

The Impact of Information Disclosure Consistency on IPO Pricing Efficiency

-- Evidence from China's A-share Market

Ruyi Shen^a

School of Public Finance and Taxation, Guangdong University of Finance and Economics,
Guangzhou, Guangdong, China

^aEmail: 2542681622@qq.com

Abstract

The core of the comprehensive registration-based IPO reform is information disclosure. However, traditional single-period static textual analysis is limited in identifying ex ante over-packaging and strategic disclosure by IPO firms driven by financing incentives. Using IPO firms listed on China's A-share registration-based boards from 2019 to 2025 as the sample, this study applies the TF-IDF algorithm and cosine similarity model in natural language processing to quantify intertemporal textual consistency between prospectuses and the first annual reports after listing, and empirically examines its effect on IPO underpricing. The results show a significant negative association between information disclosure consistency and IPO underpricing: the lower the ex post observed textual consistency, the higher the underpricing compensation required by investors. This conclusion remains robust after employing an instrumental-variable approach to alleviate endogeneity concerns. Further tests indicate that, after additionally controlling for secondary-market sentiment noise measured by post-listing first-month turnover and information-environment factors proxied by analyst coverage, the coefficient on information disclosure consistency remains significantly negative. This finding suggests that intertemporal textual consistency has independent pricing value in restraining IPO underpricing and does not merely reflect short-term secondary-market noise. Further analysis reveals a substitution effect between internal corporate governance and information disclosure consistency in reducing IPO underpricing, and this effect is more pronounced in market segments with higher information asymmetry. The study provides quantitative evidence and policy implications for regulators to introduce regulatory technology for penetrating textual scanning, for IPO firms to maintain long-term honest disclosure, and for investors to avoid pricing risks.

Keywords

IPO Underpricing; Information Disclosure Consistency; Textual Analysis; TF-IDF Algorithm; Investor Heterogeneous Beliefs.

1. Introduction

The pricing efficiency of initial public offerings (IPOs) has long been a central issue in finance. Low pricing efficiency may lead to resource misallocation in capital markets and amplify market volatility. IPO underpricing, defined as the first-day closing price premium over the offer price, is the most typical manifestation of inefficient pricing and is widely observed across capital markets. In fact, IPO underpricing is not merely a pricing error. As Sherman and Titman [1] argue, underpricing is essentially a risk premium paid by underwriters and issuers to induce informed investors to reveal private information or to compensate them for information

acquisition costs. Although such a premium has theoretical justification under information asymmetry, excessive underpricing still substantially reduces funds raised by issuers and may trigger intense post-listing speculation and even long-run performance reversals. Therefore, how to optimize public information disclosure, mitigate information asymmetry between issuers and investors, and reduce unreasonable IPO underpricing has become an important research topic.

As natural language processing has developed quickly in recent years, more scholars have started to examine how non-financial textual information affects pricing efficiency in capital markets. The seminal study by Loughran and McDonald [2] shows that textual analysis methods designed for financial settings can capture hidden fundamental signals in prospectuses and annual reports. Yet most existing studies look at disclosure text in a single period only. That is a problem. Strong financing incentives may lead issuers to use static disclosure features for over-packaging or strategic disclosure. In this setting, relying only on pre-issuance text from one point in time is no longer enough to measure the true quality of corporate information disclosure. This study therefore examines disclosure quality through dynamic textual consistency. Prior research finds that year-over-year changes made by management to the narrative parts of periodic reports contain useful economic signals about changes in firm fundamentals [3]. During the IPO process, firms may over-package prospectuses to appear more attractive, and once investors view that information as unreliable, they demand higher risk premiums during book building, which increases IPO underpricing. If a firm does engage in excessive embellishment at the IPO stage, that style of disclosure is often difficult to maintain. After listing, disclosure requirements become stricter, and firms like this may have to revise their descriptions of core business and risk factors in later annual reports. For that reason, textual differences between the prospectus and the first post-listing annual report may indicate, to some extent, whether the pre-IPO disclosure was credible, and can serve as an ex post proxy for disclosure credibility.

Intertemporal textual consistency may influence IPO pricing through two main channels, information credibility and comprehension difficulty. Classical signaling theory argues that in an opaque information environment, sending truthful and high-quality signals is costly [4]. Firms without corresponding fundamental support may face greater litigation and reputational costs if they try to keep disclosure text consistent before and after listing, so they have less reason to maintain that consistency. In this sense, intertemporal textual consistency can serve as a harder-to-imitate signal of whether corporate information is authentic and reliable. There is also a second channel. Similar to the way financial statement comparability can lower analysts' information collection and processing costs [5], consistency in the wording and structure of pre- and post-listing disclosures makes it easier for investors to identify actual operating changes without constantly adjusting how they interpret the information. As a result, investors may find the information easier to process, their value assessments may become less dispersed, and both valuation disagreement and uncertainty may fall. This would reduce the risk compensation investors require and, in turn, lessen first-day IPO underpricing.

This study makes three contributions. First, it moves the textual analysis of non-financial information beyond traditional static cross-sectional features and brings in dynamic intertemporal comparison, which adds to the literature on disclosure quality and asset pricing. Second, at the method level, it applies textual similarity algorithms to measure intertemporal consistency, addressing the weakness of single-period static text, which is relatively easy to manipulate, and offering another way to quantify non-financial information. Third, in practical terms, it provides regulators with a quantitative tool to identify over-packaged disclosure and improve regulatory efficiency under the registration-based IPO system.

