

Prospectus, Investor Attention and Stock Performance of the First Listed Companies on the Beijing Stock Exchange

Chen Han¹, Yi Wei^{1*}, Shiyue Xu¹

¹ School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China

*weiyi@email.cufe.edu.cn

ABSTRACT

Beijing Stock Exchange is China's first corporate stock exchange established with the approval of the State Council incorporated on September 3, 2021. In this paper, we studied the first 74 listed companies on the Beijing Stock Exchange on November 15, 2021 and constructed relevant textual indicators from the dimensions of prospectus and investor attention. We found that the value figures in the prospectus showed a significant positive correlation with the rise and fall of listed companies on the early days, and there is a significant negative correlation between daily stock attention and stock price. In addition, whether a company is listed in stock market for the first time has a significant positive correlation with stock price fluctuations. At the same time, it can be seen from the text linear discriminant analysis that most of the stocks that rose on the first day belonged to the software and pharmaceutical high-tech industries, and they were more inclined to disclose their own risks, employees and other basic information of the company in the prospectus. However, most of the stocks that fell on the first day belong to the image and material industries, and their financial situation is more disclosed in the prospectus.

Keywords: Beijing Stock Exchange; prospectus; investor attention; text mining

招股说明书、投资者关注度和北交所首批上市企业股价表现——基于文本挖掘的角度

韩晨¹, 卫祎^{1*}, 徐实越¹

¹ 中央财经大学管理科学与工程学院, 北京, 100081, 中国

*weiyi@email.cufe.edu.cn

摘要

北京证券交易所于2021年9月3日注册成立,是经国务院批准设立的中国第一家公司制证券交易所,受中国证监会监督管理。本文以2021年11月15日北京证券交易所上市首批74家上市企业为研究对象,以文本挖掘的方法,从招股说明书和投资者关注度两个维度出发构建了相关文本体系指标,实证分析了其与新股上市股价表现之间的关系。研究发现,招股说明书中价值数字对上市公司首日、三日和七日的涨跌幅均呈现显著正相关关系;前一日股票关注度与股票涨跌幅呈现显著的正相关,而前七日股票关注度与股票涨跌幅呈现显著的负相关;此外,公司是否为首次上市与股票涨跌幅存在显著正相关关系。同时由文本聚类结果可以看出,首日上涨的股票所属行业多为软件、医药类高科技行业,其在招股说明书中会更倾向于披露自己的风险、员工等公司基本情况;而首日下跌的股票其所属行业多为影像、材料行业,在招股说明书中更着重披露其财务情况。

关键词: 北交所; 招股说明书; 投资者关注度; 文本挖掘

1. 引言

2021 年是中国共产党成立一百周年,也是十四五开局之年,十四五规划中明确提出并且重点强调了“坚持创新驱动发展,全面塑造发展新优势”,加快发展现代产业体系,巩固壮大实体经济根基”,成强大国内市场构建新发展格局”等关键的要求。作为社会主义市场经济活力的源泉,中小企业和创新型企业每年为社会贡献了大量的方案和创新设计,是我国塑造和保持发展新优势的重要支点。但在我国的社会主义市场经济建设的过程中,中小企业却一直面临着融资难和融资贵等问题,为了促进中小企业的发展,保障创新型企业资金的来源和流动性,北交所应运而生。

北京证券交易所(简称“北交所”),于 2021 年 9 月 3 日注册成立,是经国务院批准设立的中国第一家公司制证券交易所,受中国证监会监督管理。经营范围为依法为证券集中交易提供场所和设施、组织和监督证券交易以及证券市场管理服务等业务。截止 2021 年 12 月,已经有 82 家企业在北交所挂牌交易,其中有十家为首次新上市的企业,其余皆为从新三板转移到北交所挂牌交易。新上市的十家企业市值基本都实现了大幅增长。

