely written by common lead underwriters, the same underwriter's counsel, the same issuer's counsel, and the same auditor. When more than one lead underwriter exists, **Common Lead UW** is set to the proportion of common lead underwriters (number of common underwriters divided by the total number of lead underwriters). The next four dummy variables are one if IPO i and j reside in the same one digit to four digit SIC code, respectively. **Same IPO Year** is a dummy variable identifying whether IPO i and j are issued in the same year, and **Both Tech IPOs** is a dummy indicating whether both are in Tech oriented as identified in Loughran and Ritter (2004). Finally, included are 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, for each of the following: 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	Common Lead UW	Same UW Counsel	Same Issuer Counsel	Same Auditor	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.	R^2	Obs
Panel A: Entire Document															
(1)	0.071	0.037	0.065	0.007										0.011	2,085,903
	(11.62)	(17.10)	(26.80)	(6.03)											
(2)	0.046	0.029	0.052	0.005	0.006	0.045	0.054	0.048		0.033				0.129	2,085,903
	(8.40)	(15.53)	(22.75)	(4.61)	(2.76)	(11.62)	(14.53)	(17.54)		(17.70)					
(3)	0.037	0.023	0.040	0.004	0.007	0.044	-0.006	0.051	0.069	0.015	-0.014	-0.005	-0.000	0.203	2,085,903
	(6.98)	(14.03)	(19.84)	(3.42)	(3.55)	(12.08)	(-1.52)	(19.15)	(27.23)	(7.86)	(-13.69)	(-7.77)	(-2.83)		
Panel B: Prospectus Summary Section															
(4)	0.084	0.024	0.040	0.009										0.009	2,085,903
	(17.90)	(15.99)	(18.38)	(8.74)											
(5)	0.068	0.019	0.032	0.007	-0.001	0.037	0.033	0.023		0.020				0.072	2,085,903
	(15.26)	(13.65)	(15.30)	(7.72)	(-0.64)	(11.94)	(10.69)	(9.91)		(12.13)					
(6)	0.062	0.016	0.026	0.006	-0.000	0.037	0.003	0.024	0.035	0.008	-0.007	-0.003	-0.000	0.098	2,085,903
	(13.76)	(11.88)	(13.10)	(6.82)	(-0.28)	(12.18)	(0.97)	(10.52)	(15.22)	(4.47)	(-7.61)	(-6.07)	(-0.43)		
Panel C: Risk Factors Section															
(7)	0.058	0.035	0.056	0.008										0.008	2,085,903
	(8.70)	(15.62)	(23.00)	(6.12)											
(8)	0.026	0.026	0.043	0.005	0.004	0.059	0.046	0.041		0.048				0.129	2,085,903
	(4.48)	(13.57)	(18.47)	(4.39)	(1.70)	(15.85)	(12.75)	(14.83)		(24.28)					
(9)	0.013	0.020	0.031	0.003	0.005	0.058	-0.013	0.044	0.069	0.021	-0.012	-0.008	-0.000	0.200	2,085,903
	(2.23)	(11.74)	(14.94)	(2.66)	(2.34)	(15.61)	(-3.42)	(16.57)	(24.53)	(9.72)	(-8.78)	(-10.50)	(-2.90)		
Panel D: Use of Proceeds Section															
(10)	0.026	0.046	0.109	0.019										0.013	2,085,903
	(3.27)	(15.42)	(33.25)	(11.56)											
(11)	0.008	0.039	0.098	0.017	-0.025	0.065	0.055	0.018		0.022				0.057	2,085,903
	(1.01)	(14.06)	(30.39)	(10.91)	(-10.72)	(13.71)	(11.35)	(5.53)		(8.45)					
(12)	-0.001	0.031	0.082	0.015	-0.023	0.063	-0.011	0.019	0.074	0.002	-0.028	-0.005	-0.000	0.121	2,085,903
	(-0.12)	(12.55)	(28.30)	(9.94)	(-10.56)	(13.92)	(-1.98)	(6.05)	(18.69)	(0.70)	(-16.77)	(-6.28)	(-1.08)		
Panel E: Management's Discussion and Analysis Section															
(13)	0.107	0.037	0.069	0.010										0.014	2,085,903
	(17.78)	(19.06)	(27.31)	(8.32)											
(14)	0.090	0.031	0.060	0.009	0.010	0.020	0.041	0.051		0.019				0.072	2,085,903
	(15.68)	(17.76)	(24.33)	(7.49)	(4.11)	(5.20)	(11.38)	(17.97)		(10.30)					
(15)	0.087	0.027	0.051	0.008	0.011	0.020	-0.002	0.053	0.050	0.013	-0.014	-0.001	-0.000	0.110	2,085,903
	(15.26)	(16.81)	(22.51)	(7.04)	(4.68)	(5.47)	(-0.44)	(19.04)	(19.44)	(6.55)	(-14.11)	(-2.28)	(-1.98)		