2. Literature Review

2.1. IPO Pricing Efficiency, Information Asymmetry, and Strategic Disclosure

IPO pricing efficiency has always been a core issue in the resource allocation function of capital markets. Regarding the causes of high IPO underpricing, the classical book-building theory argues that information asymmetry between issuers and outside investors is the fundamental driver of high risk premiums. IPO underpricing is essentially the price paid by underwriters and issuers to induce informed investors to reveal private information or to compensate them for information acquisition costs [1]. Although a certain degree of underpricing is theoretically reasonable under information frictions, issuers have strong incentives, driven by substantial financing benefits, to exploit information disadvantages through opportunistic strategic disclosure.

Beyond traditional earnings management, recent studies find that managers also manipulate non-financial information to conceal underlying risks. In China's A-share market in particular, managers driven by self-interest increase the complexity of annual-report text to hide bad news [6] and use textual tone management as a covert substitute for accrual manipulation [7]. Such excessive marketing and short-term hype before IPOs may increase first-day returns in the short run, but they are often accompanied by long-term performance reversals and value corrections [8]. Therefore, effectively identifying issuers' hidden strategic embellishment and reducing the incentive for private information search [9] is a key path to alleviating high IPO underpricing and resource misallocation.

2.2. Textual Analysis in Capital Markets: From Static Feature Extraction to Dynamic Consistency Comparison

To identify over-packaging during the IPO process, scholars have gradually shifted attention from easily manipulated hard financial data to narrative soft information in prospectuses. The pioneering work of Loughran and McDonald [2] confirms that textual analysis based on finance-specific contexts can effectively capture hidden incremental fundamental signals. With the evolution of large language models and other deep learning technologies, textual feature extraction has moved beyond traditional word-frequency statistics; frontier pre-trained models can more accurately capture contextual semantics and significantly reduce heterogeneous investor beliefs [10].

Although single-period static textual features have been shown to have pricing implications, in practice they can easily become templated tools used by issuers to cater to the market. Therefore, the intertemporal comparability and dynamic consistency of disclosure possess more authentic and harder-to-fake economic value. At the macro level, research shows that high financial statement comparability can substantially reduce analysts' costs of obtaining and processing information and improve overall forecast accuracy [5]. At the textual level, the dynamic trajectory across periods is itself an important basis for verifying corporate integrity. Brown and Tucker [3] find that year-over-year modifications in MD&A text trigger significant capital-market reactions. More direct domestic evidence shows that when listed firms engage in violations or concealment, the similarity of annual-report textual information decreases significantly [11]. These findings indicate that highly consistent intertemporal text objectively reflects operational stability and compliance, whereas sharp textual changes are often passive consequences of prior packaging and hidden risks.

In summary, existing studies have made substantial progress in explaining IPO underpricing, uncovering the information content of static text, and analyzing managerial opportunistic disclosure. Nevertheless, an important research gap remains: prior textual analysis mostly focuses on single-period static features or on routine intertemporal comparisons among mature listed firms. Few studies cross the institutional boundary of the IPO, where incentives

are highly distorted, to dynamically examine the evolution of textual disclosure consistency from pre-listing to post-listing stages. As discussed above, extremely high IPO underpricing is often associated with ex ante packaging, and a decline in ex post textual similarity naturally reveals prior over-embellishment. This study therefore treats intertemporal textual consistency as an ex post validation proxy for the credibility of ex ante disclosure and empirically examines its impact on IPO pricing efficiency, aiming to provide a more dynamic and multidimensional quantitative perspective for information disclosure regulation under the comprehensive registration-based system.

3. Theoretical Analysis and Research Hypotheses

3.1. Intertemporal Textual Consistency and IPO Underpricing

During IPO pricing, inherent information gaps exist between issuers and investors. To maximize financing proceeds, some firms not only engage in traditional earnings management [12] but also tend to conduct strategic disclosure and short-term hype in prospectus drafting [8]. Recent domestic evidence further shows that, driven by self-interest, managers increase annual-report textual complexity to hide bad news [6] or manipulate textual tone to support accrual-based manipulation [7]. Rational institutional investors can detect the information risks caused by such over-packaging during book building. To compensate for high private information acquisition costs [1], investors demand higher risk compensation, thereby increasing IPO underpricing.

However, text embellished through short-term hype lacks long-term fundamental support. When firms face the real operating environment after listing, management may have to substantially revise core business and risk descriptions in the first annual report. In particular, when firms engage in pre-listing over-packaging or concealment, the similarity of their ex post disclosed text declines significantly [11]. In other words, a sharp decline in intertemporal textual consistency between the prospectus and the annual report represents ex post exposure of ex ante packaging [3]. Conversely, according to signaling theory [4], maintaining high intertemporal textual consistency is a costly and hard-to-imitate truthful signal. High-quality continuous disclosure reduces investors' incentives for private information search [9].

Moreover, just as financial statement comparability reduces analysts' information processing costs [5], dynamically consistent textual information also substantially reduces external investors' analytical costs and narrows heterogeneous investor beliefs [10]. Therefore, this study uses intertemporal textual consistency as an ex post validation proxy for the degree of ex ante information packaging and proposes the following hypothesis:

H1: The lower the textual consistency between the prospectus and the first annual report after listing, the higher the IPO underpricing rate.