由投机心态驱动,北交所首日企业的抑价是受投资者关注度最高的要点之一,而 IPO 抑价受到多种因素影响。Chen 和 Zhou [1]发现 IPO 抑价受投资者情绪影响,Xu 等人[2]发现信息披露与 IPO 抑价负相关。在本文中,我们从招股说明书和投资者关注的角度研究了其对于北交所首批上市企业股价表现的影响。招股说明书是公司提交的发行上市申请文件,包含公司基本情况、风险因素、技术状况、财务状况等信息。投资者可以通过以下方式获取公司信息并做出投资决策,从而影响公司的 IPO 抑价。已有研究发现,提高信息披露质量有利于降低信息不对称程度进而缓解 IPO 抑价[3]。近年来,不少学者分别从信息含量、信息模糊性等方面,探讨了招股说明书文本“软信息”如何影响 IPO 抑价[4]。Han[5]等人通过研究科创板企业申报书的文本属性,挖掘其与企业科创板上市成功率之间的关系。然而,目前国内外鲜有文献从招股说明书可读性影响投资者信息解读成本的角度研究 IPO 市场效率问题。中文招股说明书可读性是否会通过改变投资者有效信息获取继而影响 IPO 抑价程度,这一问题亟待研究。

公众注意力衡量了一只股票受投资者关注的广泛程度,受到越多的投资者关注,那么股票的公众注意力就越高。投资者关注中一个重要的问题是投资者关注度的衡量,投资者有限关注理论的研究属于新兴领域,难点之一在于“关注”较难进行衡量[6]。Seasholes 等[7]认为涨停报道会吸引投资者关注,利用涨停板衡量作为投资者关注度,发现股票涨停事件会影响投资者注意力配置,并产生个体对涨停股票的超买行为。Yu 等人[8]利用百度指数作为投资者关注度的代理变量,研究发现投资者的有限关注能给股

票带来正向的价格压力,而这种压力会很快发生反转。

可以看出,文本属性已经成为与企业股票表现的重要研究对象之一,而北交所作为中国的最新成立的公司制证券交易所,其研究结果最能够反映其之间的关系。在本文中,我们希望通过研究北交所首批上市企业的招股说明书与其上市成功率和股价之间的关系。

2. 北交所政策解读

与上交所深交所通常上市的都是规模庞大、资源充足和信誉卓著的大企业不同,北交所设置的目的是支持优质的中小型、创新型企业的发展。与北交所相似性最大的便是早些时候设置的新三板,而北交所的政策要求在新三板的基础上更加规划化,市场准入要求相对于新三板较高,投资者门槛却降低到了 50 万元和两年以上的股市经验。反观新三板中股票质量最高、投资门槛最低的精选层也需要 100 万元的投资门槛,这无疑是投资风险降低和股票质量提升的积极信号。公司制、注册制和介于 A 股交易市场和新三板之间的交易形式和规则也基本决定了北交所的交易政策的灵活性与规范性。

对于企业申报书中高频词汇的分析有利于全面了解政策的内容、目标和方向,在接下来的内容中,我们就以北交所入市政策的三个重要的视角(帮扶中小企业、信息披露和规范化管理)为切入点,分别对高频词及其社会网络进行分析,最终得出政策实施中的要点和重点。

本节中,我们在北交所官网上下载各公司的招股说明书和北交所交易规则作为文本分析的原材料。经过数据清洗之后,通过对北京证券交易所五个角度,部门规章、发行融资、持续监管、交易管理、市场管理搜集相关政策文件。政策分析的数据来源主要是来自官网数据和学术数据库,所以数据质量较高,数据清洗工作量小,仅有两家公司数据不全,我们通过在网手动收集资料的方式补全了有关信息。在此之后,我们导入自定义财经金融词库,并进行分词,去除了停用词并且将数据进行进一步的分析和可视化操作。

通过统计关键词词频,我们得出前三十个关键词整体高频词和发行融资高频词,并绘制词云图。考虑到发行融资在企业股票发行中占有重要的地位,因此我们为发行融资相关政策做了单独的可视化分析,结合整体和发行融资部分的政策文本,得到了词频分布表和词云图如下所示:

表 1 整体高频词 (前 30)

序号	关键词	词频	序号	关键词	词频	序号	关键词	词频
1	上市公司	3353	11	上市	986	21	证券交易所	741
2	披露	2431	12	报告	968	22	情形	719
3	交易	1479	13	股份	957	23	股票	712
4	发行人	1434	14	公告	928	24	提交	635
5	发行	1214	15	审核	871	25	影响	614
6	机构	1189	16	信息	858	26	董事	613
7	文件	1094	17	业务	814	27	董事会	595
8	证券	1077	18	投资者	810	28	北京	591
9	事项	1030	19	申请	789	29	规则	590
10	事件	1006	20	证监会	751	30	意见	581

表 2 发行融资高频词 (前 30)

