

专题报告

新闻中的有限注意力和“凸显效应”因子构建——多因子选股系列研究之十七

方正证券研究所证券研究报告

分析师

曹春晓

登记编号：S1220522030005

相关研究

《日内协同股票性价比度量与“协同效应”因子构建——多因子选股系列研究之十六》
2024.03.19

《近期量化基金大幅波动原因分析及历次超额大幅回撤复盘》2024.02.22

《红利板块配置拥挤度测算及中证红利指数增强策略构建》2024.01.17

《高频因子低频化系列 2023 年表现回顾》
2024.01.14

《大市值风格指数比较分析》2023.12.28

《超预期陷阱与估值动态及“预期惯性”因子构建——多因子选股系列研究之十五》
2023.12.23

一直以来，投资者尝试利用新闻数据构建有效因子，但构建有效策略的难度较大。简单来说，策略过程中遇到了几个主要挑战，这些挑战涉及到数据量、新闻质量、打分准确性及市场定价问题。

大语言模型和情绪识别

基于智谱 GLM4 模型和新闻文本进行情感识别打分，获取大语言模型对每条新闻的情感评分，构建 glm4_news_emo 因子。回测区间内周频调仓的 glm4_news_emo 因子 IC 为 1.2%，年化 ICIR 为 1.3，多空年化收益率 15.5%，信息比率 0.95。大语言模型能够较大提升原有情感打分模型表现，在处理文本信息上具有显著优势。

新闻发布时间点的市场反馈

基于新闻发布后市场的涨跌幅，截面均值距离化后构建 news_reaction 因子，我们可以看到在周频调仓下的因子表现十分稳定，周 IC 均值在 -2.04%，年化 ICIR -2.68，IC 胜率 63%。因子多空年化收益率 20.78%，信息比率 2.14，最大回撤 -7.03%。

新闻中的注意力机制和“凸显效应”

市场的注意力始终在极度利好、极度利空两类信息上，导致利好信息被过度定价，而利空很多时候会延续跌势；市场注意力越高、热度越高的股票后市都大概率下跌。

基于 IFind 股票热度数据，等权合成到日频维度数据，周度回测下的“凸显效应”因子多空年化收益率为 33.02%，信息比率为 3.19。周度 IC 均值为 -3.13%，年化 ICIR 为 -3.32。风格剥离后得到纯“凸显效应”因子，因子多空年化收益率为 26.15%，最大回撤仅为 -4.72%，信息比率高达 3.76，周度 IC 均值 -2.44%，年化 ICIR -3.49。

热度异常时间点的动量效应

通过识别热度异常波动及其与涨跌幅方向不一致的时间点，这些时间点的涨跌幅通常具有动量效应，因此以此构建“热度异常”因子。“热度异常”因子周度 IC 为 1.37%，年化 ICIR 达到 1.97，多空年化收益 10.89%，信息比率 1.37，与主要风格因子相关性低，异质性强。

风险提示

本报告基于历史数据分析，历史规律未来可能存在失效的风险；市场可能发生超预期变化；各驱动因子受环境影响可能存在阶段性失效的风险。

正文目录

1 引言	4
2 舆情数据与因子构建	5
2.1 新闻数据介绍	5
2.2 情绪因子初步构建	6
2.3 大语言模型和情绪识别	6
2.4 新闻发布时间点的市场反馈	8
3 新闻的注意力机制和“凸显效应”	13
3.1 “凸显效应”因子	13
3.2 热度异常变化的动量效应	16
4 总结	18
5 风险提示	19

