Is Textual Information Informative to Informed Investors? Evidence from Bidding Information of Institutional Investors in IPOs

Sung Wook Joh*, Yongseok Kim*

Abstract

Building on past studies showing that textual information in an IPO prospectus (*prospectus text*) is related to IPO outcomes and initial returns, this paper investigates whether institutional investors capitalize on prospectus texts. Exploiting a unique institutional setting of IPOs in Korea, we show that IPO firms with positive tones in their prospectus texts often set higher initial offer price bands in advance of bookbuilding, yielding negative final offer price revisions. In contrast, tones and lengths in prospectuses do not affect institutional investors' average bidding prices and oversubscription ratios. While institutional investor bidding price strongly predicts short-term and long-term IPO stock performances, prospectus texts predict long-term IPO stock performance. Our findings suggest that prospectus texts are informative to institutional investors, especially in the long-run.

Keywords: Textual Analysis, Bookbuilding, Offer Price Revisions, Initial Returns, Long-term Performance

*Seoul National Univ.

1 Introduction

When a firm makes an initial public offering (IPO), the issuer and its underwriter prepare registration statements with information about it. These statements contain past records and financial statements (*hard information*) as well as textual information including self-description, future opportunities, risks, and insiders' views not reflected in the hard information (*prospectus text*) (Hanley & Hoberg, 2010; Loughran & McDonald, 2013).

To reduce litigation risk and underpricing due to information asymmetry, issuing firms have an incentive to reveal truthful information in their prospectus texts (Loughran & McDonald, 2013; Rogers, Van Buskirk, & Zechman, 2011).[1] Past studies reported that prospectus text predicted IPO outcomes, suggesting that it can be informative to investors. For example, firms whose prospectus text tones were more negative than those of other firms had greater offer price revisions and greater underpricing (Loughran & McDonald, 2013). Also, issuing firms with more information in preliminary prospectuses than other firms had smaller absolute changes in offer prices and smaller initial returns (Hanley & Hoberg, 2010). Both studies suggest that such textual information indicate risks or uncertainties in IPO firms, and hence have informational value.

Whether information in prospectus texts can be valuable to informed investors (such as institutional investors) remains an open question. Past studies suggest that institutional investors are well informed. Underwriters and IPO firms want to derive information from institutional investors during a bookbuilding process (Benveniste & Spindt, 1989; Benveniste & Wilhelm, 1990; Spatt & Srivastava, 1991). Indeed, Cornelli and Goldreich (2003) and Joh and Kim (2017) show that bidding information of informed institutional investors helps predict offer price revisions and initial returns. We argue that this textual information might have little value to institutional investors for two reasons. One, when prospectus texts include information that institutional investors do not know an issuing firm's initial offer price band might already reflect this information. In this case, the text does not provide additional value. Two, as revealing confidential information is costly to issuing firms (Hanley & Hoberg, 2012), they may disclose textual information that institutional investors already know. In

[1] CEOs often manipulate earnings before IPOs (Teoh, Welch, & Wong, 1998). They can also manipulate textual information in statements. For more details, see Huang, Teoh, and Zhang (2014); Li (2008).

this case, the relationship between textual information and initial returns may stem from retail investors' reactions to textual information but not from those of informed investors. In short, it is unclear whether textual information provides additional value to informed investors.

Exploiting a unique institutional setting of the IPO in Korea, this paper investigates whether textual information provides additional value to informed investors. Specifically, we examine whether tones and amounts of description in IPO prospectuses affect bidding behaviors of institutional investors in a bookbuilding process. In Korea, IPO firms are required to conduct two separated, sequential subscription phases for institutional and retail investors, respectively. IPO firms and underwriters have to publicly disclose final offer prices, and institutional investors' bidding information in final prospectuses before retail investors subscribe. As a result, retail investors have access to both prospectus text and institutional investors' bidding prices and oversubscription ratios. This institutional setting provides information not only on institutional investors' bids but also on how they influence retail investors' subscription behaviors.

Using all IPOs from June 2009 to December 2017 in Korea, we first show that public information such as firm size, firm age, and market returns explain much of the variations in prospectus text tone and length, especially the latter. Small firms or firms raising small IPO proceeds tend to describe themselves with fewer words or sentences. In addition, young and small firms tend to describe themselves more positively, especially when stock market returns are high. More importantly, we find that firms' textual information is linked to their initial offer prices. Firms whose prospectus text has a positive tone set their initial offer price bands higher with a narrower range, suggesting that issuing firms incorporate textual information into their initial offer price bands before bookbuilding.

Second, we find that prospectus texts do not influence institutional investors' bids. Specifically, prospectus text tone and length do not affect the average bidding prices or oversubscription ratios of institutional investors. In the presence of bookbuilding information, when prospectus text tone is positive and the average bidding price and oversubscription ratio of institutional investors are low, the ratio of a finalized offer price over initial offer price is often lower. Hence, one might argue that the prospectus text has informational value but institutional investors do

not use it. We evaluate this argument through stock performance after the IPO. We find that the bidding price of institutional investors strongly predicts the future short-term performance of IPO stocks, while prospectus text tone and length do not. However, bidding price of institutional investors, prospectus text tone, and its length all predict the long-term performances of an IPO stock. These results suggest that prospectus text has informational value in the long run.

Third, we investigate retail investors' subscription decisions after observing textual information from issuing firms and bidding information from institutional investors. When institutional investors bid high and subscribe more for a stock, retail investors often subscribe more. In contrast, prospectus texts are only weakly linked to retail investor subscription. Retail investors' behaviors suggest that they use information on institutional investors' bidding, which strongly predicts short-term IPO stock returns unlike prospectus text. Also, IPO stocks with higher retail investor subscription have higher initial returns, showing that the "mimicking institutional investors strategy" of retail investors is profitable. However, oversubscription ratio of retail investors does not predict higher long-run returns. Similar to Derrien (2005) and Cornelli, Goldreich, and Ljungqvist (2006), this finding suggests that retail investors may overreact during the IPO, which results in long-run reversals. In contrast, institutional bidding price predicts higher long-run returns.

We contribute to the literature by clarifying the mechanisms through which issuing firms incorporate textual information in initial price ranges. Without information on institutional investors' bids, past studies suggest that negative tone or insufficient information in the prospectus text leads IPO firms to low-ball their initial offer price bands, thereby yielding high offer price revisions (Hanley & Hoberg, 2010; Loughran & McDonald, 2013). Using institutions' bidding information however, our study shows that while prospectus text tone and length affect IPO firms' initial price bands, they do not affect institutional investors' bidding prices, their oversubscription ratios, or IPO firms' short-term stock returns. However, the positive link between textual information and long-term performance suggests that prospectus texts can provide informational values to investors.

Our paper is organized as follows. Section 2 briefly summarizes the institutional background. Section 3 presents our sample and methodology. We summarize our empirical results in Section 4 and conclude in Section 5.

2 Institutional Background

Table 1 presents the typical schedule for an IPO procedure in Korea. First, an issuer applies to be listed on the KOSPI or KOSDAQ market, analogous to NYSE and NASDAQ respectively. Once approved, the issuer and its underwriter file the preliminary prospectus. The preliminary prospectus contains two types of information: hard information and textual information. Hard information includes past records, financial statements, initial offer price bands, and the underwriter's estimated "essential value" of the IPO firm. For the estimation of essential values, underwriters typically employ the price multiple methods (e.g., Price Earnings Ratio (PER) valuation) or the Discounted Cash Flow (DCF) method.

Textual information (*prospectus text*) includes the firm's self-description, such as its future opportunities and risks. Issuers are required to describe their risks in a predesigned section (*Risk Factors Section*, see example in Table 2, Panel A), which can include political risk, policy risk, product market risk, labor union risk, funding liquidity risk, geographical risk, and foreign market/currency risk.

Next, underwriters present the IPO firm ("road show"). Korea's regulations require underwriters to organize a subscription phase for institutional investors and a later one for retail investors. The first subscription phase is a bookbuilding procedure, open only to qualified institutional investors. During bookbuilding, institutional investors submit the number of shares they will subscribe and their maximum bidding price, similar to a "limit bid" (Cornelli & Goldreich, 2001). After bookbuilding, the issuing firm and its underwriter finalize the offer price and allocate shares among institutional investors who bid above the offer price. If an institutional investor does not subscribe for its allocated shares without a compelling reason, regulators punish it by excluding it from future bookbuildings. Hence, institutional investors have an incentive to reveal their information as described in the information-acquisition model of Benveniste and Spindt (1989).[2]

[2] Previous studies show that the bookbuilding procedure in Korea works as described in the information-acquisition model. Joh and Kim (2017) show that favorable bids of institutional investors predict the offer price

After bookbuilding, the issuer and its underwriter submit the final prospectus. Importantly, underwriters disclose the bookbuilding information of institutional investors' bids in the final prospectus (see example in Table 2, Panel B). The distribution of institutional bidding is partially revealed as bidding price ranges, providing a sketch of the stock value estimated by institutional investors. As the final prospectus is submitted before the subscription phase for retail investors, they can access the bookbuilding information of institutional investors' bids before subscribing for the IPO stock.