Table VI: Examination of Document Word Content

OLS regressions in which dependent variable is the **Document Distance** of a given IPO prospectus section to a document consisting of a list of words related to product markets as defined by SIC code categories (Panel A), accounting terminology (Panel B), underwriting strategy (Panel C), corporate legal jargon (Panel D), and Corporate Strategy (Panel E). One observation is one section of one prospectus. There are four sections: Prospectus Summary, Risk Factors, Use of Proceeds, and MD&A. Because there are 2,043 IPOs in the sample, there are 2043 x 4 = 8172 observations in each regression. The product market word list in Panel A is based on the 1,180 word list from the four digit SIC code industry definitions excluding financial, accounting and legal terms. The accounting words in Panel B are the 329 words that appear in the list of COMPUSTAT data items. The underwriting strategies in Panel C are the 191 of the terms in the IPO terms glossary on www.investorwords.com that are most related to underwriting strategies (e.g. Green Shoe, Overallotment, Aftermarket, Firm Commitment, Best Efforts). Corporate legal jargon in Panel D are the 132 words from the corporate legal glossary listed in www.learnaboutlaw.com. The corporate strategy word list in Panel E are the 835 words in the glossary of the Strategic Management textbook of Hill and Jones (http://college.hmco.com/business/hill/strategic/6e/students/glossary/index.html). T-statistics are in parentheses and are adjusted for clustering within industries and years. Although not displayed to conserve space, controls for Tech firms, VC-backing, size, recent Nasdaq returns, and the market share of issuer and underwriting counsel are also included. See Table III for control variable definitions.

Row	Prospectus Summary Dummy	Mgmt's Disc. Dummy	Risk Factors Dummy	Use of Proceeds Dummy	Log Firm Age	UW $ Market Share	Year+Ind Fixed Effects	R^2
			Panel A: Product Market					
(1)	0.020 (19.56)				0.000 (1.60)	-0.017 (-1.67)	Yes	0.206
(2)		0.005 (3.36)			0.000 (1.60)	-0.017 (-1.67)	Yes	0.062
(3)			0.011 (13.12)		0.000 (1.60)	-0.017 (-1.67)	Yes	0.096
(4)				-0.036 (-79.05)	0.000 (1.60)	-0.017 (-1.67)	Yes	0.537
			Panel B: Accounting Vocabulary					
(5)	-0.007 (-5.40)				0.004 (6.85)	-0.004 (-0.15)	Yes	0.057
(6)		0.095 (36.05)			0.004 (6.85)	-0.004 (-0.15)	Yes	0.546
(7)			-0.036 (-23.71)		0.004 (6.85)	-0.004 (-0.15)	Yes	0.125
(8)				-0.052 (-21.88)	0.004 (6.85)	-0.004 (-0.15)	Yes	0.200
			Panel C: Underwriting Strategy					
(9)	0.002 (2.38)				-0.001 (-2.65)	0.014 (0.77)	Yes	0.079
(10)		-0.023 (-23.74)			-0.001 (-2.65)	0.014 (0.77)	Yes	0.192
(11)			0.022 (18.36)		-0.001 (-2.65)	0.014 (0.77)	Yes	0.178
(12)				-0.001 (-0.41)	-0.001 (-2.65)	0.014 (0.77)	Yes	0.078
			Panel D: Corporate Legal Jargon					
(13)	-0.016 (-22.20)				0.000 (0.50)	-0.010 (-0.60)	Yes	0.103
(14)		0.014 (5.94)			0.000 (0.50)	-0.010 (-0.60)	Yes	0.090
(15)			0.036 (30.34)		0.000 (0.50)	-0.010 (-0.60)	Yes	0.258
(16)				-0.033 (-19.02)	0.000 (0.50)	-0.010 (-0.60)	Yes	0.229
			Panel E: Corporate Strategy					
(17)	-0.003 (-1.32)				0.001 (2.04)	0.019 (0.96)	Yes	0.061
(18)		0.018 (8.70)			0.001 (2.04)	0.019 (0.96)	Yes	0.091
(19)			0.045 (31.62)		0.001 (2.04)	0.019 (0.96)	Yes	0.260
(20)				-0.059 (-26.39)	0.001 (2.04)	0.019 (0.96)	Yes	0.414