序号	关键词	词频	序号	关键词	词频	序号	关键词	词频
1	上市公司	1177	11	可转债	460	21	特定	316
2	发行人	1012	12	事项	443	22	配售	315
3	披露	861	13	提交	405	23	申请文件	314
4	机构	838	14	重组	386	24	申请	305
5	保荐	789	15	北京	380	25	发行股票	300
6	发行	718	16	投资者	336	26	意见	285
7	审核	704	17	信息	332	27	证券交易所	265
8	公告	565	18	主承销商	329	28	申购	264
9	上市	558	19	交易	319	29	中国证监会	263
10	文件	517	20	转股	319	30	股票	261



图 1 政策整体词云图



图 2 发行融资词云图

1. 中小企业的支持政策

在北交所上市规则和发行融资的高频词汇分析中,上市公司、发行人、事项、发行、机构、证券、上市等词汇都是一个完善的证券交易所应有的组成部分,北交所也不例外。但值得注意的是,在高频词汇表和词云图中,报告位于第十一位,公告位于第十四位而审核却位于较为靠后但仍然比较重要的第十五位,信息处在了第十六位,而中国证监会却处在了更加靠后的位置。北交所在股票上市过程中并没有对上市公司的信息进行核准与审批,而是只需要检验上市公司报告和披露信息的真实性即可,而检验公司信息的真实性只需要审核公司有关报表、公告以及文件,因此无论是文件、公告还是信息、审核等词都在文件中频繁的出现,而通常出现在核准制交易所中的关键词如审批、批准等词汇统统消失在了高频词的行列中。这既体现了北交所作为一个注册制交易所所体现出

的旗帜鲜明的特点,也体现出北交所为降低上市门槛和减小中小公司上市难度所做出的努力。

2. 对信息披露的重视体现出市场改革新特征

信息披露对任何一家企业和上市公司来说都是非常重要的环节，对于北交所来说，上市公司的质量也是至关重要的，而对上市公司的了解离不开信息的披露，相对于其他证券交易所，北交所对公司上市有着更加宽容的态度，但是正是如此就更加需要公司披露真实的信息反馈给市场，以提高市场效率和保证交易所的经营秩序。

在高频的关键词之中,披露作为仅次于上市公司的第二大关键词,在北交所的政策规则中出现频率达2400余次之多,而在发行融资关键词中也出现了高达861次,位居第三。与披露的高频出现相呼应的则是机构、事项、保荐等关键词的大量出现,这些关键词的大量出现与信息披露的大量强调都体现出了北交所期望利用市场自有的机制进行监督管理,用公开透明的信息和市场主体以及各种机构和投资者对上市公司进行有效的监督,这也体现出我国证券交易市场化改革中提高市场效率、发挥市场机制作用的一个方向所在。

3. 规范化管理

北交所设立的目的和初衷是为中小型企业提供融资服务,解决它们融资难融资贵的问题,但是相比于 A 股市场上上市公司来说,北交所设计的中小企业具有成长性强、风险大等特点,当只有在合适的引导时,中小企业才能够体现出它们成长性强的特点,在政策的春风下快速生长,但如果没有合适的规范,那么中小企业股份市场所蕴含的巨大风险就有暴露甚至爆发的危险,所以作为中小型企业融资场所的北交所虽然具有上市门槛较低、对中小型企业宽容度较大等特点,但却面临着比 A 股交易市场更大的制定规则的挑战。

在整体词汇中,保荐、申请、提交、影响、规则、意见等词语大量出现,而在发行融资高频词中,也有特定、配售、申请文件等具有规范市场行为和融资上市过程的关键词汇。这些词汇涵盖股票的上市、配股、转股甚至停牌、退市的全过程,可以看到上交所对于股票的规范化管理并不仅仅体现在上市和融资的阶段,而且还涉及到几乎整个股票的生命周期之中。而保荐、审计、意见等关键词也对应公司上市过程中券商包销或代销、会计公司审核、证监会金融监管机构监督等多个过程,在规范性上与上交所和深交所并无二致。

3. 数据及方法

3.1 数据来源

我们从北交所官网收集了上市公司的招股说明书, 从 wind 数据库中收集了这些上市公司的公司规

模、市盈率、资产负债率、第一股东持股率等财务指标数据。为了能够更好地分析投资者的情绪，我们又从今日头条网站爬取了反应投资者关注度的数据。最终，基于数据的完整性和指标的可获得性，我们最后选取了数据较完整的 68 家公司作为我们的样本数据。