3.2. Moderating Effect of Internal Governance Characteristics

The signaling and noise-reduction functions of textual consistency are not homogeneous across all firms. Internal corporate governance characteristics have a profound influence on IPO pricing efficiency and first-day returns [13]. In firms with higher internal governance quality, outside investors can already obtain credible assurance of firm quality through well-developed governance mechanisms, so the marginal signal value of non-financial textual information is relatively limited. By contrast, when internal governance mechanisms are weak and cannot effectively constrain managerial opportunism, investors face severe adverse-selection risk, and external public disclosure becomes an important basis for assessing firms' true value. In such a high-risk environment, the credibility signal conveyed by intertemporal textual consistency is amplified and becomes an important mechanism for alleviating information asymmetry. Accordingly, internal governance mechanisms and external information disclosure may exhibit

functional substitution in reducing IPO underpricing. This study proposes the following hypothesis:

H2: Corporate governance and information disclosure consistency exhibit a substitution effect in reducing IPO underpricing; that is, the negative effect of textual consistency on IPO underpricing is more significant among firms with weaker internal governance.

3.3. Heterogeneity across Market Boards and Information Environments

The pricing efficiency of IPO firms depends heavily on the characteristics of their market board and information environment. Under China's A-share registration-based reform, the STAR Market and ChiNext differ in positioning and listing standards. Compared with the STAR Market, which has a clearer high-tech orientation and relatively higher sponsorship thresholds, ChiNext covers broader business forms and includes more early-stage firms, leading to relatively higher information asymmetry. In market boards where information asymmetry is more severe and the value of traditional financial signals is more limited, investors rely more heavily on narrative textual information in prospectuses when conducting risk screening. Therefore, textual consistency, as a dynamic proxy that directly reveals the credibility of management's ex ante disclosure, should play a more pronounced noise-reduction and pricing-correction role in boards with greater reliance on soft information and more severe information asymmetry. This study proposes the following hypothesis:

H3: In boards with higher information asymmetry, the negative effect of textual consistency on IPO underpricing is more pronounced.

4. Research Design

4.1. Sample Selection and Data Sources

This study selects firms that completed IPOs on China's A-share registration-based boards, including the STAR Market and ChiNext, from 2019 to 2025 as the initial sample. The initial sample contains 1,244 IPO firms. The sample is screened according to the following criteria: first, observations with missing prospectus or annual-report PDF files are excluded; second, observations with PDF files that cannot be parsed or from which text extraction fails are excluded; third, observations with missing corporate-governance data are excluded. After this screening process, the final sample contains 689 valid observations. Because the consistency measure requires matched, machine-readable prospectus and annual-report texts together with complete governance variables, the screening process leads to non-trivial sample attrition. This attrition mainly reflects data-availability and text-extraction quality constraints rather than discretionary sample exclusion. Therefore, the empirical results should be interpreted for IPO firms with complete and machine-readable disclosure materials, and the potential selection issue is further acknowledged in Section 6.3.

The prospectus and annual-report textual data are obtained from CNINFO and downloaded in batches using Python web-crawling techniques. Corporate financial data, IPO data, and corporate governance data are obtained from the CSMAR database. To avoid the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles.

4.2. Variable Definitions

(1) Dependent Variable: IPO Underpricing (Underpricing)

Following the classic literature, IPO underpricing is calculated as the percentage change of the first-day closing price relative to the offer price:

$$\text{Underpricing}_i = (P_{\text{close},i} - P_{\text{offer},i}) / P_{\text{offer},i}$$

where the numerator denotes the difference between the first-day closing price and the offer price, and the denominator is the offer price. A higher value indicates lower pricing efficiency in the IPO primary market.

(2) Core Explanatory Variable: Information Disclosure Consistency (Consistency)

Drawing on Brown and Tucker's [3] approach of measuring disclosure modifications based on textual similarity, this study constructs an intertemporal textual consistency measure. Specifically, it compares the "Risk Factors" section of the prospectus with the "Management Discussion and Analysis" section of the first annual report after listing. The TF-IDF algorithm is used to transform text into vectors, and cosine similarity is then calculated. If section extraction fails, whole-document similarity between the two documents is used as an alternative. This indicator ranges from 0 to 1. A value closer to 1 indicates higher textual consistency between pre- and post-listing disclosures and a lower degree of information packaging.

(3) Moderating Variable: Corporate Governance Score (Gov_Score)

Following prior studies, this study selects eight corporate governance indicators, including CEO duality, the shareholding ratio of the largest shareholder, board shareholding, executive shareholding, board size, supervisory board size, executive compensation, and independent director ratio. The first principal component is extracted through principal component analysis to construct a composite corporate governance score. A higher score indicates a more well-developed internal governance mechanism. To examine the moderating effect of internal governance, the sample is divided into high-governance and low-governance groups based on the median governance score.

(4) Grouping Variable: Listing Board (Market_Type)

Firms are grouped according to the exchange board on which they completed the IPO, namely the STAR Market and ChiNext. The STAR Market mainly serves sci-tech innovation enterprises, whereas ChiNext covers a broader set of business forms and firms at more heterogeneous growth stages. This grouping is suitable for testing the heterogeneous effects of information disclosure consistency across market boards.