图表目录

图表 1: 新闻数据示例	5
图表 2: 新闻数量统计 (单位: 万, 2014-2024.1)	5
图表 3: 主要新闻来源新闻数量 (单位: 万, 2014-2024.1)	5
图表 4: scope_news_emo 因子十分组年化超额收益率	6
图表 5: scope_news_emo 因子 IC 统计	6
图表 6: scope_news_emo 因子全时间段 IC 表现 (2017-2024.2, 周频)	6
图表 7: glm4_news_emo 因子十分组年化超额收益率	7
图表 8: glm4_news_emo 因子 IC 统计 (2018-2020.5)	7
图表 9: glm4_news_emo 因子部分时间段 IC 表现 (月频、2018-2020.5)	7
图表 10: glm4_news_emo 和 scope_news_emo 的累积 IC 对比	7
图表 11: glm4 大模型和数库科技原始情绪打分对比	8
图表 12: 2020 年 2 月打分最高的新闻	9
图表 13: 新闻发布后观察示例	9
图表 14: news_reaction_mean 因子十分组年化超额收益	10
图表 15: news_reaction_mean 因子十分组超额曲线	10
图表 16: news_reaction 因子十分组年化超额收益率	10
图表 17: news_reaction 因子 IC 统计	10
图表 18: news_reaction 因子全时间段 IC 表现 (2014-2024.2.5, 周频)	10
图表 19: news_reaction 因子分组收益曲线 (2014-2024.2.5, 周频)	11
图表 20: news_reaction 因子和大类风格因子相关性	11
图表 21: news_reaction 因子剥离主要风格因子分组收益曲线 (2014-2024.2.5)	12
图表 22: news_reaction_std 因子十分组年化超额收益率	12
图表 23: news_reaction_std 因子 IC 统计	12
图表 24: news_reaction_std 因子全时间段 IC 表现 (2014-2024.2.5, 周频)	12
图表 25: news_reaction_std 因子分组收益曲线 (2014-2024.2.5, 周频)	13
图表 26: “凸显效应”因子十分组超额收益 (周频)	14
图表 27: “凸显效应”因子 IC 统计 (周频)	14
图表 28: “凸显效应”因子全时间段 IC 表现 (2022-2024.3.18, 周频)	14
图表 29: “凸显效应”因子分组收益曲线 (2022-2024.3.18, 周频)	15
图表 30: “凸显效应”因子和大类风格因子相关性	15
图表 31: 纯“凸显效应”因子十分组超额收益 (周频)	15
图表 32: 纯“凸显效应”因子 IC 统计 (周频)	15
图表 33: 纯“凸显效应”因子全时间段 IC 表现 (2022-2024.3.18, 周频)	16
图表 34: 纯“凸显效应”因子剥离常见风格因子分组收益曲线 (2022-2024.3.18, 周频)	16
图表 35: 热度异常变化时间点识别	17
图表 36: “热度异常”因子十分组超额收益 (周频)	17
图表 37: “热度异常”因子 IC 统计 (周频)	17
图表 38: “热度异常”因子全时间段 IC 表现 (2022-2024.3.18, 周频)	17
图表 39: “热度异常”因子剥离常见风格因子分组收益曲线 (2022-2024.3.18, 周频)	18
图表 40: “热度异常”因子和大类风格因子相关性	18

1 引言

一直以来，投资者尝试利用新闻数据构建有效因子，但构建有效策略的难度较大。简单来说，策略过程中遇到了几个主要挑战，这些挑战涉及到数据量、新闻质量、打分准确性及市场定价问题。

1) **数据量巨大且不稳定**：舆情数据源提供的信息量庞大，但其稳定性及可收集到的新闻量存在不确定性，这对数据供应商提出了较高的底层数据搜集能力、处理能力。此外，很多重复的、不及时的以及与基本面无关的新闻实际上都属于噪音，这进一步增加了因子、策略构建的难度。

2) **不同信息源的梳理**：不同的新闻来源和新闻类别对市场的影响各不相同，一些可能驱动价格动量，而另一些则可能引发价格反转。正确解读和分类这些信息源是实现策略效果的关键。

3) **新闻打分的准确性和稳定性**：是否可以准确地给新闻内容打分，以及这种打分方法的准确度和稳定性，是一个关键问题。准确的打分对于评估新闻的影响和价值至关重要。

4) **市场定价问题**：识别哪些新闻信息尚未被市场充分定价，以及预期之外新闻的识别，对投资者而言是一项挑战，如何影响市场但尚未被广泛认识到的信息，真正能推动股价上涨的信息是否真的会以新闻的形式出现？