Once retail investors subscribe for IPO stocks in the second subscription phase, the issuer and its underwriter proportionally allocate shares for retail investors. Then, the issuer and its underwriter report the IPO share allocations to each type of investor (see Table 2, Panel C).

3 Data and Methodology

3.1 IPO data

Our initial sample includes 630 completed Korean IPOs during June 2009–December 2017 time period. Before June 2009, the required form of IPO filings differed substantially. We dropped 98 Special Purpose Acquisition Companies (SPACs) and 29 foreign firms, leaving 503 IPOs.

Using the Application Programming Interface (API) of the Data Analysis, Retrieval and Transfer System (DART) operated by the Korean Financial Supervisory Service (FSS), we downloaded the html files of each IPO's final prospectus and subscription results. Then, an automated script parses the documents' numerical and textual information. The script identifies the initial price range, the essential value, the institutional bidding information, and the offer price from the prospectus. The script also identifies the subscription behaviors of each type of investors from the security issuance statement. For missing data, we manually supplement the data. As 61 firms lack sufficient information in their prospectuses (mostly limited bookbuilding information), we omitted

and initial returns. Eom (2018) shows that institutional investors tend to bid more favorably after the IPO regulation reform in Korea which allows the issuing firm and its underwriter to have more discretion on pricing.

them, leaving 442 IPOs in our sample.[3] We obtain firm-level data from DataGuide (analogous to CRSP and Compustat in the United States) and market-level and underwriter-level data from the Korea Exchange (KRX).

3.2 Variables for textual information

To construct variables for textual information, we focus on the Risk Factors Section (hereafter, RFS) of the prospectus. RFS discusses an existing and potential risk of the firm. To compare each document in the intended context, focusing on a subset of a document is a common practice in the literature (e.g., Loughran and McDonald (2011); Li, Lundholm, and Minnis (2013); Hoberg and Maksimovic (2014)).

First, we use an automated script to identify the number of sentences in the RFS,[4] which indicates the amount of soft information and is our first textual information variable, Description Length. This measure is highly correlated with the number of words in RFS which also can proxy the amount of information ($\rho = 0.9899$). Note that when we measure the amount of soft information through the number of words in the RFS, the results are similar.

We construct Description Tone via the following procedure. In the spirit of Li (2010), we use the naive Bayes algorithm to classify each sentence into predefined sentiments with associated numerical values: positive (1), neutral (0), negative (-1), and uncertain (-1).[5] Then, we compute the mean of all sentences' sentiments in the RFS. See details of the classification method and robustness check in Appendix A.

3.3 Other Variables

Firm variables include the natural logarithm of total assets ($Size$), the number of years from firm establishment to firm listing (Age), total liabilities over total assets ($Leverage$), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$), which has a value of 1 if a firm is listed in KOSPI or 0 otherwise.

---

[3] Estimated essential values are only available from September 2010. Hence, the number of observations with essential values is 328.

[4] Figures and unidentified sentences are ignored. The automated script is available upon the request.

[5] For example, the sentence "This can be an important uncertainty in investor decision making and it is difficult to apply investment decision making techniques using systematic and statistical techniques" is coded as uncertain.

IPO specific variables include the following. The natural logarithm of underwriter IPO market share in the previous year plus one reflects the underwriter $Reputation$ (Carter (1990); Carter, Dark, and Singh (1998)). IPO market timing (Baker & Wurgler, 2002) is proxied by the average return of the KOSPI index for 90-trading days before the IPO date, ending at $t-1$ ($Market\ Return$). $\lambda_{industry}$ and λ_{year} denote dummy variables for industry and year respectively.[6]

The initial offer price band in the preliminary prospectus is the interval $[P_{inital}^{min}, P_{initial}^{max}]$. We define $Initial\ Mid\ Price = (P_{inital}^{min} + P_{initial}^{max})/2$ and $Initial\ Price\ Range\ (\%) = 100 \times (P_{inital}^{min} - P_{initial}^{max})/Initial\ Mid\ Price$.

Following Cornelli and Goldreich (2003), we construct several variables regarding institutional bidding information. $Avg.Inst.Bid$ is the quantity-weighted average bidding price. $Inst.Oversubs\ (All)$ is the sum of institutional investors' oversubscription ratio over the offer price for limit bids and their oversubscription ratio for strike bids. See details in Appendix B.1. $Retail\ Oversubs$ is a retail investors' oversubscription ratio (see details in Appendix B.2).

3.4 Descriptive Statistics

Several descriptive statistics are notable. First, issuers discount their initial prices by 25% of the essential value on average (see Table 3). Second, institutional investors' average bidding price is slightly higher (8%) than the midpoint of initial prices, and the finalized offer price is slightly lower (5%) than the average bidding price. Third, IPOs are very oversubscribed both for institutional and retail investors (about 212 times the quantity to be allocated for institutional investors and 602 times that for retial investors). Fourth, the first-day return averages 30%. Lastly, $Tone\ of\ Description$ and $Amount\ of\ Description$ are cross-sectionally well-distributed and are slightly negatively correlated ($\rho = -0.089$).

4 Empirical Results

[6] Industries are classified into 9 class variables following the Korean Standard Industrial Classification (KSIC) released by a central government organization for statistics, the Statistics Korea.

In this section, we present our empirical results. For convenient interpretation, $Tone\ of\ Description$ and $Amount\ of\ Description$ are each normalized to mean 0 and standard deviation 1. Reported standard errors are White (1980)'s HC estimator throughout the paper.

4.1 The Relationship between Textual Information and Public Information

Since prospectus text information is endogenously determined by the firm, it can be correlated with public information such as firm characteristics or market returns. If they are perfectly correlated, the prospectus text provides no additional information to investors. Hence, we investigate how much variation among textual information is explained by public information with a dependent variable of $Description\ Tone$ or $Description\ Length$:

$$y = \alpha + \beta_1 KOSPI + \beta_2 Market\ Return + \beta_3 Reputation + \beta_4 Age + \beta_5 Size + \beta_6 Leverage + \beta_7 Profitability + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.1)$$

$$y = \alpha + \beta_1 KOSPI + \beta_2 Market\ Return + \beta_3 Reputation + \beta_4 Age + \beta_5 logProceeds + \beta_6 Leverage + \beta_7 Profitability + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.2)$$

Issuers that are younger, smaller (or raise smaller proceeds) often describe themselves more positively, especially when market timing is good (see Table 4). Also, issuers that are younger, larger (or raise larger proceeds) tend to provide longer prospectus texts. Issuers with higher leverage provide longer prospectus texts only in the specification with *logProceeds*. These regressions account for substantial variance (0.222 and 0.529), suggesting that public information explains much of the differences in prospectus texts.

4.2 Textual Information and Initial Offer Prices

Past studies have examined the link between prospectus text and initial offer prices indirectly via price revisions. For example, the frequency of negative and uncertain words in the preliminary prospectus predicts upward offer price revision, which suggests that issuers negatively describe a firm's future prospects and operations to lowball initial prices before bookbuilding (Loughran and McDonald (2013)). Longer prospective texts are linked to smaller absolute sizes of offer price revision, which suggests greater accuracy of the initial prices (Hanley and Hoberg (2010)).

To examine the link between prospectus text and initial offer prices directly, we use an appropriate denominator to scale initial prices, $Essential\ Value$, which is objectively estimated by underwriters from PER or DCF valuation models. Specifically, we estimate regression models

$$y = \alpha + \beta_1 Description\ Tone + \beta_2 Description\ Length + \gamma Control + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.3)$$

where the dependent variable is $Initial\ Price\ Range$ (%) which shows how narrowly (or widely) issuers set their initial prices. In addition, $Initial\ Mid\frac{Price}{Essential}Value$ (%) is the dependent variable, reflecting the issuer's willingness to be paid in the IPO market. In all regression specifications, $Control$ includes $KOSPI$, $Market\ Return$, $Reputation$, Age, $Size$, $Leverage$, and $Profitability$. β_1 and β_2 of equation (3.3) capture the relation between prospectus text and initial prices of the IPO.

When $Description\ Tone$ is more positive or $Description\ Length$ is smaller, issuers tend to set their initial price ranges narrowly (see Table 5). Specifically, a one standard deviation greater $Description\ Tone$ is linked to a 0.62 smaller $Initial\ Price\ Range$ (%), while a one standard deviation less $Description\ Length$ is linked to a 0.71 greater $Initial\ Price\ Range$ (%). Also, $Profitability$ has a significant, negative coefficient, suggesting that profitable firms tend to have narrow initial price ranges.