(5) Control Variables

Following existing studies on IPO pricing efficiency [14,15], this study controls for three categories of factors that may affect IPO underpricing:

The first category consists of firm characteristics, including firm size (Size, the natural logarithm of total assets), profitability (ROA, return on assets), financial leverage (Lev, debt-to-asset ratio), and firm age (Age, the number of years from establishment to IPO).

The second category consists of issuance characteristics, including issue size (Issue_Size, the natural logarithm of actual proceeds raised) and online lottery rate (Lottery_Rate). A lower lottery rate indicates stronger market demand for the stock.

The third category consists of intermediary reputation variables, including auditor reputation (Big4, whether the auditor is one of the Big Four accounting firms) and underwriter reputation (UW_Top10, whether the lead underwriter is among the top ten securities firms).

Detailed definitions are reported in the variable definition table below:

Table 1. Variable Definition

Variable Category	Variable Name	Symbol	Definition and Calculation
Dependent variable	IPO underpricing	Underpricing	(First-day closing price - Offer price) / Offer price
Core explanatory variable	Information disclosure consistency	Consistency	TF-IDF cosine similarity between the prospectus and the first post-listing annual report; the value ranges from 0 to 1, and a value closer to 1 indicates higher textual consistency.
Moderating variable	Corporate governance score	Gov_Score	The first principal component extracted from eight governance indicators: CEO duality, largest-shareholder ownership, board ownership, executive ownership, board size, supervisory board size, executive compensation, and independent director ratio.
Grouping variable	Listing board	Market_Type	Dummy variable: STAR Market = 1; ChiNext = 0.
	High-tech industry	Hi_tech	Dummy variable equal to 1 if the firm belongs to C27, C39, I65, M73, or M74 under the CSRC 2012 industry classification, and 0 otherwise.
Mechanism variable	Analyst coverage	Analyst_Coverage	Number of analysts or analyst teams issuing earnings forecasts for the firm within one year after IPO; missing values are filled with the median.
	Average daily turnover	avg_turnover	Average daily turnover during the first month after listing (30 trading days, based on tradable shares, %).
Control variables			
Firm characteristics	Firm size	Size	Natural logarithm of total assets at the end of the year before IPO.
	Return on assets	ROA	Net profit in the year before IPO / total assets at period end.
	Leverage	Lev	Total liabilities / total assets at the end of the year before IPO.
	Firm age	Age	Number of years from firm establishment to IPO year.
Issuance characteristics	Issue size	Issue_Size	Natural logarithm of actual proceeds raised.
	Online lottery rate	Lottery_Rate	Online offering lottery rate.
Intermediary reputation	Big Four auditor	Big4	Dummy variable equal to 1 if the auditor is an international Big Four accounting firm, and 0 otherwise.
	High-reputation underwriter	UW_Top10	Dummy variable equal to 1 if the lead underwriter is among the top ten securities firms in that year, and 0 otherwise.

4.3. Model Specification

To test the core hypothesis H1, the following baseline regression model is constructed:

$$\text{Underpricing}_i = \alpha + \beta_1 \text{Consistency}_i + \gamma' \text{Controls}_i + \varepsilon_i$$

where the dependent variable is IPO underpricing, the core explanatory variable is information disclosure consistency, and the control-variable vector includes the variables described above. According to H1, the coefficient on Consistency is expected to be significantly negative; that is, higher information disclosure consistency is associated with lower IPO underpricing.

To test H2, the moderating effect of corporate governance, the sample is divided into high-governance and low-governance groups according to the median governance score. Separate regressions are then estimated, and the significance and magnitude of the consistency coefficient are compared across the two groups. If the negative effect of consistency is more significant in the low-governance group, it indicates a substitution effect between internal governance and information disclosure, implying functional complementarity in reducing IPO underpricing.

To test H3, the heterogeneous effect of the information environment, the sample is divided by listing board into STAR Market and ChiNext groups, and the same model specification is estimated separately. If the absolute value of the consistency coefficient is larger and statistically significant in the board with higher information asymmetry, the pricing-correction effect of information disclosure consistency is more pronounced.

To further rule out interference from short-term secondary-market trading sentiment and external information-environment differences, this study augments the baseline regression model with the post-listing first-month average daily turnover (*avg_turnover*) and analyst coverage (*Analyst_Coverage*), constructing the following mechanism test that controls for secondary-market noise:

$$\text{Underpricing}_i = \alpha + \beta_1 \text{Consistency}_i + \beta_2 \text{avg_turnover}_i + \beta_3 \text{Analyst_Coverage}_i + \gamma' \text{Controls}_i + \varepsilon_i$$

Here, *avg_turnover* is the average daily turnover during the first month after listing, capturing short-term secondary-market trading sentiment and investor disagreement; *Analyst_Coverage* is the number of analysts or analyst teams issuing earnings forecasts for the firm within one year after the IPO, capturing the external information environment. If the coefficient on Consistency remains significantly negative after controlling for these secondary-market noise and information-environment variables, information disclosure consistency has independent pricing explanatory power for IPO underpricing.