3.2 回归分析

回归分析是我们常用的探究变量间相关关系的统计方法。在本课题中，我们关注所构建的数据指标对北交所上市公司股价表现的影响。为了探究所构建的数据指标对目标指标的影响，我们选用回归分析进行探究。通过回归分析并派出自变量的影响后，我们可以得知我们所构建的指标对股价涨幅是否存在影响，如果有影响其效果具体如何。

3.3 LDA 模型

LDA 是一种无监督的文本聚类模型，每一个主题为一个类，每一个文档被解释成基于各个主题的一个分布的向量（类似于模糊集合的概念），是一种软分类。运用 LDA 的方法，我们将北交所上市企业的招股说明书分为两类：首日涨的企业和首日跌的企业分别进行模型训练，来观察涨跌的招股说明书都有着怎样的特点，以此来探究北交所上市公司首日涨跌背后的具体原因。

3.4 K-means 聚类

K-means 聚类是高效的无监督指标，结合可视化技术，聚类结果可以帮助我们了解研究对象的特征。为了更好了解北交所上市公司的特点，我们从聚类的角度出发，对 80 余家首批上市的北交所公司进行 K-means 聚类，聚类的指标为具有代表性的公司财务指标，来探究传统财务指标的视角下北交所公司的主要特点。

4. 实证分析

4.1 指标介绍

本文我们希望从招股说明书和投资者关注度两个维度研究其与北交所首批上市企业股价表现的关系。具体文本指标体系如下表所示：

表 3 文本指标

指标		含义
招股说明书	文本长度	文本总字数
	特异性	文本命名实体占比
	价值数字	文本具有信息价值的数字占比
投资者关注度	上市前一	
	日投资者关注度	前一日相关企业总发帖数
	上市前七	前七日相关企业总发帖数

日投资者关注度	前七日发帖文本积极词汇占比
积极情感	前七日发帖文本消极词汇占比
消极情感	
其它	是否首次上市
	二元变量，区分新三板平移企业

4.2 描述性统计

投资者关注程度的描述统计结果如表 4 示。从图中数据中我们可以看到，投资者的积极情感指数高于消极情感指数，证明投资者基本看好此次上市公司的证券。

为了能够更加清晰地展示投资者情绪对股价走势的影响情况，我们制作了相应的可视化图表。七日内股票大体走势图如图 3 所示。我们可以看到，首次上市的股票首日全部上涨，而非首次上市的股票大部分出现了首日下跌情况，仅有少部分首日实现上涨。从首日到第三日，保持上涨态势的基本还是首次上市的股票，其余股票出现了下跌。从三日到第七日的股价走势情况与前几日相似，首次上市的股票基本保持着上涨态势，而之前出现下跌情况的股票中有一小部分转为上涨态势。

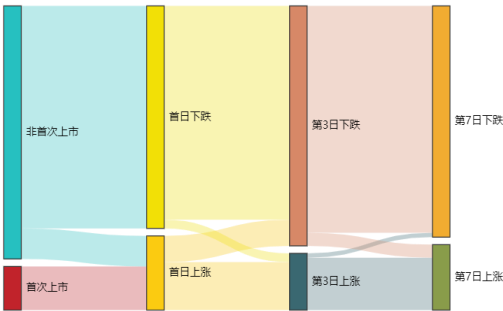


图 3 七日股票走势图

表 4 样本指标描述性统计

指标	最小值	最大值	平均值	方差
文本长度	71743.0	602178.0	250854.30	517837387
特异性	.000011	.005190	.00303815	.000
价值数字	.009286	.033582	.02171182	.000
前 1 日关注度	1.0	40.0	10.485	86.164
前 7 日关注度	1.0	47.0	13.676	129.416
积极指标	.04893	.102886	.071203	.000
消极情感	0	.0173945	.006804	.000
是否首次上市	.0	1.0	.147	.127

4.3 财务指标聚类及可视化分析

为了更好了解北交所上市公司的特点,我们利用 K-means 的方法对 80 余家首批上市的北交所公司进行公司特征的描述。其中,所选择聚类的指标为具有代表性的公司财务指标,来探究传统财务指标的视角下北交所公司的主要特点。