A firm whose prospectus text's $Description\ Tone$ is more positive than those of other firms often has higher $Initial\ Mid\ Price/Essential\ Value$ (%). For example, a one standard deviation greater $Description\ Tone$ is linked to a 1.87 greater $Initial\ Mid\ Price/Essential\ Value$ (%). Larger firms with lower leverage than other firms often set higher initial prices. In summary, prospectus text attributes are linked to initial prices.

4.3 Textual Information and Institutional Investors' Bids

Current theoretical models do not provide clear answers on how institutional investors respond to prospectus texts, but the research literature suggests two divergent investor responses when they estimate the value of an IPO firm and the number of its shares to subscribe. On the one hand, institutional investors can extract information from the text to inform their estimates. As IPO firms

and underwriters can partially adjust institutional investors' information in their final offer price (Hanley, 1993), textual information can affect initial returns or long-run performance of stocks.

On the other hand, institutional investors may rationally ignore textual information for two reasons. One, IPO firms and underwriters may have fully incorporated information from the prospectus text to set initial offer prices before bookbuilding. Two, the institutional investors might already know the information in the prospectus text. In either of these cases, institutional investors do not utilize the prospectus text.

Following Cornelli and Goldreich (2003), we construct two variables for institutional bids. $Avg.Inst.Bid$ is a quantity-weighted average bidding price, and $Inst.Oversubs$ (All) is institutional investors' oversubscription ratio. Using each variable as a dependent variable ($Avg.Inst.Bid/Initial\ Mid\ Price$ (%) or $\log Inst.Oversubs$), we estimate the following regression models.

$$y = \alpha + \beta_1 Description\ Tone + \beta_2 Description\ Length + \beta_3 Initial\ Price\ Range\ (\%) + \gamma Control + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.4)$$

β_1 and β_2 of this equation (3.4) capture the marginal effect of prospectus text information on institutional bids.

$Avg.Inst.Bid/Initial\ Mid\ Price$ (%) is not linked to either prospectus text attribute (see Table 6, column 1). However, both $Avg.Inst.Bid$ and $Initial\ Mid\ Price$ are endogenously determined (as shown in Table 5), so regression coefficient values of their ratio are difficult to interpret. To address this issue, we decompose $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) into two parts using the log transformation

$$\log\frac{Avg.Inst.Bid}{Initial\ Mid\ Price} = \log\frac{Avg.Inst.Bid}{Essential\ Value} + \log\frac{Initial\ Mid\ Price}{Essential\ Value}$$

and conduct analysis using samples with Essential value information.

$Description\ Tone$ is only linked to $\log\frac{Initial\ Mid\ Price}{Essential\ Value}$, and Description length is not linked to initial prices, institutional investors' bids, or their oversubscription ratios. Hence, our results suggest that prospectus texts do not influence institutional investors' bids.

4.4 Price Revisions after the Bookbuilding

Next, we examine how the issuer and its underwriter adjust the offer price after gathering the bookbuilding information via the following regression model.

$$\frac{Offer\ Price}{Initial\ Mid\ Price}(\%) = \alpha + \beta_1 Description\ Tone + \beta_2 Description\ Length + \beta_3 Initial\ Price\ Range\ (\%) + \beta_4 \log(Inst.Oversubs) + \beta_5 \frac{Avg.Inst.Bid}{Initial\ Mid\ Price}(\%) + \gamma Control + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.5)$$

where $Offer\ Price$ is a finalized offer price after the bookbuilding. β_1 and β_2 of equation (3.5) capture the marginal effect of prospectus text information on price revision after incorporating institutional investors' bidding information. Like Cornelli and Goldreich (2003), we also investigate how the offer price adjusts bids after bookbuilding by examining β_4 and β_5.

Table 7 presents the regression results with offer price revisions as the dependent variable. Institutional investors' bidding information is not included in column (1), similar to Loughran and McDonald (2013). Firms with more positive *Description tones* of prospectus texts than other firms are more likely to revise their offer prices downward (see Table 7, column 1), consistent with the findings of Loughran and McDonald (2013).

Regressions with independent variables capturing information on institutional investors' bids, account for much more of the variance (adjusted R^2 = 78%, see Table 7, columns 2, 3, and 4; compared to adjusted R^2 = 6% without these independent variables, see column 1). Coefficient values for *Description tones* are consistently negative and significant. After including institutional investors' oversubscription ratio in the model, *Description length* has a significant, positive regression coefficient.

For institutional investors' bidding information, coefficient values for $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) and $\log Inst.Oversubs\ (All)$ are positive and statistically significant in all specifications. When institutional investors bid higher or subscribe for more shares of an IPO, issuers often revise their prices upward.

In summary, the bookbuilding information is related to price revisions, suggesting that issuers and underwriters use this information when they finalize the offer price. Since institutional investors' bids do not respond to a description tone as shown in Section 4.3, Description tone is negatively linked to offer price revision while description length is positively linked to it.

4.5 Retail investors' Subscription after the Offer Price Revision

After seeing a summary of institutional investors' bids and the finalized offer price, retail investors subscribe for the IPO. $Retail\ Oversubs$ is the natural logarithm of retail investors' oversubscription ratio and is the dependent variable in the following regression.

$$\log Retail\ Oversubs = \alpha + \beta_1 Tone\ of\ Description + \beta_2 Amount\ of\ Description + \beta_3 Initial\ Price\ Range\ (\%) + \beta_4 \log(Inst.Oversubs) + \beta_5 \frac{Avg.Inst.Bid}{Initial\ Mid\ Price}(\%) + \beta_6 \frac{Offer\ Price}{Initial\ Mid\ Price}(\%) + \gamma Control + \lambda_{industry} + \lambda_{year} + \epsilon \quad (3.6)$$

Note that the traditional offer price revision variable, $Offer\ Price/Initial\ Mid\ Price$, can be decomposed into two parts: how institutional investors value the IPO firm compared to initial prices ($Avg.Inst.Bid/Initial\ Mid\ Price\ (\%)$) and how issuers adjust the offer price to institutional bidding prices ($Offer\ Price/Avg.Inst.Bid\ (\%)$). β_1 and β_2 of equation (3.5) capture the marginal effects of prospectus text information on retail investors' subscription. β_4, β_5, and β_6 show how retail investors reflect bookbuilding information when they subscribe for IPO stocks.

Both *Description tone* and *Description length* are not linked to $Retail\ Oversubs$ (see Table 8). Offer price revisions ($Offer\ Price/Initial\ Mid\ Price\ (\%)$) and size are significantly linked to $Retail\ Oversubs$, so retail investors tend to subscribe for more IPO stocks of smaller firms and those with greater upward price revisions.

Next, we decompose offer price revision into two parts, how issuers adjust the offer price to bidding prices ($Offer\ Price/Avg.Inst.Bid\ (\%)$) and how institutional investors evaluate the IPO ($Avg.Inst.Bid/Initial\ Mid\ Price\ (\%)$). When institutional investors bid high, retail investors tend

to subscribe more (See Table 8, column 2). The results also show that the explanatory power of offer price revision stems more from how institutional investors evaluate the IPO rather than how issuers adjust the offer price to bidding prices.

Institutional investors' oversubscription ratio is also a strong predictor for retail investors' oversubscription ratio, accounting for substantially more variance (see Table 8, column 3: adjusted R^2 = 44% compared to 21% in column 2). In this model, *Description Length* is significantly, positively linked to retail investors' oversubscription ratio, suggesting that retail investors subscribe more when firms' prospectus texts are longer. The coefficient for $Offer\ Price/Initial\ Mid\ Price$ (%) is significant and negative.

Retail investors seem to think that institutional investors' high oversubscription ratio has more information, compared to bidding prices (see Table 8, column 4). As they put more emphasis on oversubscription ratio of institutional investors, $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) yields a negative impact on the subscription when it is included with oversubscription ratio. The coefficient for $Offer\ Price/Avg.Inst.Bid$ (%) is negative and significant, suggesting that retail investors tend to subscribe less when the offer price relatively high compared to the average institutional bidding price.

In summary, when institutional investors subscribe more, retail investors subscribe more. Also, prospectus text description tone is not linked to retail investor subscription, while description length's link to it is not robust.

4.6 Short-term and Long-term Performance of IPO Stocks

Finally, we examine whether prospectus texts or bookbuilding information are linked to short-term or long-term performance of IPO stocks.

$$y = \alpha + \beta_1 Description\ Tone + \beta_2 Description\ Length + \beta_3 Initial\ Price\ Range\ (\%) + \beta_4 \log Inst.Oversubs + \beta_5 \frac{Avg.Inst.Bid}{Initial\ Mid\ Price}(\%) + \beta_6 \frac{Offer\ Price}{Initial\ Mid\ Price}(\%) + \beta_7 \log Retail\ Oversubs + \gamma Control + \lambda_{industry} + \lambda_{year} + \epsilon. \quad (3.7)$$

$Initial\ Return_t$ is $100 \times (P_t\ - Offer\ Price)/Offer\ Price$ where P_t is the stock price of the IPO firm at t trading days after the IPO date.