4.4. Descriptive Statistics

Table 2 reports descriptive statistics for the main variables. After winsorization and missing-value processing, the final sample contains 689 observations. The mean IPO underpricing rate is 1.158, with a standard deviation of 1.244, indicating substantial variation in first-day returns among IPOs on China's A-share registration-based boards. The mean information disclosure consistency is 0.178, with a median of 0.140 and a range from 0 to 0.695, indicating considerable cross-firm variation in textual disclosure consistency before and after listing. The distributions of the remaining variables are within reasonable ranges and are not discussed individually here.

Table 2. Descriptive Statistics of Main Variables

Variable	Obs.	Mean	Std. Dev.	Min	25th pct.	Median	75th pct.	Max
IPO underpricing	689	1.158	1.244	-0.250	0.303	0.787	1.731	6.934
Information disclosure consistency	689	0.178	0.179	0.000	0.018	0.140	0.248	0.695
Firm size	689	20.740	0.815	19.373	20.188	20.584	21.149	23.516
Return on assets	689	0.057	0.069	-0.202	0.022	0.046	0.086	0.280
Leverage	689	0.354	0.175	0.062	0.212	0.341	0.471	0.806
Firm age	689	15.498	5.586	5.000	11.000	15.000	19.000	32.000
Issue size	689	11.365	0.728	10.001	10.869	11.290	11.806	13.367
Online lottery rate	689	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Big Four auditor	689	0.080	0.271	0.000	0.000	0.000	0.000	1.000
High-reputation underwriter	689	0.444	0.497	0.000	0.000	0.000	1.000	1.000
Corporate governance score	689	0.627	1.510	-2.422	-0.559	0.447	1.721	4.145
High-tech industry	689	0.444	0.497	0.000	0.000	0.000	1.000	1.000

5. Empirical Results

5.1. Baseline Regression Analysis

Table 3 reports the baseline regression results for the effect of information disclosure consistency on IPO underpricing. To examine how different groups of control variables affect the results, three models are estimated. Model (1) includes only the core explanatory variable, information disclosure consistency. The coefficient is -0.392 and is significant at the 10% level, which gives initial support for H1. Model (2) adds firm-level control variables. After these variables are included, the coefficient on consistency is -0.380 and remains significantly negative at the 10% level.

Model (3) further includes issuance characteristics and intermediary reputation variables, and the model's explanatory power increases substantially ($R = 0.256$). The consistency coefficient is -0.445 and significant at the 5% level. The result becomes clearer here.

Among the control variables, issue size is significantly negative at the 1% level, which is in line with expectations. When the proceeds raised are larger, firms tend to provide more information, and the degree of underpricing is lower. Taken together, the results suggest that the restraining effect of information disclosure consistency on IPO underpricing remains after firm characteristics, issuance characteristics, and intermediary reputation factors are added step by step.

Table 3. Regression Results for the Effect of Information Disclosure Consistency on IPO Underpricing

Variable	(1) Baseline	(2) Firm Controls	(3) Full Model
consistency	-0.392*	-0.380*	-0.445**
	(0.229)	(0.227)	(0.203)
Size		-0.047	0.444***
		(0.072)	(0.091)
ROA		-0.571	0.538
		(0.646)	(0.648)
Lev		0.002	-0.612*
		(0.372)	(0.371)
Age		0.011	-0.010
		(0.008)	(0.007)
Issue_Size			-0.993***
			(0.080)
Lottery_Rate			-99.177
			(368.807)
Big4			0.239
			(0.173)
UW_Top10			0.320***
			(0.085)
Constant	1.227***	2.063	3.527**
	(0.065)	(1.370)	(1.553)
Observations	689	689	689
R-squared	0.003	0.007	0.256

5.2. Robustness Tests

To check whether the main result is reliable, this study runs several robustness tests, reported in Table 4. In the first test, the winsorization criterion is shifted from the 1%/99% range to the

5%/95% range, and the full model is estimated again. The consistency coefficient is -0.444 and significant at the 5% level. The result is similar.

The second test removes observations with zero consistency scores, since these cases may mean that the two texts are not fully comparable. After re-estimation, the consistency coefficient is -0.385 and significant at the 10% level. A third check focuses on the sample period. Because firms listed in 2024-2025 have relatively short post-listing histories, the sample is limited to 2019-2023 and the model is re-estimated. The consistency coefficient is -0.339 and significant at the 10% level.

Across these three tests, the sign and significance of the consistency coefficient remain broadly stable, which suggests that the core conclusion does not change.

Table 4. Robustness Tests

Variable	(1) Alternative winsorization	(2) Excluding consistency = 0	(3) 2019-2023 subsample
consistency	-0.444**	-0.385*	-0.339*
	(0.194)	(0.231)	(0.201)
Size	0.360***	0.452***	0.330***
	(0.070)	(0.101)	(0.073)
ROA	1.271*	0.246	0.840
	(0.747)	(0.751)	(0.566)
Lev	-0.194	-0.624	-0.255
	(0.273)	(0.420)	(0.293)
Age	-0.007	-0.012	-0.004
	(0.006)	(0.008)	(0.006)
Issue_Size	-0.901***	-0.988***	-0.959***
	(0.066)	(0.091)	(0.082)
Lottery_Rate	-5.565	-108.817	287.594
	(333.463)	(392.517)	(342.219)
Big4	0.275*	0.374*	0.275*
	(0.146)	(0.204)	(0.161)
UW_Top10	0.253***	0.375***	0.291***
	(0.071)	(0.095)	(0.086)
Constant	3.908***	3.302*	5.098***
	(1.210)	(1.720)	(1.137)
Observations	689	579	645
R-squared	0.265	0.254	0.260

5.3. Endogeneity Test

Information disclosure consistency and IPO underpricing may be related through omitted variables or reverse causality, so endogeneity is a concern here. To address this, this study uses the mean information disclosure consistency of other firms in the same industry and year as an instrumental variable and applies two-stage least squares regression. Table 5 compares the OLS and IV-2SLS results.