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

OLS (Panel A) and Tobit (Panels B and C) regressions with yearly fixed effects are presented for 2,043 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 $\Delta P+$ and $\Delta P-$ (Panels B and C) are its positive and negative truncated components. Each regression utilizes a different independent variable depending on the section used. In the first row in each panel, the "% of document field" refers to the explanatory variable **Whole 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 Edgar). In the remaining four rows in each panel, this column refers to the percentage of characters allocated to each of four key sections: Prospectus Summary, Risk Factors, Uses of Proceeds, and MD&A. **Log Firm Age** is the log of 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). **UW $ market share** is the lead underwriter's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Issuer Law $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. The **VC Dummy** is equal to one if a firm is VC financed. **Auditor $ Market Share** is the auditor's dollar market share in the past calendar year. **Pre-File Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the filing date. The **Log Dollars Filed** is the log of the original filing amount. **Year and Industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries. The **Tech Dummy**, based on Loughran and Ritter (2004), is also included but not shown to conserve space. T-statistics are in parentheses and are adjusted for clustering within industries and years.

Prospectus Row Section	% of Document	Log Firm Age	UW $ Market Share	UW Law $ Market Share	Iss Law $ Market Share	Auditor $ Market Share	VC Dummy	Pre-File Nasdaq Return	Log Dollars Filed	Year+Ind Fixed Effects	R^2
Panel A: ΔP (OLS)											
(1) Whole Document	0.039	-0.013	2.255	-0.258	0.463	0.026	0.027	0.117	-0.021	Yes	0.191
	(1.86)	(-2.33)	(6.96)	(-0.98)	(1.63)	(0.41)	(1.79)	(1.48)	(-2.09)		
(2) Prospectus Summary	-0.846	-0.012	2.349	-0.241	0.463	0.027	0.019	0.114	-0.016	Yes	0.194
	(-2.93)	(-2.19)	(7.27)	(-0.91)	(1.70)	(0.44)	(1.36)	(1.45)	(-1.75)		
(3) Risk Factors	0.358	-0.009	2.357	-0.216	0.439	0.039	0.024	0.123	-0.015	Yes	0.193
	(2.40)	(-1.65)	(7.00)	(-0.82)	(1.62)	(0.63)	(1.67)	(1.61)	(-1.64)		
(4) Use of Proceeds	-2.377	-0.013	2.218	-0.259	0.457	0.010	0.025	0.116	-0.021	Yes	0.192
	(-1.90)	(-2.29)	(6.94)	(-0.98)	(1.63)	(0.17)	(1.72)	(1.47)	(-2.14)		
(5) Management's Discussion	0.513	-0.017	2.296	-0.240	0.513	0.017	0.028	0.118	-0.019	Yes	0.195
	(2.41)	(-2.88)	(7.05)	(-0.92)	(1.84)	(0.29)	(1.85)	(1.48)	(-2.03)		
Panel B: $\Delta P+$ (Tobit)											
(6) Whole Document	0.045	-0.016	3.052	-0.356	0.547	-0.005	0.069	0.211	-0.002	Yes	0.225
	(1.38)	(-1.80)	(7.18)	(-1.09)	(1.54)	(-0.05)	(4.08)	(1.80)	(-0.18)		
(7) Prospectus Summary	-1.891	-0.014	3.198	-0.328	0.518	-0.007	0.051	0.203	0.005	Yes	0.233
	(-4.15)	(-1.55)	(7.88)	(-1.01)	(1.54)	(-0.07)	(3.16)	(1.74)	(0.43)		
(8) Risk Factors	0.489	-0.010	3.175	-0.302	0.511	0.017	0.065	0.219	0.004	Yes	0.228
	(2.27)	(-1.12)	(7.29)	(-0.94)	(1.49)	(0.18)	(3.91)	(1.95)	(0.41)		