For the long-run return, the dependent variable is either $Market-to-book_t$ or $BHAR[0,T]$. $Market-to-book_t$ is a market-to-book ratio of equity t months after the IPO and $BHAR[0,T]$ is a buy-and-hold abnormal return of the stock at T trading days after the IPO defined as $\prod_{t=0}^{T}(1+R_t) - \prod_{t=0}^{T}(1+R_{m,t})$ where R_t is daily stock return of the IPO firm and $R_{m,t}$ is daily market index return at t trading days after the IPO. β_1 and β_2 of equation (3.6) captures the marginal effect of prospectus text information on stock performances. Unlike past studies (e.g., Loughran and McDonald (2013)), the bidding and subscription behaviors of investors are public information in this study. We capture the relation between initial returns and bidding and subscription behaviors of investors in β_4, β_5, β_6, and β_7.

While the models in column (1), (3), and (5) include offer price revision as covariates, those in (2), (4), and (6) decompose it into two parts and include institutional investors' bidding information and retail investors' subscription information (see Table 9). Column (1), (3), and (5) show that offer price revision is a strong predictor of initial returns. For example, a one percent greater $Offer\ Price/Initial\ Mid\ Price$ (%) is linked to a 0.848% greater first-day return (see Table 9, column 1). This pattern is consistent with the partial adjustment phenomenon in the literature (Hanley, 1993).

When $Offer\ Price/Initial\ Mid\ Price$ (%) is decomposed into two parts in column (2), (4), and (6), $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) has significant, positive coefficients, consistent with Cornelli and Goldreich (2003). Moreover, retail investors' oversubscription ratio is a strong predictor of initial returns, while institutional investors' oversubscription ratio is not.

In all specifications, prospectus text information does not predict initial returns. In the short-run however, the stock price after the IPO may deviate from the fundamental value due to investor sentiment (Derrien (2005); Cornelli et al. (2006)). For this reason, the value of textual information may not be reflected for the short-run.

Table 10 presents the regression results on the long-term performance of the IPO stocks. The dependent variable in column (1) and (2) is a market-to-book ratio of equity, while that in column (3), (4), and (5) is a buy-and-hold abnormal return. The coefficient for $Avg.Inst.Bid/Initial\ Mid\ Price\ (\%)$ is positive and significant in 4 specifications, suggesting that institutional investors' bidding prices have a predictive power even for long-run stock returns. In contrast, retail investors' oversubscription ratio, a strong predictor for initial returns, is not signficant in all specifications. This pattern is consistent with Derrien (2005) and Cornelli et al. (2006), who suggest retail investors' optimism can raise the IPO stock price in the short-run, followed by long-run reversals.

Prospectus text information weakly predicts long-run performances of IPO stock., both Description Tone and Descriptive Length are positively and significantly linked to market-to-book ratio at 6 months after the IPO but not 12 months after the IPO. Descriptive Length is positively and significantly linked to buy-and-hold abnormal return until 120 days after the IPO date.

In summary, our findings show that even though textual information has no predictive power for returns in the short-run, it predicts returns in the long-run, suggesting that it can be useful for institutional investors in the long-run.

5 Conclusion

Using the unique institutional setting of the IPO in Korea, we test whether prospectus text information affects institutional investors' bids in bookbuilding. As institutional investors are considered informed investors, analyzing their response to prospectus texts can indicate how informative they are.

Our findings show that prospectus text information is linked to initial offer prices, offer price revision after bookbuilding, and long-term stock outcomes, but not institutional investors' bids or short-term stock returns. Hence, prospectus texts have information value for institutional investors.

As institutional investors' bidding prices predict IPO returns, this result implies that retail investors can benefit from an investment strategy that mimics institutional investors' bids. In fact, retail investors utilize institutional investors' public bid information, but not prospectus text

information. The subscription behaviors of retail investors may reflect the cost of processing information, as extracting information from bidding prices is typically easier than reading and interpreting hundreds pages of the prospectus. So, our findings imply that public disclosure of the bookbuilding information can be a useful investor protection tool that benefits retail investors and reduces uninformed investors' information disadvantage.

Appendix:

A.1. Details in the classification method

To measure description tone, we employ the naive Bayes algorithm. This supervised machine learning model is frequently used in the accounting and finance literature on the text classification (e.g., Li (2010); Buehlmaier and Whited (2018)) and longstanding in computational linguistics (Hastie, Friedman, & Tibshirani, 2001). Moreover, without an established dictionary such as Loughran and McDonald (2011)'s, the dictionary approach which counts predefined target words in a document is hard to implement.

The first step of our approach is treating a sentence as a bag of words: it ignores structure and sequence of words in a sentence but just counts the frequency of words in the sentence. In this sense, two distinguished sentences, v_1= "Tom loves Emma" and v_2= "Emma loves Tom", are considered as the same bag of words, $\omega = [Tom: 1, loves: 1, Emma: 1]'$. In general, if we have n sentences and J distinct words to consider, a bag of words approach is a function g which transforms a sentence v_i into a J-vector ω_i.

The second step is a classification. Taking $g(v_i) = \omega_i \in \Omega$ as an input, the text classifier in this paper can be defined as a function

$$f: \Omega \to S \; where \; S = \{positive, neutral, negative, uncertain\}$$

which conducts the classification by maximization of the conditional probability

$$\arg\max_{s \in S} P(s|\omega_i). \; (A.1)$$

The naive algorithm is based on the naive assumption that frequencies of words, ω_{ij} for $j = 1, \dots, J$ are independently distributed. Hence, by the Bayes rule, equation (A.1) can be rewritten as

$$\arg\max_{s \in S} P(\omega_i|s) \cdot \frac{P(s)}{P(\omega_i)} = \prod_j P(\omega_{ij}|s) \cdot P(s). \; (A.2)$$

Finally, we estimate $\hat{P}(\omega_{ij}|s)$ and $\hat{P}(s)$ as sample analog in training sample.

To conduct the naive Bayes algorithm, it is necessary to construct appropriate training data. Among about 90,000 identified sentences in our sample, we randomly extract 10,000 sentences and manually read and classify each sentence into four tones. We remove every punctuation in sentences

and tokenize each sentence into the bag of words. To reduce dimensionality, we limit the number of distinct words in sentences to 20,000 in order of frequency, hence rarely appearing words are ignored in the model. Then, we train our classification model with training data as described above and classify remaining tokenized sentences with this model.

A.2 Cross-validation test on the classification model

To test the out-of-sample performance of our classification model, we conduct the 5-fold cross-validation test. First, we equally split our training sample into five subsets. Then, we pick one subset as a test sample and train the classification model using the remaining four subsets. To confirm the performance of the model in out-of-sample, we calculate the accuracy rate of classification in a test sample. Among four sentiments predefined, we treat negative and uncertain as same when calculating accuracy rates. We repeat the above procedure five times by picking another subset as a test sample.

Table A1 shows accuracy rates from 5-fold cross-validation test. An average of accuracy rates on out-of-sample is evaluated as 75.85%, which is high enough to accept the validity of the model.

A.3 Robustness test on the description tone

Since our measure can be subjective, we additionally employ two independent coders, coder A and coder B, to construct different training sample and estimate the tone measure using the classification model trained by their classification. Coder A is doing her Master's in Finance at Seoul National University. Coder B is doing her undergraduate study in Business Administration at Seoul National University. Both coders are native Korean.

Table A2 shows the correlation of tones between coders. Even though three coders have heterogenous academic and social background, correlation as high as 0.87 suggests our measure is consistent.

Appendix B

We illustrate how we construct variables for institutional and retail investors based on the example in Table 2.

B.1 Variables for institutional investors

As seen in Panel B of Table 2, the issuer and its underwriter disclose bidding information with n bidding price ranges. For bidding price range i among ranges except the range for strike bids, the disclosure presents the minimum and the maximum of bidding prices for the range ($[P_i^{min}, P_i^{max}]$), the quantity of bids for the range (Q_i), but not the actual distribution of bidding within the range. We assume that the quantity of bids Q_i in bidding price range

$$Avg.Inst.Bid = \frac{\sum_i Q_i \cdot \overline{P_i}}{\sum_i Q_i}$$

where $\overline{P_i} = (P_i^{min} + P_i^{max})/2$ for bidding price range i for $i \notin the\ range\ for\ strike\ bids$. However, the main obstacle of our assumption is to treat the highest range and the lowest range. Since $(P^{min} + \infty)/2$ is not defined for the highest range and it is unrealistic to assume that bids are uniformly distributed within $[0, P^{max}]$ for the lowest range. For the convenience, we transform the highest range from $[P^{min}, \infty]$to $[P^{min}, P^{min}]$ and the lowest range from $[0, P^{max}]$ to $[P^{max}, P^{max}]$.

