

# 科创板信息披露质量、网络安全关注与 IPO 抑价：基于招股说明书的文本挖掘分析

陈豪

武汉纺织大学管理学院，湖北武汉，430074；

**摘要：**信息披露是注册制改革的核心，本文以注册制在中国首次实施为背景，以科创板上市公司招股说明书为文本挖掘对象，构建中文金融信息披露质量衡量指标，本研究通过实证分析探讨了招股说明书中的信息披露质量与网络安全关注度对 IPO 抑价的影响。结果表明，科创板企业在招股说明书中披露的信息质量越高，IPO 抑价的程度越低；其中，网络安全治理与合规性信息的披露对投资者的风险评估起到了重要作用，进一步减少了 IPO 抑价。此外，研究还表明，招股说明书的信息披露质量对 IPO 抑价的影响在非国有企业中更为明显。本文扩展了中文金融文本挖掘的研究视角，并从信息披露和网络安全的角度进一步深入探讨了 IPO 抑价问题，为监管机构优化信息披露规定提供了新的启示。

**关键词：**科创板；信息披露质量；IPO 抑价；网络安全；文本挖掘

**DOI：**10.69979/3041-0673.25.06.036

## 引言

信息披露是科创板注册制的核心<sup>[1]</sup>，招股说明书披露的内容分为数字信息和文本信息，投资者往往更关注于企业财务状况、经营业绩等数字信息，而文本信息的复杂性导致投资者忽视了这类信息。近年来已有很多学者对招股说明书文本信息做出相关研究，招股说明书中的文本信息的披露能够反映企业内在价值，对指导投资者投资有积极作用<sup>[2]</sup>；提高了资本市场定价效率，抑制了首日发行收盘价远高于发行价的现象<sup>[3]</sup>；其中不同情绪语调的词汇出现频率能够预测企业未来表现力，对市场定价及投资者投资都起到一定指导作用<sup>[4]</sup>。

随着当今时代信息爆炸和计算机技术的不断进步，相关学者总结出大数据的 4V 属性，分别是数据量大、类型多变、价值稀疏和处理速度快。大数据的价值稀疏这一特点使得数据量增加时边际价值并未有明显提升，数据中伴有大量离群数据。本文的研究旨在对数量庞大的招股说明书文本数据进行挖掘，从中探究信息披露质量与发行定价效率间的关系。

为此需要回答以下研究问题：

研究问题 1：科创板招股说明书信息披露质量与网络安全关注能否影响股票发行的定价效率？

研究问题 2：在不同产权性质的企业中，影响有何区别？

## 1 文献综述

随着互联网的普及与计算机技术的不断提升，文本

挖掘逐渐应用于金融领域用于刻画关注度、情绪语调、可读性等方面。王克敏<sup>[5]</sup>等通过研究中国上市公司年报文本信息的信息复杂性，从管理者自利角度探究其与公司业绩的关系；越来越多的学者对于企业 IPO 的分析开始转向文本信息，通过对企业招股说明书中“软信息”的情绪语调、可读性等方面的研究探寻企业 IPO 抑价的原因。周佰成<sup>[3]</sup>等通过分析招股说明书文本信息构建金融文本词典来分析招股书文本可读性对首发抑价的影响；张飞<sup>[2]</sup>等从招股说明书模糊信息视角出发，通过量化文本迷糊信息研究其如何影响 IPO 首日收益。

IPO 抑价是指 IPO 首日收盘价格远远高于发行价格，新股抑价的现象几乎存在于世界各个国家的股票市场中<sup>[6]</sup>，但在中国市场尤为突出。对于抑价之谜的解释理论繁多，大体可以分为两类：一是信息不对称理论，如“赢者的诅咒”理论、询价理论、信号理论、委托代理理论，一方面，信息不对称理论探讨了不同主体之间的信息差异，例如有信息的投资者与无信息的投资者、投资者与承销商、投资者与发行人，以及发行人与承销商之间的矛盾；另一方面，制度性因素理论则提出了如避免诉讼假说、Ruud 的价格支持理论等观点，解释了 IPO 抑价的形成机制。以往研究多基于公司公开披露的财务信息、新闻媒体报导等测度公司的信息披露程度，而通过研究招股说明书中的非财务信息来解释 IPO 抑价的研究相对不足。董秀良<sup>[7]</sup>等通过采用随机前沿模型分析科创板新股定价的合理性，研究发现科创板上市公司首次