(9) Use of Proceeds	-4.789	-0.016	2.960	-0.378	0.523	-0.035	0.064	0.208	-0.005	Yes	0.227
	(-2.24)	(-1.79)	(7.01)	(-1.17)	(1.48)	(-0.39)	(3.95)	(1.79)	(-0.43)		
(10) Management's Discussion	0.790	-0.021	3.121	-0.328	0.627	-0.007	0.070	0.216	-0.000	Yes	0.231
	(2.59)	(-2.29)	(7.22)	(-1.02)	(1.76)	(-0.08)	(4.05)	(1.85)	(-0.04)		
Panel C: $\Delta P-$ (Tobit)											
(11) Whole Document	0.039	-0.002	2.561	-0.127	0.441	-0.054	-0.009	-0.016	-0.040	Yes	0.142
	(1.46)	(-0.33)	(6.14)	(-0.52)	(1.49)	(-0.63)	(-0.55)	(-0.20)	(-4.67)		
(12) Prospectus Summary	0.191	-0.002	2.570	-0.108	0.467	-0.051	-0.006	-0.013	-0.037	Yes	0.141
	(0.57)	(-0.26)	(6.20)	(-0.44)	(1.58)	(-0.60)	(-0.36)	(-0.16)	(-4.38)		
(13) Risk Factors	0.034	-0.001	2.597	-0.107	0.455	-0.051	-0.008	-0.014	-0.037	Yes	0.141
	(0.19)	(-0.19)	(6.17)	(-0.43)	(1.53)	(-0.60)	(-0.50)	(-0.17)	(-4.25)		
(14) Use of Proceeds	-1.029	-0.002	2.558	-0.116	0.447	-0.060	-0.009	-0.016	-0.038	Yes	0.141
	(-0.76)	(-0.27)	(6.21)	(-0.47)	(1.52)	(-0.70)	(-0.56)	(-0.20)	(-4.38)		
(15) Management's Discussion	0.651	-0.006	2.595	-0.113	0.498	-0.067	-0.007	-0.021	-0.038	Yes	0.150
	(3.64)	(-0.95)	(6.29)	(-0.46)	(1.72)	(-0.80)	(-0.44)	(-0.26)	(-4.61)		

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

OLS regressions with yearly fixed effects are presented for 2,043 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 variables are Panel A) the **Initial Return** which is the actual return from the IPO offer price to the first CRSP reported closing price and Panel B) **1-year Post-IPO Return** is the monthly alpha over the first year after the IPO computed as the intercept of a regression of excess monthly stock returns (raw returns minus the riskless thirty-day T-bill rate) on the three Fama-French factors (MKT, HML, SMB) plus momentum (UMD). Each regression utilizes a different independent variable depending on the section used. In the first row in each panel, the "% of document field" refers to the explanatory variable **Whole 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 Edgar). In the remaining four rows in each panel, this column refers to the percentage of characters allocated to each of four key sections: Prospectus Summary, Risk Factors, Uses of Proceeds and MD&A. **Log Firm Age** is the log of the IPO year minus the firm's founding date, as used in Field and Karpoff (2002). **UW $ market share** is the lead underwriter's dollar market share in the past calendar year. **Issuer Law $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Auditor $ Market Share** is the auditor's dollar market share in the past calendar year. The **VC Dummy** is equal to one if a firm is VC financed. The **Log Dollars Filed** is the log of the original filing amount. **Pre-File Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the filing date. The **Tech Dummy**, based on Loughran and Ritter (2004), is also included but not shown to conserve space. **Year and Industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries. T-statistics are in parentheses and are adjusted for clustering within industries and years.