The assumption of uniform distribution is also applied for the institutional investors' oversubscription ratio, when the offer price $P^{\text{offer}} \in (P_i^{min}, P_i^{max})$ for some range i. Specifically, we define

$$Inst.Oversubs\ (Limit) = \frac{\sum_i Q_i^*}{Q_{Inst.\ Initial\ Assignment}} \quad (B.2)$$

where

$$Q_i^* = \begin{cases} Q_i & \text{if } P^{\text{offer}} \le P_i^{min} \\ Q_i \times \dfrac{P_i^{max} - P^{\text{offer}}}{P_i^{max} - P_i^{min}} & \text{if } P^{\text{offer}} \in (P_i^{min}, P_i^{max}) \\ 0 & \text{otherwise} \end{cases}$$

for $i \notin the\ range\ for\ strike\ bids$ and $Q_{Inst.\ Initial\ Assignment}$ is the number of the IPO shares assigned to institutional investors. We also define

$$Inst.Oversubs\ (Strike) = \frac{Q_{i \in the\ range\ for\ strike\ bids}}{Q_{Inst.\ Initial\ Assignment}}\ (B.3)$$

$$Inst.Oversubs\ (Strike) = Inst.Oversubs\ (Limit) + Inst.Oversubs\ (Strike).(B.4)$$

B.2 Variable for retail investors

As seen in Panel C of Table 2, we do not observe a direct measure for subscription of retail investors except few cases. Hence, we need to calculate the implied subscription ratio of retail investors. We define

$$Retail\ Oversubs = \frac{Q_{Inst.\ and\ Retail\ Subs} - \sum_i Q_i}{Q_{Retail\ Initial\ Assignment}}\ (B.4)$$

where $\boldsymbol{Q_{Inst.\ Retail\ Subs}}$ is the total number of shares subscribed by institutional, $\boldsymbol{Q_i}$ is the quantity of bids in bidding price range $\boldsymbol{i}$ for $\boldsymbol{i \in all\ bidding\ price\ ranges}$, and retail investors and $\boldsymbol{Q_{Retail\ Initial\ Assignment}}$ is the number of the IPO shares assigned to retail investors. However, like our example, there are cases where $\boldsymbol{Q_{Inst.\ and\ Retail\ Subs} \leq \sum_i Q_i}$. We replace $\boldsymbol{\sum_i Q_i}$ with $\boldsymbol{Q_{Inst.\ and\ Retail\ Subs}}$ in those cases.

References

Baker, M., & Wurgler, J. (2002). Market Timing and Capital Structure. *The Journal of Finance, 57*(1), 1--32.

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

Benveniste, L. M., & Wilhelm, W. J. (1990). A comparative analysis of IPO proceeds under alternative regulatory environments. *Journal of Financial Economics, 28*(1-2), 173--207.

Buehlmaier, M. M. M., & Whited, T. M. (2018). Are Financial Constraints Priced? Evidence from Textual Analysis. *Review of Financial Studies, 55*, 2219--2236.

Carter, R. B. (1990). Initial Public Offerings and Underwriter Reputation. *The Journal of Finance, 45*(4), 1045--1067.

Carter, R. B., Dark, F. H., & Singh, A. K. (1998). Underwriter Reputation, Initial Returns, and the Long-Run Performance of IPO Stocks. *The Journal of Finance, 53*(1), 285--311.

Cornelli, F., & Goldreich, D. (2001). Bookbuilding and Strategic Allocation. *The Journal of Finance, 56*(6), 2337--2369.

Cornelli, F., & Goldreich, D. (2003). Bookbuilding: How Informative Is the Order Book? *The Journal of Finance, 58*(4), 1415--1443.

Cornelli, F., Goldreich, D., & Ljungqvist, A. (2006). Investor Sentiment and Pre-IPO Markets. *The Journal of Finance, 61*(3), 1187--1216.

Derrien, F. (2005). IPO Pricing in Hot Market Conditions: Who Leaves Money on the Table? *The Journal of Finance, 60*(1), 487--521.

Eom, C. (2018). Institutional bidding behaviors during IPO bookbuilding: Evidence from Korea. *Journal of Corporate Finance, 48*, 1--15.

Hanley, K. W. (1993). The underpricing of initial public offerings and the partial adjustment phenomenon. *Journal of Financial Economics, 34*(2), 231--250.

Hanley, K. W., & Hoberg, G. (2010). The Information Content of IPO Prospectuses. *Review of Financial Studies, 23*(7), 2821--2864.

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

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The Elements of Statistical Learning*.

Hoberg, G., & Maksimovic, V. (2014). Redefining Financial Constraints: A Text-Based Analysis. *Review of Financial Studies, 28*(5), 1312--1352.

Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone Management. *The Accounting Review, 89*(3), 1083--1113.

Joh, S. W., & Kim, Y.-H. (2017). Effects of Institutional Investors \textquoteright Bidding Information on Offer Prices and Initial Returns of IPOs. *Asia-Pacific Journal of Financial Studies, 46*(1), 116--154.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221--247.

Li, F. (2010). The Information Content of Forward-Looking Statements in Corporate Filings-A Na \ \i ve Bayesian Machine Learning Approach. *Journal of Accounting Research, 48*(5), 1049--1102.

Li, F., Lundholm, R., & Minnis, M. (2013). A Measure of Competition Based on 10-K Filings. *Journal of Accounting Research, 51*(2), 399--436.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35--65.

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

Rogers, J. L., Van Buskirk, A., & Zechman, S. L. C. (2011). Disclosure Tone and Shareholder Litigation. *The Accounting Review, 86*(6), 2155--2183.

Spatt, C., & Srivastava, S. (1991). Preplay Communication, Participation Restrictions, and Efficiency in Initial Public Offerings. *The Review of Financial Studies, 4*(4), 709--726.

Teoh, S. H., Welch, I., & Wong, T. J. (1998). Earnings Management and the Long-Run Market Performance of Initial Public Offerings. *The Journal of Finance, 53*(6), 1935--1974.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817.

Figure 1
IPO process in Korea

	Player	Action	Publicly Revealed Information
I	• Issuing firm • Underwriter	Filling the preliminary prospectus	1. Hard information a. Past records b. Financial statements c. Estimated essential value d. Initial price range 2. Soft information a. Future opportunities b. Risks
		↓	
II	• Institutional investors	Bidding for the IPO stocks	
		↓	
III	• Issuing firm • Underwirter	Filing the final prospectus	1. Bidding price and quantity 2. Final issue price
		↓	
IV	• Retail investors	Subscribing for the IPO stocks	
		↓	
V	• Issuing firm • Underwriter	1. Allcating shares a. Discretionarily for the institutional b. Proportionally for the retal 2. Filing the security issuance statement	Subscripti 1. on of each type of investors 2. Allcation of shares for each types of investors

Table 2
Disclosure of Hyundai Rotem (KRX:064350)

Panel A: Description in the Risk Factors Section (Abbreviated)

1. Business Risk
(a) Our primary customers are the government and related agencies, and the loss of a major customer or a reduction in revenue from a particular customer can have a negative impact on out business…
…
(b) Due to the cyclical nature of the railway and plant indisturies in which we operate, our business activities may be subject to significant volatility. …
…
2. Firm-specific Risk
(a) …
…
(d) Our business performance may be affected by exchange rate fluctuations. …
…

Panel B: Bidding information

Bidding Prie Range (KRW)	N. of Bidding	Q. of Bidding
23,000–	227	719,143,395
22,000–23,000	11	5,313,292
21,000–22,000	32	29,475,663
20,000–21,000	68	47,970,958
19,000–20,000	20	6,581,891
18,000–19,000	16	5,309,345
17,000–18,000	16	5,335,852
0–17,000	2	1,888,176
Strike bids	65	122,273,133

Panel C: Subscription information (Abbreviated)

	Initial Assignment		Subscription		Allocation	
	Q	Ratio	Q	Ratio	Q	Ratio
The ESOP Association	5,412,000	20.0	5,412,000	3.5	5,412,000	20.0
Inst. and Retial Investors	21,648,000	80.0	311,383,620	96.5	21,648,000	80.0

Table 3
Descriptive Statistic

442 Korean IPOs had sufficient bookbuilding information in the final prospectus during 2009–2017. Essential value is only available in the prospectus since September 2010, hence, the number of observations including the essential value is 328. Description tone and Description length are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. See Subsection 3.3 and Appendix B for other variables.