公开发行的定价显著高于其内在价值,且高估程度明显超过了同期的创业板市场。科创板设立以来,涌现出更多学者开始探究新的政策背景下 IPO 抑价能否以及如何得到改善,本文将在理论上论证管理者操纵招股说明书信息披露质量的原因,在实证上运用文本挖掘的方法,对招股说明书信息披露质量与 IPO 抑价之间的关系进行了全面研究。

## 2 理论假设与模型构建

### 2.1 理论假设

管理者在操纵招股说明书可读性进行印象管理时,主要采取两种方式:隐藏和归因。隐藏是指隐藏公司可能存在的风险、较差的业绩等负面信息。IPO 公司管理者通过使用大量复杂词汇和专业术语,降低了招股说明书的可读性。模糊公司经营的真实状况和发展前景<sup>[8]</sup>。归因是指在公司运营状态较好时,管理者将此功劳归因于自身的努力,而当公司业绩不理想时,管理者通常将其归因于外部环境因素。根据优序融资理论,招股说明书可读性降低导致信息不对称程度加深,投资者的信息解读成本上升,从而导致 IPO 抑价程度提高。基于以上分析,本文提出假设:

H1: 科创板 IPO 公司招股说明书文本可读性越好, IPO 抑价越低。

随着科技的发展,网络安全成为上市公司面临的一个重要挑战。在科创板这样的高科技创新型企业中,网络安全的保障措施对投资者的信心至关重要。企业在招股说明书中对网络安全的披露质量越高,投资者对其未来风险的评估就越低,进而减少 IPO 抑价。高质量的网络安全信息披露表明公司有完善的治理结构和有效的风控机制,有助于提升投资者对其运营可持续性的信心,从而降低抑价。

H2: 招股说明书中网络安全关注度越高, IPO 抑价越低。

当信息不对称导致 IPO 抑价时,招股说明书的可读性通过影响投资者的信息获取,从而影响 IPO 抑价。然而,国有企业的 IPO 抑价通常源于完成经济或政策目标,而非纯粹依赖资本市场融资。因此,招股说明书的可读性在国有企业 IPO 抑价中的作用并不成立。基于以上分析,本文提出假设:

H3: 信息可读性与 IPO 抑价的负向关系在非国有企业中更加显著。

### 2.2 模型构建

为了验证前述假设,本文构建了一个实证模型,研

究招股说明书文本可读性与 IPO 抑价之间的关系。

$$\text{Underprice} = \beta_0 + \beta_1 \text{Readability} + \beta_2 \text{Security} + \beta_3 \text{Controls} + \varepsilon \quad (1)$$

其中, Underprice 为 IPO 抑价程度; Readability 为招股说明书文本信息可读性,包括文本长度、文本可理解性、会计术语数、转折词数四个衡量指标; Security 为网络安全关注度,衡量企业在招股说明书中披露的网络安全治理措施、合规性以及数据保护等相关内容的全面性与透明度。Controls 为控制变量。

## 3 数据获取与主模型验证

### 3.1 样本选择与数据来源

本文选取 2019 年 7 月 22 日至 2022 年 10 月 15 日在科创板上市的 480 家企业为初始样本,利用网络爬虫技术从巨潮资讯网获取科创板上市企业招股说明书。基于此,本文对初步样本进行了以下筛选:首先,考虑到金融行业的特殊性,剔除该行业的上市公司;其次,鉴于招股说明书 PDF 文档的获取与转换问题,排除了无法下载或复制的样本。然后,去除了缺失关键数据的样本;最后,剔除了因财务状况异常或其他问题而被标记为 ST 的公司。IPO 公司上市前的财务数据主要来自 CSMAR 数据库,缺失的数据则通过东方财富网和公司年报手动补充。