Prospectus Row Section	% of Document	Log Firm Age	UW $ Market Share	UW Law $ Mkt Share	Iss Law $ Mkt Share	Auditor $ Mkt Share	VC Dummy	Pre-File Nasdaq Return	Log Dollars Filed	Year+Ind Fixed Effects	R^2	Obs
					Panel A: Initial Return							
(1) Whole Document	0.072 (1.88)	-0.052 (-2.88)	4.647 (2.68)	-0.180 (-0.32)	1.872 (1.65)	-0.227 (-1.47)	0.141 (3.55)	0.353 (2.07)	-0.049 (-3.25)	Yes	0.243	2,043
(2) Prospectus Summary	-2.926 (-2.78)	-0.049 (-2.86)	4.925 (2.77)	-0.152 (-0.28)	1.840 (1.66)	-0.227 (-1.47)	0.113 (3.34)	0.337 (2.01)	-0.038 (-2.88)	Yes	0.249	2,043
(3) Risk Factors	1.262 (2.42)	-0.038 (-2.80)	4.956 (2.70)	-0.065 (-0.12)	1.754 (1.62)	-0.185 (-1.25)	0.131 (3.43)	0.368 (2.27)	-0.035 (-2.41)	Yes	0.249	2,043
(4) Use of Proceeds	-8.565 (-3.87)	-0.052 (-2.92)	4.460 (2.55)	-0.217 (-0.40)	1.815 (1.60)	-0.290 (-1.87)	0.132 (3.41)	0.342 (2.02)	-0.056 (-3.83)	Yes	0.246	2,043
(5) Management's Discussion	0.679 (1.41)	-0.056 (-3.41)	4.718 (2.69)	-0.145 (-0.26)	1.949 (1.72)	-0.236 (-1.55)	0.143 (3.58)	0.355 (2.10)	-0.045 (-3.23)	Yes	0.244	2,043
					Panel B: One-Year Post-IPO Return							
(6) Whole Document	0.001 (0.08)	0.005 (1.58)	0.240 (2.34)	-0.134 (-2.15)	-0.109 (-0.92)	0.002 (0.05)	-0.013 (-2.15)	0.026 (0.90)	-0.001 (-0.18)	Yes	0.039	2,028
(7) Prospectus Summary	0.027 (0.21)	0.005 (1.59)	0.239 (2.39)	-0.134 (-2.18)	-0.108 (-0.91)	0.002 (0.05)	-0.013 (-2.14)	0.027 (0.91)	-0.001 (-0.17)	Yes	0.039	2,028
(8) Risk Factors	-0.139 (-2.06)	0.003 (1.18)	0.214 (2.09)	-0.142 (-2.33)	-0.091 (-0.76)	-0.002 (-0.05)	-0.012 (-1.94)	0.025 (0.86)	-0.002 (-0.43)	Yes	0.042	2,028
(9) Use of Proceeds	-0.215 (-0.40)	0.005 (1.58)	0.235 (2.31)	-0.136 (-2.20)	-0.111 (-0.95)	0.000 (0.01)	-0.014 (-2.19)	0.026 (0.89)	-0.001 (-0.23)	Yes	0.039	2,028
(10) Management's Discussion	0.153 (2.53)	0.003 (1.25)	0.246 (2.42)	-0.134 (-2.18)	-0.100 (-0.84)	-0.002 (-0.04)	-0.013 (-2.13)	0.026 (0.90)	-0.001 (-0.27)	Yes	0.042	2,028

Table IX: Summary of Prospectus Filing Patterns

The abbreviations **PS**, **RF**, **USE**, and **MD&A** represent Prospectus Summary, Risk Factors, Use of Proceeds, and MD&A, 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. For each section, the total distance of the given amendment from the previous filing is reported. 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 one. Document distance is bounded to be in the interval (0,1). The total number of IPOs for which the given number of prospectuses are filed is reported in the last column.