Panel A: Distribution

	N	Mean	St. Dev.	Min	Q1	Q3	Max
Tone	442	-0.17	0.12	-0.57	-0.26	-0.08	0.18
N. of Sentences	442	181.29	105.37	30	105.2	221	732
KOSPI	442	0.19	0.39	0	0	0	1
Age	442	14.70	9.99	0.88	9.01	16.77	67.25
Size	442	17.97	1.44	15.46	17.04	18.57	24.98
Leverage	442	0.50	0.26	0.03	0.34	0.63	2.90
Profitability	442	0.14	0.15	-0.66	0.08	0.21	0.79
Initial Range	442	0.17	0.05	0.04	0.13	0.20	0.48
Proceeds (USD) (1USD = 1,000KRW)	442	71,050.35	218,153.90	2,240	10,027.4	40,185.6	2,661,717
Initial Mid Price/ Essential Value	328	0.75	0.10	0.43	0.69	0.83	1.01
Avg. Bid Price/ Initial Mid Price	442	1.08	0.14	0.56	1.01	1.16	1.81
Inst. Oversubs. Ratio (Limit)	442	187.39	193.31	1.27	36.92	301.93	833.02
Inst. Oversubs. Ratio (Strike)	442	24.79	42.23	0.00	0.00	39.31	298.44
Inst. Oversubs. Ratio (All)	442	212.18	207.29	1.27	48.36	318.52	941.25
Offer Price/Avg. Bid Price	442	0.95	0.08	0.59	0.90	1.00	1.24
Retail Oversubs	442	602.28	2,065.05	0.27	70.43	740.55	42,424.87
Initial Return	442	0.30	0.43	-0.33	-0.03	0.54	1.60

Panel B: Correlation

	Description							
	Tone	Length	KOSPI	Age	Size	Leverage	Profitability	Proceeds
Tone	1	-0.098	-0.367	-0.210	-0.383	-0.062	0.050	-0.268
Length		1	0.215	0.054	0.369	0.203	-0.172	0.429
KOSPI			1	0.288	0.709	0.132	-0.134	0.406
Age				1	0.331	-0.015	-0.158	0.105
Size					1	0.232	-0.115	0.578
Leverage						1	-0.432	0.052
Profitability							1	-0.067
Proceeds								1

Table 4
Determinants of Textual Information

All regressions include 442 Korean IPO observations during 2009-2017. The dependent variable in first two columns is a normalized tone of sentences in the Risk Factors Section of the prospectus (Description tone). The dependent in last two column is a normalized number of sentences in the Risk Factors Section of the prospectus (Description length). See Appendix A for details in textual variables. In column (1) and (3), we control for several firm characteristic variables: the natural logarithm of total assets (Size), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets (Leverage), operating income scaled by total assets (Profitability), and KOSPI dummy (KOSPI) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. Column (2) and (4) replace Size with logProceeds. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year (Reputation) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ (Market Return). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis.

	Description tone		Description length	
	(1)	(2)	(3)	(4)
KOSPI	-0.383**	-0.433***	0.082	0.170
	(0.160)	(0.150)	(0.166)	(0.161)
Market timing	1.652**	1.569**	0.422	0.544
	(0.718)	(0.710)	(0.353)	(0.344)
Reputation	-0.014	-0.012	-0.012	-0.015
	(0.028)	(0.027)	(0.036)	(0.036)
Age	-0.009**	-0.013***	-0.008**	-0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Size	-0.166***		0.244***	
	(0.044)		(0.050)	
Log(Proceeds)		-0.172***		0.245***
		(0.047)		(0.047)
Leverage	0.016	-0.080	0.146	0.289**
	(0.182)	(0.201)	(0.141)	(0.139)
Profitability	0.013	0.123	-0.215	-0.369
	(0.304)	(0.303)	(0.288)	(0.292)
Constant	1.593**	1.511*	-5.042***	-4.812***
	(0.784)	(0.775)	(0.814)	(0.737)
Year dummies	Y	Y	Y	Y
Industry-fixed	Y	Y	Y	Y
N	442	442	442	442
Adjusted R^2	0.222	0.222	0.529	0.525

*** p < .01, ** p < .05, * p <.10

Table 5
The IPO Information and Initial Prices

The dependent variable in column (1) is $100 \times (P_{initial}^{min} - P_{initial}^{max})/\text{Initial Mid Price}$ where Initial Mid Price is defined as $(P_{initial}^{min}+P_{initial}^{max})/2$ for initial prices in the preliminary prospectus denoted as the initial interval $[(P_{initial}^{min}, P_{initial}^{max}]$. The dependent variable in column (2) is $\text{Initial Mid Price}/\text{Essential Value}$ (%). Essential Value denotes underwriters' estimation for the IPO firm value by employing the price multiple method (e.g., Price Earnings Ratio (PER) valuation) or the Discounted Cash Flow (DCF) method. Description tone and Description length are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. We control for several firm characteristic variables: the natural logarithm of total assets (Size), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets (Leverage), operating income scaled by total assets (Profitability), and KOSPI dummy (KOSPI) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year (Reputation) and an average returns of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ (Market Return). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis. The number of observations decreases from 442 in column (1) to 328 in column (2), since Essential value is only available since September 2010.

	Initial Price Range (%)	Initial Mid Price/Essential Value (%)
	(1)	(2)
Description tone	-0.620** (0.294)	1.868*** (0.592)
Description length	0.714** (0.318)	-0.919 (0.850)
KOPSI	0.868 (1.041)	0.613 (2.013)
Market timing	-3.235 (4.273)	9.656 (9.041)
Reputation	-0.054 (0.131)	-0.657 (6.516)
Age	-0.026 (0.027)	-0.100 (0.066)
Size	-0.328 (0.296)	1.984*** (0.626)
Leverage	-0.842 (1.062)	-5.500** (2.687)
Profitability	-7.093*** (2.341)	-5.521 (4.221)
Constant	26.177*** (5.252)	56.093*** (11.117)
Year dummies	Y	Y
Industry-fixed	Y	Y
N	442	328
Adjusted R^2	0.036	0.121

*** p < .01, ** p < .05, * p <.10

Table 6
The IPO Information, Initial Prices, and Institution's Behavior

The dependent variable in column (1) is the level of quantity-weighted average institutional investors bidding price to $Initial\ Mid\ Price$ where $Initial\ Mid\ Price$ is defined as $(P^{min}_{\text{initial}}+P^{max}_{\text{initial}})/2$ for initial prices in the preliminary prospectus denoted as the interval $\left[(P^{min}_{\text{initial}}, P^{max}_{\text{initial}}\right]$. Using $\log Avg.Inst.Bid/Initial\ Mid\ Price = \log Avg.Inst.Bid/Essential\ Value - \log Initial\ Mid\ Price/Essential\ Value$ (%), we run regressions of $\log Avg.Inst.Bid/Essential\ Value$, $\log Initial\ Mid\ Price/Essential\ Value$ (%), and $\log Avg.Inst.Bid/Initial\ Mid\ Price$ as a dependent variable in column (2), (3), and (4) respectively. $Essential\ Value$ denotes underwriters' estimation for the IPO firm value by employing the price multiple method (e.g., Price Earnings Ratio (PER) valuation) or the Discounted Cash Flow (DCF) method. The dependent variable in column (5) is the natural logarithm of institutional investors' oversubscription ratio for limit and strike bids. $Tone\ of\ Description$ and $Amount\ of\ Description$ are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. $Initial\ Price\ Range$ (%) denotes $100 \times (P^{min}_{\text{initial}} - P^{max}_{\text{initial}})/Initial\ mid\ Price$. We control for several initial initial firm characteristic variables: the natural logarithm of total assets ($Size$), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets ($Leverage$), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year ($Reputation$) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ ($Market\ Return$). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis. Regressions in column (1) and (5) include 442 Korean IPO observations during 2009–2017. The number of observations decreases to 328 in column (2), (3), and (4), since Essential value is only available since September 2010.

	Avg. Inst. Bid/Initial Mid Price (%)	log(Avg. Inst. Bid/Essential Value (%))	log(Initial Mid Price/Essential Value (%))	log(Avg. Inst. Bid/Initial Mid Price (%))	log(Inst. Oversubs (All))
	(1)	(2)	(3)	(4)	(5)
Description tone	-0.880 (0.757)	0.016 (0.011)	0.026*** (0.009)	-0.011 (0.008)	0.053 (0.061)
Description length	0.642 (0.887)	-0.011 (0.015)	-0.015 (0.012)	0.003 (0.010)	-0.034 (0.089)
Initial Price Range (%)	0.047 (0.129)	0.0001 (0.002)	-0.002 (0.002)	0.002 (0.001)	-0.021** (0.009)
KOSPI	-5.264** (2.224)	-0.028 (0.033)	0.012 (0.028)	-0.040 (0.026)	-0.074 (0.214)
Market Timing	1.556 (9.496)	0.124 (0.134)	0.138 (0.136)	-0.013 (0.100)	0.950 (0.692)
Reputation	0.401 (0.265)	0.163 (0.105)	-0.010 (0.094)	0.172** (0.081)	0.026 (0.034)
Age	-0.062 (0.073)	-0.001 (0.001)	-0.002 (0.001)	0.0002 (0.001)	-0.006 (0.006)
Size	-0.564 (0.747)	0.010 (0.011)	0.029*** (0.009)	-0.018** (0.008)	-0.144** (0.068)
Leverage	-0.525 (3.217)	-0.115*** (0.041)	-0.094** (0.043)	-0.021 (0.031)	-0.263 (0.292)
Profitability	1.695 (6.935)	-0.132* (0.075)	-0.111* (0.062)	-0.022 (0.055)	0.187 (0.525)
Constant	87.700*** (13.931)	4.220*** (0.193)	4.077*** (0.159)	4.748*** (0.141)	4.104*** (1.231)
Year dummies	Y	Y	Y	Y	Y
Industry-fixed	Y	Y	Y	Y	Y