2 舆情数据与因子构建

2.1 新闻数据介绍

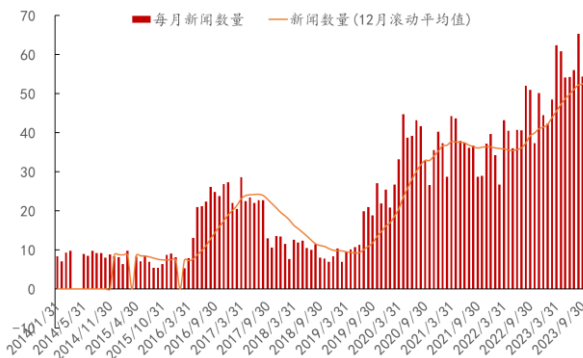
在本报告中，我们尝试使用各种方式解决引言中提到的问题，我们首先使用了数库科技公司提供的舆情数据为基础，进行一系列研究，尝试解决以上问题。图表 1 为数库科技提供的新闻数据重要字段，其中有股票代码、新闻 ID、新闻发布时间、新闻与股票相关性、文本情感打分和新闻标题文本。文本情感打分为一个情感打分分类，0 为中性概率，1 为正面概率，2 为负面概率，三个概率和为 100%。这些字段构成了新闻舆情数据的核心，帮助我们初步了解舆情数据。

图表1:新闻数据示例

stockcode	chinesename	newsid	newsts	relevance	emotiondetail	newstyle_cn
300750_SZ_EQ	宁德时代	2850951	2014-06-19T07:41:40+0800	0.0942	{0=0.8861, 1=0.0722, 2=0.0417}	路翔股份(002192)定增方案出炉 吕向阳认购4亿成实际控制人
300750_SZ_EQ	宁德时代	4007344	2015-11-13T06:47:17+0800	0.0788	{0=0.9711, 1=0.0284, 2=5.0E-4}	动力电池“白名单”发布 多家上市公司受益
300750_SZ_EQ	宁德时代	6824570	2016-12-31T02:45:00+0800	0.0833	{0=0.633, 1=0.3654, 2=0.0016}	【干货】第五批推荐目录车型配套动力电池 释放多个信号
300750_SZ_EQ	宁德时代	9990739	2017-12-30T02:03:00+0800	0.0863	{0=0.7408, 1=0.1497, 2=0.1096}	动力电池遭遇利润滑坡：后补贴时代车企腹背受敌
300750_SZ_EQ	宁德时代	11585366	2018-12-31T13:42:00+0800	0.0313	{0=0.7104, 1=0.2774, 2=0.0122}	国内自主汽车零部件产业转型升级的战略机遇期来临
300750_SZ_EQ	宁德时代	16523639	2019-12-31T22:43:08+0800	0.0164	{0=0.0018, 1=0.9975, 2=7.0E-4}	一天大事：中国12月官方制造业PMI50.2 好于预期
300750_SZ_EQ	宁德时代	27828518	2020-12-31T23:02:22+0800	0.1548	{0=0.0, 1=1.0, 2=0.0}	收购股权+增资 科创新源8000万元获得瑞泰克55%股权
300750_SZ_EQ	宁德时代	40027779	2021-12-31T23:56:15+0800	0.6	{0=0.822, 1=0.17, 2=0.008}	预计明年发布：华为、宁德时代、长安合作的阿维塔 11 细节公布
300750_SZ_EQ	宁德时代	53693079	2022-12-31T22:52:03+0800	0.1266	{0=0.783, 1=0.0833, 2=0.1338}	大卖已成定局？恒驰5两个月交付324辆
300750_SZ_EQ	宁德时代	75221309	2023-12-31T22:57:45+0800	0.0201	{0=0.0186, 1=0.9796, 2=0.0018}	大国有“数”2023，往“高”攀登，向“新”发展
300750_SZ_EQ	宁德时代	77287609	2024-01-31T23:45:01+0800	0.3214	{0=0.9201, 1=0.0585, 2=0.0215}	但斌：新能源的发展并不妨碍传统能源的需求