Amend-ment	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	215,222	0.000	13,175	0.000	38,358	0.000	1,740	0.000	28,553	0.000	2,043
2	226,890	0.035	13,801	0.287	39,368	0.095	2,000	1.074	31,948	0.263	2,015
3	233,634	0.017	13,975	0.140	40,087	0.044	2,028	0.520	33,384	0.094	1,937
4	240,738	0.010	13,998	0.077	41,506	0.024	2,017	0.294	35,032	0.054	1,624
5	251,054	0.008	14,048	0.053	43,465	0.019	2,024	0.216	37,454	0.034	1,113
6	265,164	0.005	14,759	0.056	46,166	0.015	2,069	0.252	41,351	0.036	578
7	268,794	0.005	14,450	0.045	47,451	0.014	2,042	0.194	42,093	0.025	294
8	275,497	0.003	14,282	0.042	49,295	0.009	2,146	0.125	45,397	0.027	129
9	314,610	0.003	15,917	0.033	56,142	0.008	2,420	0.126	55,773	0.011	51
10	303,171	0.002	14,656	0.026	57,890	0.006	2,377	0.067	55,573	0.008	18
11	300,594	0.001	17,530	0.056	55,904	0.007	3,285	0.009	60,165	0.003	7
12	311,267	0.004	18,676	0.085	55,018	0.005	3,482	0.100	66,446	0.050	6
13	178,831	0.220	10,892	1.098	30,796	0.587	5,210	0.274	23,226	0.328	1
14	186,392	0.010	12,183	0.166	30,943	0.007	5,213	0.015	24,838	0.042	1
15	186,581	0.000	12,183	0.000	30,943	0.000	5,213	0.000	24,838	0.000	1

50

Table X: Determinants of Changes in the Prospectus

OLS regressions with yearly fixed effects are presented for 2,043 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 corresponds to either prospectus as a whole or to one of each of the four sections: Prospectus Summary, Risk Factors, Uses of Proceeds or MD&A. In Panel A, **Changes in Document Allocation**, is the change in the natural logarithm of the total number of characters in the prospectus document and the percentage of characters allocated to each of four key sections. In Panel B, **Changes in Document Distance** is the total distance of the given amendment from the previous filing and is measured as the dot product of the two normalized word vectors, one for each document being compared. Each vector corresponds to the frequency of word roots normalized to have a length of one. The **Log Dollars Filed** is the log of the original filing amount. All change variables are cumulative and are based on changes from the initial filing to the final amendment. **Log Firm Age** is the log of 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 **VC Dummy** is equal to one if a firm is VC financed. **UW $ market share** is the lead underwriter's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Issuer Law $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. **Auditor $ Market Share** is the auditor'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. The **Tech Dummy**, based on Loughran and Ritter (2004), is also included but not shown to conserve space. *T*-statistics are in parentheses and are adjusted for clustering within industries and years.