### 3.2 主要变量定义

#### (1) IPO 抑价程度

本文使用上市首日收益率来衡量 IPO 抑价程度,用如下公式计算:

$$UP = \frac{P_1 - P_0}{P_0} \quad (2)$$

其中,  $P_0$  表示发行价格;  $P_1$  表示收盘价格。

#### (2) 招股说明书文本信息可读性与网络安全关注

本文参考周佰成等<sup>[3]</sup>的研究方法,从多个维度评估招股说明书文本的可读性与网络安全关注度。具体来说,本文将文本长度 (Length) 作为评估招股说明书可读性的一个指标。借鉴薛爽等的思路,文本可理解性、会计专业术语、逆接成分也可作为衡量可读性的标准。文本可理解性 (Understand) 参照已有词典,统计招股说明书中含现代汉语常用词数;会计术语密度 (AT-density),会计专业术语密度越大,文本可读性越低,参照王克敏 (2018) 的做法,构建词库,统计招股说明书中会计术语数量。三是逆接成分 (Adverse),文本逆接成分越多,越难理解,统计招股说明书中逆接成分数量来衡量。本文借鉴已有网络安全词典统计招股说明书中的

网络安全相关词汇来衡量企业对网络安全关注度 (Security)。借鉴以往研究选取以下变量作为控制变量：公司规模 (Size)、公司年龄 (Age)、发行价格 (Offerprice)、中签率 (Lottery)、加权净资产收益率 (ROE)、资产负债率 (Lev)。

## 4 实证结果

### 4.1 描述性统计

本文对实证研究中的各变量进行了描述性统计分析。数据显示，科创板在中国证券市场的平均 IPO 抑价率为 136.2%，标准差为 1.405，显著高于发达国家的资本市场水平。在招股说明书披露质量方面，招股说明书总字符数均值为 363220 字，标准差为 74685，表明招股说明书篇幅普遍较长，且各公司招股说明书篇幅存在较大差异；转折词、会计术语、常用词数均值分别为 255、

16789、12773，这表明招股说明书中存在较多难以理解的会计术语词汇，较少使用汉语常用词。

### 4.2 可读性与网络安全关注的影响

本文计算了各变量的方差膨胀因子，结果显示所有计算值均低于 5。表 1 中所列为多元回归分析的结果，第 (1) 列为各控制变量与 IPO 抑价的回归结果，其中中签率 (Lottery)、资产负债率 (Lev)、发行价格与 IPO 抑价程度之间存在显著的负相关关系。公司年龄 (Age) 与 IPO 抑价程度成显著负向统计关系。表中 (2) 到 (6) 列分别表示招股说明书可读性各个指标对 IPO 抑价的影响。其中文本长度 (Length)、转折词数 (Adverse)、会计术语数 (AT) 与 IPO 抑价程度呈现显著正向统计关系，网络安全关注 (Security) 与 IPO 抑价程度呈现显著负向统计关系，验证了假设 H1 与 H2。

表 1 招股说明书可读性与网络安全关注对 IPO 抑价的影响

	(1)	(2)	(3)	(4)	(5)	(6)
Length		0.0002* (1.809)				
Adverse			0.003*** (3.858)			
AT				0.0007*** (3.847)		
Understand					-0.000	
Security						0.002** (-1.342)
控制变量	控制	控制	控制	控制	控制	控制
样本量	448	448	448	448	448	448
调整后R <sup>2</sup>	0.099	0.089	0.111	0.111	0.084	0.111
F 值	7.749	7.144	8.977	8.965	6.834	6.965