N	442	328	328	328	442
Adjusted R^2	0.108	0.055	0.120	0.082	0.270

*** p < .01, ** p < .05, *p <.10

Table 7
Bookbuilding Information and Price Revision

All regressions include 442 Korean IPO observations during 2009–2017. The dependent variable is a level of the finalized offer price to $Initial\ Mid\ Price$ where $Initial\ Mid\ Price$ is defined as $(P^{min}_{\text{initial}}+P^{max}_{\text{initial}})/2$ for initial prices in the preliminary prospectus denoted as the interval $\left[(P^{min}_{\text{initial}}, P^{max}_{\text{initial}}\right]$. $Tone\ of\ Description$ and $Amount\ of\ Description$ are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. $Initial\ Price\ Range$ (%) denotes $100 \times (P^{min}_{\text{initial}} - P^{max}_{\text{initial}})/Initial\ mid\ Price$. For institutional investors' bidding information, $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) denotes the level of quantity-weighted average institutional investors bidding price to $Initial\ Mid\ Price$ and $\log Inst.Oversubs\ (All)$ denotes the natural logarithm of institutional investors' oversubscription ratio for limit and strike bids. We control for several firm characteristic variables: the natural logarithm of total assets ($Size$), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets ($Leverage$), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year (Reputation) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ ($Market\ Return$). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis.

	Issue Price/Initial Mid Price (%)			
	(1)	(2)	(3)	(4)
Description tone	-1.708** (0.759)	-0.945** (0.448)	-2.228*** (0.459)	-1.471*** (0.396)
Description length	1.381 (1.056)	0.825 (0.591)	1.718** (0.701)	1.187** (0.538)
Initial Range (%)	-0.028 (0.126)	-0.069 (0.090)	0.177** (0.085)	0.048 (0.077)
Avg. Inst. Bid/Initial Mid Price (%)		0.867*** (0.042)		0.563*** (0.050)
log(Inst. Oversubs (All))			9.885*** (0.477)	4.914*** (0.544)
KOSPI	-4.614* (2.519)	-0.049 (1.534)	-3.887** (1.625)	-1.288 (1.417)
Market Timing	11.686 (8.343)	10.337* (6.193)	2.291 (5.193)	6.140 (4.887)
Reputation	0.579* (0.301)	0.232* (0.131)	0.326* (0.174)	0.227* (0.118)
Age	-0.061 (0.077)	-0.008 (0.039)	-0.0003 (0.050)	0.004 (0.035)
Size	-0.149 (0.860)	0.340 (0.476)	1.273** (0.515)	0.876** (0.423)
Leverage	-2.018 (3.741)	-1.563 (2.037)	0.580 (2.288)	-0.431 (1.904)
Profitability	2.898 (7.563)	1.428 (3.337)	1.050 (3.735)	1.024 (2.683)
Constant	59.715*** (15.347)	-16.336* (8.697)	19.142** (9.559)	-9.847 (7.388)
Year dummies	Y	Y	Y	Y
Industry-fixed	Y	Y	Y	Y
N	442	442	442	442
Adjusted R^2	0.058	0.710	0.652	0.777

*** p < .01, ** p < .05, *p <.10

Table 8
The IPO Information, Bidding Information, and Retail Investor's Behavior

All regressions include 442 Korean IPO observations during 2009–2017. The dependent variable is the natural logarithm of retail investors' oversubscription ratio. $Tone\ of\ Description$ and $Amount\ of\ Description$ are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. $Initial\ Mid\ Price$ is defined as $(P^{min}_{\text{initial}}+P^{max}_{\text{initial}})/2$ for initial prices in the preliminary prospectus denoted as the interval $\left[(P^{min}_{\text{initial}}, P^{max}_{\text{initial}}\right]$ and $Initial\ Price\ Range$ (%) is defnied as $100\times(P^{min}_{\text{initial}}-P^{max}_{\text{initial}})/Initial\ mid\ Price$. $Offer\ Price/Initial\ Mid\ Price$ (%) and $Offer\ Price/Avg.Inst.Bid$ (%) are the level of the finalized offer price to $Initial\ Mid\ Price$ and $Avg.Inst.Bid$ respectively, where $Avg.Inst.Bid$ is quantity-weighted average institutional investors bidding price. For institutional investors' bidding information, $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) denotes the level of $Avg.Inst.Bid$ to $Initial\ Mid\ Price$ and $\log Inst.Oversubs\ (All)$ denotes the natural logarithm of institutional investors' oversubscription ratio for limit and strike bids. We control for several firm characteristic variables: the natural logarithm of total assets ($Size$), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets (Leverage), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year ($Reputation$) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ ($Market\ Return$). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis.

	log(Retail Oversubs)			
	(1)	(2)	(3)	(4)
Description tone	0.076 (0.097)	0.050 (0.096)	-0.149* (0.085)	-0.159* (0.085)
Description length	0.051 (0.134)	0.077 (0.130)	0.221** (0.099)	0.225** (0.097)
Initial Range (%)	-0.008 (0.018)	-0.012 (0.017)	0.018 (0.015)	0.017 (0.015)
Issue Price/Initial Mid Price (%)	0.053*** (0.007)		-0.036*** (0.008)	
Issue Price/Avg. Inst. Bid (%)		0.013 (0.012)		-0.053*** (0.011)
Avg. Inst. Bid/Initial Mid Price (%)		0.060*** (0.007)		-0.027*** (0.007)
log(Inst. Oversubs (All))			1.389*** (0.103)	1.365*** (0.101)
KOSPI	0.051 (0.350)	0.126 (0.340)	-0.256 (0.296)	-0.245 (0.296)
Market Timing	0.403 (1.053)	0.824 (1.035)	0.120 (0.827)	0.215 (0.818)
Reputation	-0.003 (0.054)	0.001 (0.054)	0.013 (0.035)	0.015 (0.035)
Age	0.006 (0.009)	0.006 (0.009)	0.009 (0.008)	0.010 (0.008)
Size	-0.409*** (0.096)	-0.388*** (0.092)	-0.222*** (0.073)	-0.215*** (0.072)
Leverage	-0.199 (0.408)	-0.259 (0.382)	-0.013 (0.277)	-0.026 (0.269)
Profitability	-0.172 (0.661)	-0.148 (0.678)	-0.174 (0.587)	-0.112 (0.589)
Constant	8.929*** (1.832)	5.941*** (2.119)	8.525*** (1.387)	12.476*** (1.687)
Year dummies	Y	Y	Y	Y

Industry-fixed	Y	Y	Y	Y
N	442	442	442	442
Adjusted R^2	0.209	0.233	0.440	0.448

*** p < .01, ** p < .05, *p <.10

Table 9
Initial Returns

All regressions include 442 Korean IPO observations during 2009–2017. Dependent variables are $100 \times (P_t - Offer\ Price)/Offer\ Price$ where P_t is a price of IPO stocks at t trading days after the IPO date. t is 0, 3, and 7 for column (1) and (2), (3) and (4), and (5) and (6) respectively. $Tone\ of\ Description$ and $Amount\ of\ Description$ are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. $Initial\ Price\ Range$ (%) is defnied as $100 \times (P^{min}_{\text{initial}} - P^{max}_{\text{initial}})/Initial\ mid\ Price$ where $Initial\ Mid\ Price$ is defined as $(P^{min}_{\text{initial}} + P^{max}_{\text{initial}})/2$ for initial prices in the preliminary prospectus denoted as the initial interval $\left[(P^{min}_{\text{initial}}, P^{max}_{\text{initial}}\right]$. $Offer\ Price/Initial\ Mid\ Price$ (%) and $Offer\ Price/Avg.Inst.Bid$ (%) are the level of the finalized offer price to $Initial\ Mid\ Price$ and $Avg.Inst.Bid$ respectively, where $Avg.Inst.Bid$ is quantity-weighted average institutional investors bidding price. For institutional investors' bidding information, $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) denotes the level $of\ Avg.Inst.Bid$ to $Initial\ Mid\ Price$ and $\log Inst.Oversubs\ (All)$ denotes the natural logarithm of institutional investors' oversubscription ratio for limit and strike bids. $\log Retail\ Oversubs$ is the natural logarithm of retail investors' oversubscription ratio. We control for several firm characteristic variables: the natural logarithm of total assets ($Size$), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets ($Leverage$), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year ($Reputation$) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ ($Market\ Return$). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis.