Row	Dependent Variable	ΔP+	ΔP−	Log Dollars Filed	Log Firm Age	VC Dummy	UW $ Market Share	UW Law Market Share	Iss Law Market Share	Auditor Market Share	Pre-Offer Nasdaq Return	Year+Ind Fixed Effects	R^2
	Panel A: Changes in Document Allocation												
(1)	Whole Document	-0.009 (-0.86)	-0.017 (-1.02)	0.004 (1.28)	-0.005 (-1.88)	-0.013 (-3.24)	-0.003 (-0.03)	0.146 (2.02)	-0.029 (-0.35)	-0.039 (-1.60)	-0.029 (-1.34)	Yes	0.105
(2)	Prospectus Summary	0.002 (2.19)	-0.007 (-2.75)	-0.000 (-0.20)	-0.001 (-2.21)	0.001 (1.83)	-0.018 (-1.33)	-0.018 (-2.01)	0.001 (0.09)	0.001 (0.16)	-0.001 (-0.17)	Yes	0.057
(3)	Risk Factors	0.001 (0.32)	-0.003 (-0.93)	0.000 (0.62)	0.001 (1.61)	0.001 (1.13)	0.032 (1.82)	0.012 (0.77)	0.020 (1.05)	0.012 (2.20)	0.001 (0.27)	Yes	0.177
(4)	Use of Proceeds	0.000 (0.80)	0.001 (0.81)	0.000 (0.67)	0.000 (0.02)	-0.000 (-1.32)	-0.009 (-2.25)	-0.006 (-1.69)	-0.002 (-1.36)	-0.003 (-1.53)	-0.002 (-1.50)	Yes	0.112
(5)	Management's Discussion	-0.003 (-1.36)	-0.009 (-1.71)	-0.001 (-1.04)	-0.001 (-0.76)	-0.002 (-1.60)	-0.001 (-0.04)	-0.025 (-1.24)	0.003 (0.12)	-0.008 (-1.03)	-0.005 (-0.67)	Yes	0.049
	Panel B: Changes in Document Distance												
(6)	Whole Document	-0.005 (-0.48)	-0.071 (-4.01)	0.000 (0.13)	-0.006 (-2.41)	0.000 (0.05)	0.010 (0.12)	-0.001 (-0.02)	0.087 (0.98)	-0.010 (-0.43)	0.029 (1.25)	Yes	0.150
(7)	Prospectus Summary	-0.023 (-0.80)	-0.220 (-4.21)	0.010 (1.36)	-0.003 (-0.46)	0.005 (0.38)	0.216 (0.68)	0.203 (1.06)	0.315 (1.18)	0.051 (0.69)	0.099 (1.16)	Yes	0.128
(8)	Risk Factors	0.000 (0.01)	-0.196 (-3.30)	0.003 (0.38)	-0.016 (-2.17)	-0.001 (-0.10)	-0.223 (-0.73)	0.024 (0.13)	0.381 (1.32)	-0.041 (-0.47)	0.141 (1.89)	Yes	0.084
(9)	Use of Proceeds	0.073 (1.62)	-0.194 (-2.74)	0.003 (0.43)	-0.000 (-0.05)	-0.015 (-0.97)	0.194 (0.51)	0.267 (0.82)	0.042 (0.11)	-0.248 (-2.97)	-0.135 (-1.35)	Yes	0.083
(10)	Management's Discussion	-0.054 (-2.06)	-0.180 (-2.77)	0.004 (0.55)	-0.011 (-1.50)	0.002 (0.15)	0.444 (1.74)	0.040 (0.21)	0.033 (0.11)	-0.051 (-0.75)	-0.019 (-0.23)	Yes	0.095

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

OLS regressions with yearly fixed effects are presented for 2,043 IPOs issued in the US from February 1996 to October 2005 excluding: firms with an issue price less than five dollars, ADRs, financial firms, dual class IPOs, unit IPOs, and REITs. The dependent variables are Panel A) the **Initial Return** which is the actual return from the IPO offer price to the first CRSP reported closing price and Panel B) **1-year Post-IPO Return** is the monthly alpha over the first year after the IPO computed as the intercept of a regression of excess monthly stock returns (raw returns minus the riskless thirty-day T-bill rate) on the three Fama-French factors (MKT, HML, SMB) plus momentum (UMD). Each regression in the table utilizes a different independent variable for both Δ **Percent of Document** and **Normal** Δ **Distribution** depending on whether the entire prospectus or a key section is used: Prospectus Summary, Risk Factors, Uses of Proceeds or MD&A. Δ **Percent of Document**, is the change in the natural logarithm of the total number of characters in the prospectus document and for each section. The total distance for a given amendment from the previous filing is represented by **Normal** Δ **Distribution** which 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 one. All change variables are cumulative, and are based on changes from the initial filing to the final amendment. **Log Firm Age** is the log of 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). **UW $ market share** is the lead underwriter's dollar market share in the past calendar year. **UW Law $ Market Share** is the underwriting firm's legal counsel's dollar market share in the past calendar year. **Iss Law $ Market Share** is the issuer firm's legal counsel's dollar market share in the past calendar year. **Auditor $ Market Share** is the auditor's dollar market share in the past calendar year. The **VC Dummy** is equal to one if a firm is VC financed. **Pre-Offer Nasdaq Return** is the NASDAQ return for the 30 trading days preceding the offer date. The **Log Dollars Filed** is the log of the original filing amount. **Year and Industry** fixed effects are also included, where industry definitions are based on the Fama-French 48 industries. The **Tech Dummy**, based on Loughran and Ritter (2004), is also included but not shown to conserve space. *T*-statistics are in parentheses and are adjusted for clustering within industries and years.

