











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在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.