在指标数据中,我们对数据进行了 0-1 标准化,消除量纲和数据尺度的影响,使得结论更加可靠、有效。为了确定 K-means 最佳 K 值,在这里采用了手肘法的思想进行确定,通过计算,我们得到如下所示结果:

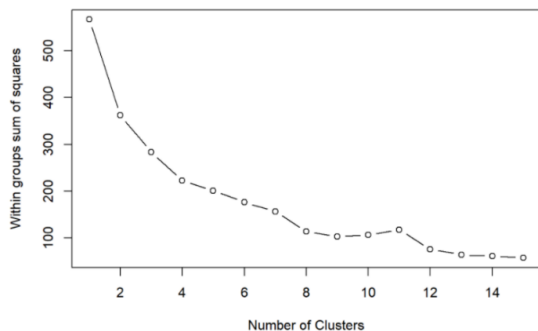


图 4 手肘法确定 k 值

根据上图,我们初步选取聚类个数为 4,并进行下一步的处理。为了更好的对聚类结果进行展示,我们利用了 R 绘制了段剖面图如图 5 所示。图中为我们展示了 4 个类样本数量,每一类各项指标与总体指标均值的大小关系。

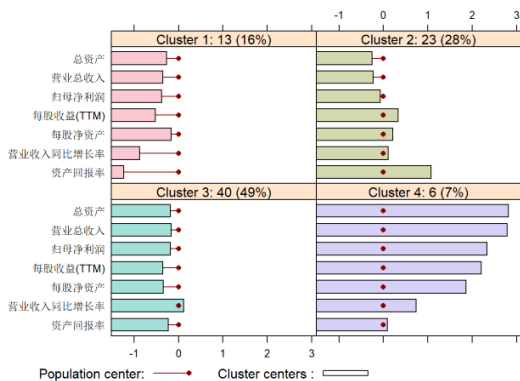


图 5 财务指标剖面图

由段剖面图我们可知,各类的指标均值与总体指标均值的相互关系。从财务的角度,我们将北交所上市公司大致分为了 4 类并直观得看出各类公司的财务特点。我们可以发现 Cluster1 中,其财务指标较差,是一些小体量,低投资回报、低收益的公司。而 Cluster2 是基本达到了整体平均水平,中等体量,但是有着非常高资产回报率的公司,这一类公司突出特点是有极高的资产回报率。Cluster3 是较小体量,但收入成长性较强的一类公司企业,虽然各指标表现都马马虎虎,但其有着较强的成长趋势。Cluster4 中是

大体量,财务指标表现非常强势的企业。

如果在传统的分析中,甚至是在量化投资的领域,类别 4 都是被策略构建者所青睐的指标,那么我们进一步去观察各类中的股价表现情况。

在 Cluster1 中,我们发现这些公司总体指标表现都低于平均水平,是一些体量较小,营业收入同比增长较低、资产回报率较低的公司,在传统意义上,这些股票都不是适合投资的股票,但我们观察这些聚类结果为 1 的公司发现了很多其首日涨幅表现非常好的公司,如志晟信息首日涨幅为 238%、恒合股份为 148.5%、中寰股份达到 123.048%等等,可见财务指标已不具反映其涨跌的代表性,探究其原因,在聚类 1 中大部分为新科技类型企业,这类企业都是高投入高成长高预期的企业,但暂时其财务状况并不是很好,但投资者依然对其有较高的期待,从而带来了良好的股价表现。在 Cluster4 中,也有不少企业其股价表现很糟糕,如长虹能源为-13.032%,颖泰生物为-6.775%,同力股份为-5.093%,但是其财务指标势头相当猛烈。可见,传统的财务指标在股价反应上存在局限性,而我们的探究视角为打破这一局限性提供了有价值的思考 and 实践。

通过聚类分析,我们将北交所的公司分为小体量企业、高回报企业、高成长性企业以及大体量企业四类,每类的具体特征我们也给出了相应的分析。可以看出,财务指标并不能有效地反映股票表现,尤其是对于北交所所针对的科技性、概念性强的企业来说,其当期的盈利状况可能不太好,但市场对其成长发展都有着非常乐观的预期。

4.3 回归分析: 申报书与市场关注和不同时间股价表现的关系

在这一部分,我们关注所构建的数据指标对北交所上市公司股价表现的影响。当探究所构建的数据指标对目标指标的影响时,我们往往有多种方法可供选择,在这里我们选用回归分析进行探究。