Row	Prospectus Section	Δ Percent of Document	Normal. Δ Distribution	Log Firm Age	UW $ Market Share	UW Law Market Share	Iss Law Market Share	Auditor Market Share	VC Dummy	Pre-Offer Nasdaq Return	Log Dollars Offered	Year+Ind Fixed Effects	R²	Obs
						Panel A: Initial Return								
(1)	Whole Document	-0.315 (-2.48)	-0.449 (-2.38)	-0.065 (-3.11)	3.219 (2.02)	-0.433 (-0.89)	1.773 (1.48)	-0.315 (-1.99)	0.145 (3.51)	1.182 (4.92)	0.034 (1.84)	Yes	0.268	2,043
(2)	Prospectus Summary	1.175 (1.59)	-0.100 (-2.37)	-0.060 (-3.02)	3.258 (2.07)	-0.442 (-0.92)	1.773 (1.50)	-0.295 (-1.86)	0.148 (3.57)	1.198 (4.95)	0.035 (1.88)	Yes	0.264	2,043
(3)	Risk Factors	-0.536 (-0.80)	-0.106 (-2.05)	-0.063 (-3.10)	3.194 (2.02)	-0.478 (-1.00)	1.793 (1.48)	-0.299 (-1.93)	0.149 (3.54)	1.204 (4.97)	0.035 (1.91)	Yes	0.264	2,043
(4)	Use of Proceeds	1.618 (1.78)	-0.040 (-1.38)	-0.061 (-3.04)	3.228 (2.05)	-0.471 (-0.95)	1.748 (1.49)	-0.307 (-1.98)	0.149 (3.52)	1.193 (4.99)	0.035 (1.92)	Yes	0.263	2,043
(5)	Management's Discussion	0.018 (0.03)	-0.195 (-3.80)	-0.063 (-3.13)	3.309 (2.07)	-0.467 (-1.00)	1.749 (1.48)	-0.307 (-2.00)	0.149 (3.49)	1.173 (4.92)	0.033 (1.85)	Yes	0.268	2,043
						Panel B: One-Year Post-IPO Return								
(6)	Whole Document	-0.026 (-0.98)	0.014 (0.46)	0.004 (1.49)	0.135 (1.28)	-0.141 (-2.24)	-0.120 (-1.08)	-0.004 (-0.11)	-0.013 (-2.11)	-0.098 (-3.54)	0.004 (1.21)	Yes	0.044	2,028
(7)	Prospectus Summary	0.049 (0.21)	-0.007 (-0.60)	0.004 (1.52)	0.138 (1.30)	-0.141 (-2.28)	-0.116 (-1.04)	-0.003 (-0.08)	-0.013 (-2.08)	-0.096 (-3.42)	0.004 (1.16)	Yes	0.044	2,028
(8)	Risk Factors	0.198 (1.32)	0.007 (0.70)	0.004 (1.48)	0.131 (1.24)	-0.147 (-2.32)	-0.125 (-1.13)	-0.006 (-0.14)	-0.013 (-2.08)	-0.097 (-3.48)	0.004 (1.16)	Yes	0.045	2,028
(9)	Use of Proceeds	-1.114 (-1.03)	0.009 (1.18)	0.004 (1.42)	0.126 (1.17)	-0.150 (-2.35)	-0.122 (-1.09)	-0.003 (-0.08)	-0.013 (-2.10)	-0.095 (-3.42)	0.004 (1.18)	Yes	0.045	2,028
(10)	Management's Discussion	-0.059 (-0.71)	0.009 (0.64)	0.004 (1.52)	0.130 (1.23)	-0.147 (-2.31)	-0.119 (-1.07)	-0.004 (-0.09)	-0.013 (-2.13)	-0.096 (-3.47)	0.004 (1.21)	Yes	0.044	2,028