Table 5. Endogeneity Test: OLS versus IV-2SLS

Variable	(1) OLS	(2) IV-2SLS
consistency	-0.445**	-3.170***
	(0.203)	(0.759)
Size	0.444***	0.536***
	(0.091)	(0.101)
ROA	0.538	0.660
	(0.648)	(0.687)
Lev	-0.612*	-0.695*
	(0.371)	(0.386)
Age	-0.010	-0.010
	(0.007)	(0.008)
Issue_Size	-0.993***	-1.002***
	(0.080)	(0.086)
Lottery_Rate	-99.177	-679.541
	(368.807)	(450.034)
Big4	0.239	0.134
	(0.173)	(0.192)
UW_Top10	0.320***	0.220**
	(0.085)	(0.094)
Constant	3.527**	2.464
	(1.553)	(1.678)
Observations	689	689
R-squared	0.256	0.108
First-stage F statistic	—	125.72
Weak-instrument test	—	Passed

Note: Robust standard errors are reported in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The first-stage regression reports a Kleibergen-Paap Wald F statistic of 125.72, which is well above the 10% critical value of 16.38 proposed by Stock and Yogo [16]. This largely rules out a weak-instrument problem. In the second-stage regression, the coefficient on information disclosure consistency is -3.170 and remains significantly negative at the 1% level. The sign is the same as in the baseline OLS regression. Taken together, the IV results suggest that the main conclusion still holds after accounting for potential endogeneity.

5.4. Mechanism Test Controlling for Secondary-Market Noise

Table 6. Mechanism Test Controlling for Secondary-Market Noise

Variable	(1) Underpricing
consistency	-0.481**
	(0.202)
avg_turnover	0.017***
	(0.005)
Analyst_Coverage	0.022***
	(0.007)
Size	0.389***
	(0.090)
ROA	0.347
	(0.623)
Lev	-0.605*
	(0.367)
Age	-0.009
	(0.007)
Issue_Size	-0.947***
	(0.091)
Lottery_Rate	-63.494
	(352.870)
Big4	0.244
	(0.176)
UW_Top10	0.327***
	(0.084)
Constant	3.629**
	(1.658)
Observations	689
R-squared	0.279

Note: Robust standard errors are reported in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The preceding analysis confirms that information disclosure consistency significantly restrains IPO underpricing. To further verify that this conclusion is not driven by post-listing short-term secondary-market trading sentiment or differences in the external information environment, this study further adds the post-listing first-month average daily turnover and analyst coverage to the baseline regression. Average daily turnover captures secondary-market trading activity, short-term sentiment, and investor disagreement, whereas analyst coverage captures the information environment after listing. If the coefficient on consistency remains significantly negative after controlling for these factors, intertemporal textual consistency between the prospectus and the first post-listing annual report is not merely a concomitant result of secondary-market noise but an information-quality signal with independent pricing value.

Table 6 reports the extended regression results after controlling for secondary-market noise. The results show that, after further controlling for post-listing first-month average daily turnover and analyst coverage, the coefficient on information disclosure consistency is -0.481 and significantly negative at the 5% level. Meanwhile, the coefficient on turnover is 0.017 and significantly positive at the 1% level, while the coefficient on analyst coverage is 0.022 and significantly positive at the 1% level. This suggests that short-term secondary-market trading activity and external information attention both affect IPO first-day underpricing.

More importantly, after adding short-term secondary-market trading noise and information-environment variables, the consistency coefficient remains significantly negative. In other words, the restraining effect of information disclosure consistency on IPO underpricing still holds, and its pricing value appears to be independent of short-term secondary-market noise.

5.5. Moderating Effect Analysis

To test H2, namely whether internal corporate governance affects the restraining effect of information disclosure consistency on IPO underpricing, the sample is divided into high-governance and low-governance groups based on the sample median of the corporate governance score, and separate regressions are run for each group. The idea is straightforward. If the negative effect of information disclosure consistency is stronger in the low-governance group, this points to a substitution effect between internal governance mechanisms and external information disclosure. In that case, both help blueuce IPO underpricing, but when one source of assurance is weaker, the other matters more.

Table 7 reports the group regression results. In the high-governance group, the coefficient on information disclosure consistency is -0.238, but it does not pass conventional significance tests ($p = 0.389$). This suggests that when firms already have relatively sound internal governance mechanisms, investors can to some extent infer firm quality from governance arrangements such as equity checks and balances and board structure. Under these conditions, the extra signal provided by textual consistency between the prospectus and annual report is limited. Put differently, internal governance has already taken over part of the pricing-correction role that external information disclosure might otherwise play.