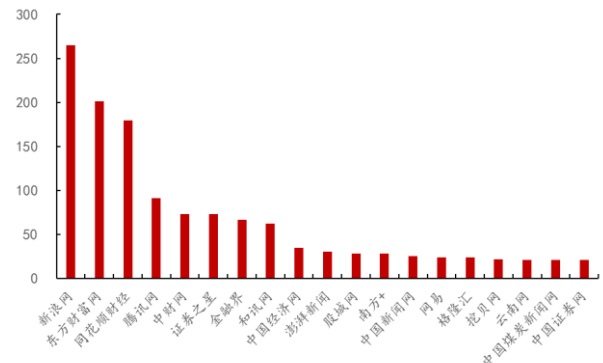
资料来源：Wind，数库科技，方正证券研究所

图表2:新闻数量统计（单位：万，2014-2024.1）



资料来源：Wind，数库科技，方正证券研究所

图表3:主要新闻来源新闻数量（单位：万，2014-2024.1）



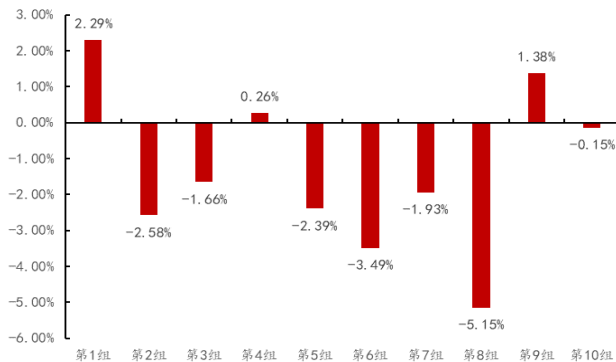
资料来源：Wind，数库科技，方正证券研究所

通过图表 2 和 3，可以观察到数库科技在过去一年中提供的月度新闻数据量已经突破 50 万条。该公司采取了高级数据清洗技术对原始文本资料进行优化处理，进而利用机器学习和深度学习算法对数据集进行评分与分类，以确保高覆盖度和数据质量。数据来源广泛，涵盖了新浪网、腾讯网和澎湃新闻等主流互联网新闻门户，以及东方财富、同花顺财经和中国经济网等专业财经网站。

2.2 情绪因子初步构建

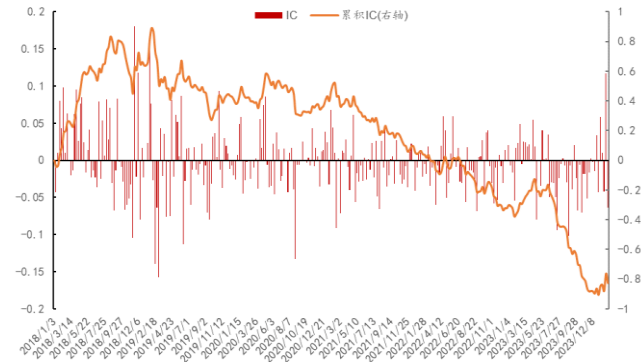
基于新闻舆情数据中的 `emotiondetail` 字段，我们计算一个简易版本的新闻情绪因子，具体构建方式为每次回看过去 7 天的新闻数据，筛选股票相关性 `relavence>0.3` 的新闻，然后计算单条新闻的 `news_emo`，具体计算方式如下，如 `emotiondetail` 为 $\{0=0.0309, 1=0.9668, 2=0.0023\}$ ，其中 0 代表中性打分概率，1 为正面打分概率，2 为负面，我们的计算方式为： $\text{factor} = 0*0.0309+1*0.9668+(-1)*0.0023$ ，再等权合成到股票维度作为新闻情绪因子，按照周度频率进行调仓测试，具体回测效果如下。