北交所的股价表现是关注的核心问题。在这里,我们选择了上市首日、3 日以及 7 日的涨幅。考虑到北交所上市对于股价的促进作用具有时效性,且短期内股价的涨幅集中体现投资者对一支股票的情绪倾向,其涨幅的大小反映了市场对该股票的看好程度,同时,对于价格的涨跌也是普通个体投资者、公司管理层、机构投资者与市场监督管理机构所关心的核心问题。故我们将涨幅作为回归模型中的因变量。

我们选择了 3 个时间窗口并获取其涨幅情况,分别为首日涨幅、3 日涨幅与 7 日涨幅分别构建了 3 个回归模型。这样做的目的如下:(1)不同的时间窗口反映着不同的投资者情绪情况,随着时间的增长,投资者情绪往往会逐渐趋于平静,这时,指标对涨幅的影响也许会有不同的效果;(2)增加试验次数,减小实验的偶然性,使结论更稳健。以文本指标为自变量,

财务指标为控制变量，通过回归分析我们分别得到了短期—1日、中期—3日、中长期—7日三个回归模型的结果，如下表所示。

表 5 1 日回归结果

指标	标准化系数	T	显著性
文本长度	-0.080	-0.269	0.789
特异性	-0.461	-1.340	0.186
价值数字	0.596	1.542	0.089
前 1 日投资者关注度	1.307	2.844	0.006
前 7 日投资者关注度	-1.252	-2.375	0.021
积极指标	0.229	0.454	0.652
消极情感	0.148	0.759	0.451
是否首次上市	0.866	7.860	0.000
净资产收益率(TTM)	0.151	0.808	0.422
公司规模	-0.527	-0.551	0.584
市盈率	-0.12	-1.066	0.291
资产负债率	-0.094	-0.542	0.524
第一大股东持股率	0.070	0.412	0.682

表 6 3 日回归结果

指标	标准化系数	T	显著性
文本长度	0.215	0.558	0.579
特异性	-0.342	-0.765	0.447
价值数字	1.315	2.623	0.011
前 1 日投资者关注度	1.297	2.175	0.034
前 7 日投资者关注度	-1.214	-1.774	0.082
积极指标	-0.052	-0.080	0.937
消极情感	0.164	0.650	0.519
是否首次上市	0.766	5.359	0.000
净资产收益率(TTM)	-0.039	-0.162	0.872
公司规模	-1.593	-1.283	0.205
市盈率	-0.085	-0.578	0.566
资产负债率	-0.069	0.364	0.717
第一大股东持股率	0.180	0.814	0.419

表 7 7 日回归结果

指标	标准化系数	T	显著性
文本长度	0.160	0.400	0.691
特异性	-0.516	-1.110	0.272
价值数字	1.317	2.525	0.014
前 1 日投资者关注度	1.429	2.305	0.025
前 7 日投资者关注度	-1.322	-1.858	0.069
积极指标	-0.022	0.032	0.974
消极情感	0.212	0.808	0.423
是否首次上市	0.719	4.834	0.000
净资产收益率(TTM)	0.016	0.064	0.949
公司规模	-1.520	-1.178	0.244
市盈率	-0.067	-0.436	0.664
资产负债率	-0.113	-0.574	0.568
第一大股东持股率	0.162	0.707	0.483

由回归结果可以看出，价值数字对于北交所上市公司首日、三日和七日的涨跌幅都是呈现显著的正相关。其中，价值数字是指企业招股说明书中能够为市场提供实质价值的数字的数量。因此，企业在招股说明书中提供有价值的信息越多，企业在上市初期股票走势会更为乐观。

对于关注度，前一日关注度与股票涨跌幅呈现显著的正相关，然后前七日关注度与股票涨跌幅呈现显著的负相关。可以发现随着投资者对股票关注度时间窗口的增长，关注度与股票收益率由正相关过渡到负相关关系。可以认为投资者更倾向于短期的投机行为，长期的关注度挤压不利于现时股票收益。

最后，公司是否首次上市北交所也对股票涨跌幅有着显著影响。首次在北交所上市的股票，相比于从新三板平移来的股票，股票收益率总体偏高。认为原新三板的企业，其股票价格已经经历了较长时间的市