The pattern is different in the low-governance group. There, the coefficient on information disclosure consistency is -0.652 and significantly negative at the 5% level ($p = 0.027$). Its absolute value is much larger than that in the high-governance group and also larger than the full-sample baseline coefficient (-0.445). This matters. When internal governance is weak and cannot effectively constrain managerial opportunism, investors appear to rely more heavily on external disclosure when judging firm value. In this setting, textual consistency carries more weight as a signal, so its effect on correcting IPO underpricing is stronger.

Table 7. Group Regression of the Corporate Governance Moderating Effect

Variable	(1) High governance	(2) Low governance
consistency	-0.238	-0.652**
	(0.275)	(0.293)
Controls	Yes	Yes
Observations	344	345
R-squared	0.226	0.310

Overall, these results support the "substitution effect" hypothesis between internal governance and external information disclosure. Where internal governance is incomplete, investors face greater information asymmetry and adverse-selection risk, and high intertemporal textual disclosure consistency becomes a more important channel for assessing the cblueibility of ex

ante information. That lowers the risk compensation requiblue by investors. By contrast, when internal governance already provides enough assurance,the marginal pricing value of textual consistency declines. This helps explain how information disclosure and corporate governance interact in IPO pricing, and it provides empirical evidence that the effect of disclosure differs across governance environments.

5.6. Heterogeneity Analysis

Table 8. Heterogeneity Analysis by Listing Board

Variable	(1) STAR Market	(2) ChiNext
consistency	-0.356*	-0.645**
	(0.208)	(0.369)
Size	0.508***	0.804***
	(0.124)	(0.128)
ROA	-0.301	2.915***
	(0.793)	(1.013)
Lev	-0.921**	-0.945**
	(0.477)	(0.492)
Age	-0.007	-0.010
	(0.010)	(0.009)
Issue_Size	-0.961***	-1.204***
	(0.099)	(0.132)
Lottery_Rate	-1,649.5**	-3,904.7***
	(748.8)	(938.0)
Big4	0.189	0.093
	(0.226)	(0.250)
UW_Top10	0.148	0.326***
	(0.109)	(0.126)
Constant	2.714	-0.927
	(2.083)	(1.904)
Observations	377	312
R-squared	0.282	0.339

To test H3, this study splits the sample into STAR Market and ChiNext groups based on listing board and then runs separate regressions for each group. The results are shown in Table 8. For the STAR Market group, the coefficient on information disclosure consistency is -0.356 and is significant at the 10% level. For the ChiNext group, the consistency coefficient is -0.645 and is significant at the 5% level. The difference is clear. The absolute value of the consistency coefficient in ChiNext (0.645) is much larger than that in the STAR Market (0.356).

6. Conclusion

6.1. Research Conclusion

Using IPO firms listed on China's A-share registration-based boards from 2019 to 2025, this study adds to the IPO pricing literature by examining intertemporal textual consistency as a dynamic disclosure signal, rather than treating text as a fixed feature at one point in time. The results show that firms whose prospectuses and first post-listing annual reports are more consistent have lower IPO underpricing, and this relation remains stable across alternative specifications and endogeneity tests. This finding adds to signaling theory by suggesting that disclosure continuity across the IPO boundary is a relatively costly signal of disclosure credibility and is harder for lower-quality firms to imitate.

The study further finds that the pricing effect of textual consistency is not just a by-product of secondary-market trading noise. Even after controlling for first-month turnover and analyst coverage, textual consistency still reduces IPO underpricing. That matters. It suggests that cross-period disclosure consistency carries information for primary-market pricing that is not fully explained by post-listing sentiment or outside information attention. In that sense, the study expands the textual analysis literature by showing that dynamic consistency captures part of information quality that other post-listing indicators do not fully reflect.

The evidence also points to a governance-substitution mechanism. Where internal governance is weak, investors appear to rely more on external disclosure when judging firm quality, so the marginal pricing effect of textual consistency becomes stronger. By contrast, when governance mechanisms are already relatively complete, the added role of textual consistency is smaller. This provides a basis for more differentiated information-disclosure supervision, especially for firms with weaker governance arrangements, where disclosure continuity across the IPO process may deserve closer monitoring.

Finally, the stronger effect in market segments with higher information asymmetry suggests that the value of disclosure consistency depends on the surrounding information environment. Under the registration-based IPO system, dynamic textual supervision may therefore be a useful regulatory tool for identifying over-packaging, improving information disclosure quality, and improving IPO pricing efficiency.

6.2. Policy Implications

The core of the registration-based reform is information disclosure. The findings of this study extend the textual analysis framework in asset pricing and offer policy implications for capital-market participants.

Regulators should move toward deeper and more dynamic textual supervision. Under the registration-based system, regulatory attention should not stop at a one-time compliance review of disclosure text at the issuance stage. Instead, regulators can use regulatory technology and natural language processing to carry out broad consistency checks across issuers' prospectuses and post-listing periodic reports over time. Where firms show unusual intertemporal textual deviations, regulators should set up timely inquiry procedures and targeted inspection mechanisms. This matters. From an ex post verification perspective, these measures can help identify possible over-packaging, reinforce issuers' primary responsibility for information disclosure, and reduce the market problem of "over-packaging before listing and performance reversal after listing" at the source. Textual consistency indicators could also be added to the information disclosure quality evaluation system so that differentiated regulation has a clearer quantitative basis.