### 4.3 分样本回归结果

根据公司产权性质，本文将样本分为国有企业和非国有企业两组，进一步分析招股说明书文本可读性与不同产权性质企业 IPO 抑价的关系。实证分组回归结果如表 2 所示，其中 (1) — (4) 为非国有企业的回归结果，

(5) — (8) 为国有企业的回归结果。回归分析显示，国有企业样本中各可读性指标的回归系数均不显著，而在控制了影响抑价率的主要因素后，非国有企业样本中三项指标与 IPO 抑价之间呈现显著相关关系，这与全样本回归结果一致，验证了假设 3。

表 2 产权性质的影响

变量	非国有企业				国有企业			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Length	0.0002** (2.13)				0.0003 (1.1608)			
Adverse		0.004*** (4.34)				-0.0011 (-0.421)		
AT			0.009*** (4.069)				0.0001* (1.828)	
Understand				-0.002 (-0.89)				-0.0001 (-0.137)
控制变量	控制	控制	控制	控制	控制	控制	控制	控制
样本量	410	410	410	410	38	38	38	38
调整后R <sup>2</sup>	0.1	0.129	0.125	0.1	0.232	0.202	0.278	0.198

## 5 讨论与结论

本文以科创板上市公司为研究对象,分析了招股说明书文本信息披露质量与 IPO 抑价之间的关系,并进一步探讨了网络安全信息披露对 IPO 抑价的影响。研究结果表明,招股说明书的信息披露质量显著降低 IPO 抑价,且这一效应在非国有企业中更为突出。同时,网络安全治理及合规性信息的披露对投资者的风险评估起到了至关重要的作用,并显著降低了 IPO 抑价。

网络安全信息的披露质量在当代企业治理中愈发重要。随着数字化转型的加速,网络安全问题已成为企业长期稳定发展的关键因素。本文的实证结果表明,科创板企业在招股说明书中披露网络安全治理措施和合规性信息时,能够有效提升投资者的信任,降低企业在信息披露过程中的不确定性。这不仅有助于提升企业的透明度,也使投资者在评估企业潜在风险时更加全面,从而降低 IPO 抑价。

因此,网络安全信息披露的质量,作为一种对公司长期治理的体现,不仅关乎企业的合规性,还直接影响投资者的决策行为。企业在招股说明书中详细披露网络安全措施、数据保护能力、网络风险管理策略等信息,将有助于减少因信息不对称所带来的市场波动,并引导投资者做出更加理性的投资决策。

### 参考文献

[1] 巫岑,饶品贵,岳衡.注册制的溢出效应:基于股价

同步性的研究[J].管理世界,2022,38(12):177-202.

[2] 张飞,周孝华.招股书模糊信息对 IPO 首日收益的影响研究[J].管理工程学报,2020,34(04):34-43.

[3] 周佰成,周阔.招股说明书可读性影响 IPO 抑价了吗?[J].外国经济与管理,2020,42(03):104-117+135.

[4] Guo Haifeng, Wang Ying, Wang Bo, Ge Yuanjing. Does prospectus AE affect IPO underpricing? A content analysis of the Chinese stock market [J]. International Review of Economics and Finance, 2022, 82.

[5] 王克敏,王华杰,李栋栋,戴杏云.年报文本信息复杂性与管理者自利——来自中国上市公司的证据[J].管理世界,2018,34(12):120-132+194. DOI:10.19744/j.cnki.11-1235/f.2018.0038.

[6] Stoll, H., & Curley, A. (1970). Small Business and the New Issues Market for Equities. Journal of Financial and Quantitative Analysis, 5(3), 309-322.

[7] 董秀良,刘佳宁,满媛媛.注册制下科创板首发定价合理性及高回报成因研究[J].上海财经大学学报,2020,22(06):65-78.

[8] 包也.信息披露印象管理行为概念、方式和量化方法:文献综述[J].市场周刊,2021,34(11):29-33.