场行为的调整，其收益率会更偏向于小于首次上市的公司股票。

4.4 LDA 主题模型

接下来，我们将从企业招股说明书的角度入手，去分析北交所股票首日涨跌的原因。企业招股说明书中详细披露了公司的情况，如营业情况、财务情况等。由前文可知，股票的首日涨跌情况最能体现市场对该股票的情绪，故而我们选择首日涨跌情况为研究目标。利用 LDA 主题模型聚类的方法，我们将北交所上市企业的招股说明书分为两类：首日涨的企业和首日跌的企业进行模型训练。最终结果如下表所示：

表 8 LDA 主题聚类结果

主题	代表词
0 基础词	披露 风险 上市 北交所
1 财务	财务 净利润 评估
2 营业	流动 生产 原材料 产品
3 审计	比率 委员会 员工 测算
4 行业类型	遥感 图像 成像 信息安全
5 行业技术	软件 生产 硬件 汽车
6 医疗行业	医疗 医药 科技 功能
7 材料行业	材料 包装 运营 北京
8 医疗用品	药物 医疗器械 灌注 过敏

表 9 文本-主题-词分布

类别	首日上涨股票	首日下跌股票
文档-主题	[0 3 1 5 8]	[1 4 0 3 7]
主题-词	[0.49 0.29 0.08 0.02 0.02]	[0.31 0.23 0.15 0.13 0.04]

从中可以看出，对于首日上涨的股票其所属行业，其更多为软件、医药类高科技行业，说明高新技术产业是北交所扶持、投资者关注的重点对象。而首日下跌的股票其所属行业多为影像、材料此类传统行业，技术门槛相对更低。除此之外，二者对于招股说明书中的信息披露也有区别，首日股票上涨的公司更倾向于披露自己的风险、员工等公司基本情况，真正尽到了充分、完备的进行信息披露；首日股票下跌的公司招股说明书中偏向于披露其财务情况，而这种信息披露并非公司信息披露准则所提倡的。

5. 结论

本文以 2021 年 11 月 15 日北京证券交易所开市首批 74 家上市企业为研究对象，以文本挖掘的方法，从招股说明书和投资者关注度两个维度出发构建了相关文本体系指标，实证分析了其与新股上市股价表

现之间的关系。其中，招股说明书中的价值数字对于北交所上市公司首日、三日和七日的涨跌幅都是呈现显著的正相关；上市前一日关注度与股票涨跌幅呈现显著的正相关，而上市前七日关注度与股票涨跌幅呈现显著的负相关；首次在北交所上市的股票，相比于从新三板平移来的股票，股票收益率更高。

LDA 聚类结果说明，首日上涨的股票其所属行业多为软件、医药类高科技行业。而首日下跌的股票其所属行业多为影像、材料行业。除此之外，对于首日股票上涨的公司更倾向于披露自己的风险、员工等公司基本情况；而首日股票下跌的公司招股说明书中着重披露其财务情况。

本文从企业招股说明书出发，结合实证分析了北交所首批上市股价表现及投资者关注度之间的关系。在未来的工作中，我们将结合更多的社交媒体文本数据和企业的申报材料进行相关的探索。

REFERENCES

- [1] Chen, P., Zhou X. (2016) Institutional Investor Private Information, Retail Investor Sentiment and IPO First-day Return. *Chinese Journal of Management Science*.
- [2] Xu, G., Ma, C., Liu, W., Jia, Y. (2018) Information disclosure and IPO first day return. *Chinese Journal of Management Science*, 26(10): 10-19.
- [3] Yao, Y., Zhao, M. (2016) Chinese styled risk disclosure, disclosure level and the market reaction. *Economic Research Journal*, 7: 158-172.
- [4] Blankespoor, E. (2019) The impact of information processing costs on firm disclosure choice: evidence from the XBRL mandate. *Journal of Accounting Research*, 57: 919-967.
- [5] Han, C., Liang, D., Huang, J., Wei, L. (2022) The impact of disclosure characteristics of the registration statement for Science and Technology Innovation Board on the IPO underpricing, *Procedia Computer Science*, 199: 238-245
- [6] Jia, C., Zhao, Y., Sun, M., Wang, B. (2010). Limited attention of investors and the liquidation of non-tradable Shares. *Journal of Financial Research*, 11:108-122.
- [7] Seasholes, M., Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance*, 14(5), 590-610.
- [8] Yu, Q., Zhang, B. (2012). Limited attention of investors and stock returns - An Empirical Study Based on Baidu Index. *Journal of Financial Research* 8:152-162.