IPO firms should give up short-term textual packaging and maintain honest disclosure over the longer run. The results indicate that exaggerated textual packaging may draw investor attention in the short term, but reputational costs linked to low ex post consistency are eventually priced by the market. In other words, the market does adjust. IPO firms need to understand that consistent, objective, and truthful textual disclosure is itself a credible signal of firm quality. It reduces information asymmetry and helps build reputation in the capital market. Lower IPO financing costs may follow, and firms can also build a more reliable reputational base for later market value management and subsequent capital operations.

Investors, meanwhile, should treat textual consistency as a useful reference when judging the quality of corporate information. In the complex IPO market, investors should not simply follow the formulaic optimistic tone often found in prospectuses. They need to read that language alongside industry characteristics and the governance environment, so they can assess firms' "ex ante disclosure credibility" from more than one angle. This is especially relevant for firms with weak governance mechanisms and low information transparency. In such cases, investors

should watch changes in textual disclosure consistency before and after listing and use those changes as part of investment risk assessment and pricing. Doing so may help investors avoid the winner's curse and make more forward-looking, rational investment decisions.

6.3. Research Limitations and Future Directions

Although this study finds significant empirical evidence on intertemporal textual consistency and IPO pricing efficiency, several limitations remain because of data and scope constraints, and these point to several directions for later research.

At the sample level, this study examines China's A-share registration-based boards in a Chinese-language setting. The results may not transfer directly to other markets. Legal systems, investor structures, and social cultures differ across countries, so the cross-market validity of these conclusions still needs to be tested. Later research could apply the intertemporal dynamic consistency framework to multilingual and cross-border IPO samples, then examine whether different institutional settings change the pricing effect of textual consistency.

The final sample was also formed after excluding observations with missing prospectus or annual-report PDFs, text-parsing failures, and missing governance data. Those exclusions were necessary, but they create a limitation. If excluded firms differ systematically from retained firms in disclosure quality, sample attrition may have introduced selection bias. In the current dataset, the full exclusion logs and the characteristics of all excluded firms are not available, so this study cannot carry out a complete Heckman correction. A useful next step would be to report attrition by reason, compare included and excluded firms by industry, size, and listing board, and, where possible, conduct formal selection-bias corrections.

At the measurement level, information disclosure consistency in this study is measured with the TF-IDF word-frequency algorithm, which mainly captures lexical overlap. This approach is common in textual analysis research and the results remain stable under different tests, but the method is less effective at identifying contextual semantic similarity. Future studies could use pre-trained deep learning models such as BERT to map text into high-dimensional semantic space and then calculate similarity, which may improve the accuracy and information content of consistency measures. It may also be useful to assign different weights to different sections. For example, the risk factors section could be given a higher weight so that disclosure quality is measured more precisely.

At the mechanism-test level, this study mainly checks the stability of the pricing effect of textual consistency by controlling for short-term secondary-market trading noise and the external information environment. Even so, this does not cover every possible channel. The effect of information disclosure consistency on IPO underpricing may also work through channels such as reducing expected litigation risk, attracting media attention, and increasing analyst forecast consensus. Future work could test these channels more directly and build a fuller mechanism model so the route from information disclosure consistency to capital-market pricing efficiency is specified more clearly.

At the practical application level, this study provides quantitative evidence that regulators may use technical tools to identify information over-packaging under the registration-based system. What remains less clear is how textual consistency indicators should be built into regulatory procedures, how reasonable threshold standards should be set, and how these indicators should be linked with other regulatory tools. These issues still need further study.

References

- [1] Sherman, A. E., & Titman, S. (2002). Building the IPO order book: Underpricing and participation limits with costly information. *Journal of Financial Economics*, 65(1), 3–29.

- [2] 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.
- [3] Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2), 309–346.
- [4] Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- [5] De Franco, G., Kothari, S. P., & Verdi, R. S. (2011). The benefits of financial statement comparability. *Journal of Accounting Research*, 49(4), 895–931.
- [6] Wang, K., Wang, H., Li, D., & Dai, X. (2018). Complexity of annual report text information and managerial self-interest: Evidence from Chinese listed companies. *Management World*, 34(12), 120–132, 194. <https://doi.org/10.19744/j.cnki.11-1235/f.2018.0038>
- [7] Wang, H., & Wang, K. (2018). Accrual manipulation and tone manipulation of annual report text information. *Accounting Research*, (4), 45–51.
- [8] Chahine, S., Colak, G., Hasan, I., & Mazboudi, M. (2020). Investor relations and IPO performance. *Review of Accounting Studies*, 25(2), 474–512.
- [9] Brown, S., & Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2), 443–477.
- [10] Gu, X., He, X., & Liu, H. (2025). Risk information disclosure effect on IPO pricing efficiency: Evidence from China. *International Review of Economics & Finance*, 104492.
- [11] Tan, J., & Wang, X. (2022). Corporate fraud and annual report text manipulation. *China Soft Science*, (3), 99–111.
- [12] 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.
- [13] Teti, E., & Montefusco, I. (2022). Corporate governance and IPO underpricing: Evidence from the Italian market. *Journal of Management and Governance*, 26(3), 851–889.
- [14] Loughran, T., & Ritter, J. R. (2004). Why has IPO underpricing changed over time? *Financial Management*, 33(3), 5–37.
- [15] Lowry, M., & Murphy, K. J. (2007). Executive stock options and IPO underpricing. *Journal of Financial Economics*, 85(1), 39–65.
- [16] Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press.