	Initial Return (%)		Return [0,3] (%)		Return [0,7] (%)	
	(1)	(2)	(3)	(4)	(5)	(6)
Description tone	-0.035	-0.868	-1.153	-2.254	-1.456	-2.655
	(2.175)	(2.111)	(2.382)	(2.336)	(3.022)	(3.016)
Description length	2.446	2.567	5.598	6.022	6.215	6.746
	(2.685)	(2.642)	(3.870)	(3.845)	(5.146)	(5.135)
Initial Range (%)	-0.073	-0.090	0.185	0.156	-0.198	-0.234
	(0.348)	(0.315)	(0.460)	(0.411)	(0.525)	(0.485)
Issue Price/Initial Mid Price (%)	0.848***		1.016***		0.998***	
	(0.141)		(0.198)		(0.211)	
Issue Price/Avg. Inst. Bid (%)		-0.152		-0.290		-0.435
		(0.275)		(0.354)		(0.396)
Avg. Inst. Bid/Initial Mid Price (%)		0.768***		0.998***		1.027***
		(0.222)		(0.317)		(0.343)
log(Inst. Oversubs (All))		0.008		0.872		1.172
		(2.765)		(3.318)		(3.726)
log(Retail Oversubs)		5.291***		4.548***		4.229***
		(0.922)		(0.998)		(1.173)
KOSPI	6.867	7.999	12.842	14.298*	18.311**	19.933**
	(6.872)	(6.273)	(8.156)	(7.441)	(9.014)	(8.329)
Market Timing	63.070**	67.626***	58.314**	65.584**	50.554	58.831*
	(26.736)	(25.577)	(29.390)	(27.673)	(34.636)	(33.194)
Reputation	0.328	0.396	0.576	0.677	0.895	1.006
	(1.533)	(1.454)	(1.549)	(1.476)	(1.453)	(1.393)

Age	0.246	0.230	0.361	0.358	0.271	0.275
	(0.195)	(0.193)	(0.228)	(0.228)	(0.235)	(0.234)
Size	-8.399***	-5.825***	-10.454***	-7.906***	-11.272***	-8.738***
	(2.205)	(2.138)	(2.719)	(2.683)	(3.082)	(3.027)
Leverage	3.465	3.596	4.628	4.370	6.663	6.261
	(8.755)	(8.700)	(11.224)	(11.277)	(12.017)	(12.120)
Profitability	-34.271**	-32.695**	-38.511*	-36.710*	-43.600*	-41.645*
	(15.773)	(15.285)	(20.141)	(19.727)	(22.273)	(21.733)
Constant	145.034***	74.581	142.227***	76.562	136.043***	79.344
	(41.337)	(48.825)	(46.219)	(53.814)	(52.393)	(61.148)
Year dummies	Y	Y	Y	Y	Y	Y
Industry-fixed	Y	Y	Y	Y	Y	Y
N	442	442	442	442	442	442
Adjusted R^2	0.137	0.219	0.137	0.210	0.109	0.170

*** p < .01, ** p < .05, * p <.10

Table 10
Long-run Performance of IPO firms

The dependent variable in column (1) and (2) is a market-to-book equity ratio of the IPO firm t months after the IPO and the depedent variable in column (3), (4), and (5) is $\prod_{t=0}^{T}(1+R_t) - \prod_{t=0}^{T}(1+R_{m,t})$ where R_t is daily stock return of the IPO firm and $R_{m,t}$ is daily market index return at t trading days after the IPO. t is 6 months and 12 months for column (1) and (2), and T is 90, 120, and 180 for column (3), (4), and (5) respectively. $Tone\ of\ Description$ and $Amount\ of\ Description$ are a normalized tone of sentences and a normalized number of sentences in the Risk Factors Section of the prospectus, respectively. See Appendix A for details in textual variables. $Initial\ Price\ Range$ (%) is defnied as $100 \times (P^{min}_{\text{initial}} - P^{max}_{\text{initial}})/Initial\ mid\ Price$ where $Initial\ Mid\ Price$ is defined as $(P^{min}_{\text{initial}} + P^{max}_{\text{initial}})/2$ for initial prices in the preliminary prospectus denoted as the initial interval $\left[(P^{min}_{\text{initial}}, P^{max}_{\text{initial}}\right]$. $Offer\ Price/Initial\ Mid\ Price$ (%) and $Offer\ Price/Avg.Inst.Bid$ (%) are the level of the finalized offer price to $Initial\ Mid\ Price$ and $Avg.Inst.Bid$ respectively, where $Avg.Inst.Bid$ is quantity-weighted average institutional investors bidding price. For institutional investors' bidding information, $Avg.Inst.Bid/Initial\ Mid\ Price$ (%) denotes the level $of\ Avg.Inst.Bid$ to $Initial\ Mid\ Price$ and $\log Inst.Oversubs\ (All)$ denotes the natural logarithm of institutional investors' oversubscription ratio for limit and strike bids. $\log Retail\ Oversubs$ is the natural logarithm of retail investors' oversubscription ratio. We control for several firm characteristic variables: the natural logarithm of total assets ($Size$), the number of years from the firm establishment to the firm listing (Age), total liabilities over total assets ($Leverage$), operating income scaled by total assets ($Profitability$), and KOSPI dummy ($KOSPI$) which indicates 1 if a firm is listed in KOSPI and 0 if a firm is listed in KOSDAQ. We also control for underwriter's reputation measured by the natural logarithm of 1+market share of underwriter in IPO market in previous year ($Reputation$) and an average return of the KOSPI index on 90-trading days prior to the IPO date, ending at $t-1$ ($Market\ Return$). Year dummies and industry-fixed effects are included for all regressions. White (1980)'s standard errors are reported in parenthesis.

	Market-to-book		BHAR (%)		
	6 months	12 months	[0, 90]	[0, 120]	[0,180]
	(1)	(2)	(3)	(4)	(5)
Description tone	0.499** (0.251)	0.047 (0.117)	-0.160 (3.277)	-0.270 (3.593)	3.542 (7.994)
Description length	0.689** (0.327)	0.911* (0.494)	11.128** (4.692)	15.204*** (4.925)	14.272* (7.387)
Initial Range (%)	-0.001 (0.043)	0.024 (0.026)	-0.500 (0.519)	-0.494 (0.551)	-0.822 (0.813)
Issue Price/Avg. Inst. Bid (%)	0.082** (0.035)	0.006 (0.024)	-0.275 (0.478)	0.032 (0.498)	0.254 (0.651)
Avg. Inst. Bid/Initial Mid Price (%)	0.067** (0.026)	0.043** (0.020)	0.772* (0.401)	0.888** (0.419)	1.328** (0.518)
log(Inst. Oversubs (All))	-0.303 (0.381)	-0.265 (0.299)	-3.346 (5.676)	-4.793 (6.292)	-26.555 (17.951)
log(Retail Oversubs)	0.007 (0.137)	-0.090 (0.147)	-0.025 (2.266)	-0.003 (2.465)	8.185 (7.726)
KOSPI	2.401*** (0.751)	0.605 (0.570)	10.974 (10.024)	3.292 (11.119)	-19.962 (21.354)
Market Timing	-1.608 (2.623)	0.145 (1.483)	-18.285 (37.825)	-0.719 (42.569)	31.144 (56.181)
Reputation	0.133 (0.129)	0.129** (0.051)	4.843* (2.643)	5.978** (2.571)	1.572 (2.388)
Age	-0.025 (0.026)	-0.007 (0.010)	0.259 (0.285)	0.359 (0.306)	0.713* (0.390)
Size	-1.764*** (0.465)	-0.820*** (0.170)	-9.674** (3.986)	-9.644** (4.264)	-2.099 (6.909)
Leverage	9.526 (6.157)	0.702 (0.798)	16.713 (16.447)	9.257 (15.807)	22.549 (21.302)
Profitability	-9.487*** (3.330)	-2.725* (1.609)	-41.354 (25.979)	-54.359** (27.224)	-61.358** (28.068)
Constant	20.455*** (6.874)	17.494*** (3.939)	275.536*** (82.074)	375.000*** (87.105)	269.103* (161.566)

Year dummies	Y	Y	Y	Y	Y
Industry-fixed	Y	Y	Y	Y	Y
N	442	425	442	442	432
Adjusted R^2	0.340	0.154	0.138	0.184	0.084

$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$

Table A1
5-fold Cross-validation

	Accuracy
Fold 1	72.95%
Fold 2	75.40%
Fold 3	78.50%
Fold 4	76.45%
Fold 5	75.95%

Table A2
Correlation Matrix of Tones between Coders

	Coder A	Coder B	Coder C
Coder A	1	0.930	0.887
Coder B		1	0.911
Coder C			1