图表4: scope_news_emo 因子十分组年化超额收益率



资料来源：Wind，数库科技，方正证券研究所

图表5: scope_news_emo 因子 IC 统计



资料来源：Wind，数库科技，方正证券研究所

图表6: scope_news_emo 因子全时间段 IC 表现（2017-2024.2，周频）

因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
1.20%	0.07	1.30	1.92	0.63	15.50%	0.16	-6.00%	0.95

资料来源：Wind，数库科技，方正证券研究所

通过图表 4 和 5，我们观察到利用原始数据所提供的新闻情绪打分构建的因子在长期表现上呈现出轻微的负相关性，这种趋势不是特别显著，波动性较大。因子的十分组单调性并不突出，且 IC 值的波动较大。

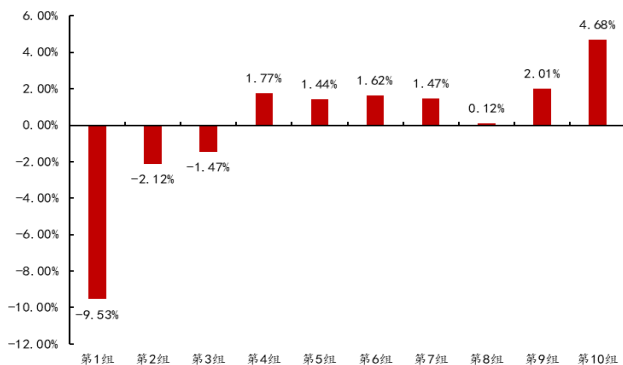
2.3 大语言模型和情绪识别

在初步构建的因子未能达到预期表现时，我们首先考虑的是情绪打分的准确性问题。考虑到对大量新闻内容进行精确的情绪评估本身具有挑战性。因此，本报告采用大语言模型对新闻文本进行情绪评分，以期提高因子的表现。

我们测试了一系列的国产大模型、openai 的 gpt3.5、4.0，最终选择了智谱 ai 开发的 GLM-4 模型。作为一款新一代认知智能大模型，GLM-4 搭载了 128K 上下文窗口，能够处理单次提示下达 300 页的文本量。这一能力在总结信息、内容抽取、复杂推理及编程等多个场景下展现了对复杂长文本的卓越处理能力。技术上，GLM-4 解决了长上下文信息处理中常见的精度下降问题，实现在 needle test 中几乎 100%的精确召回率。在多项大模型评测中，GLM-4 的性能接近或达到 GPT-4 的水平，包括 MMLU、GSM8K、MATH、BBH、HellaSwag 及 HumanEval 等重要指标。在指令跟随能力方面，GLM-4 同样展现了优异表现，测试指标达到 GPT-4 水平的 88%至 100%。出于种种考虑，我们最终采用 GLM-4 大模型进行新闻舆情打分。同样的方式，我们通过大模型对所有基础新闻文本进行情绪打分，等权合成作为因子，回看周期同样为过去 7 天，进行周频、月频回测。由于测

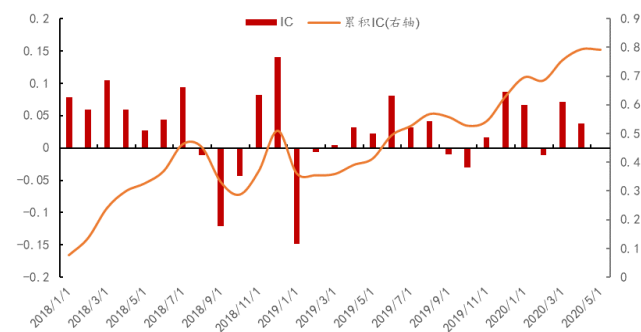
试成本等因素的限制，本报告将仅展示 2018 年至 2020 年 5 月期间大模型舆情因子的回测结果。

图表7: glm4_news_emo 因子十分组年化超额收益率



资料来源: Wind, 数库科技, 方正证券研究所

图表8: glm4_news_emo 因子 IC 统计(2018-2020.5)



资料来源: Wind, 数库科技, 方正证券研究所

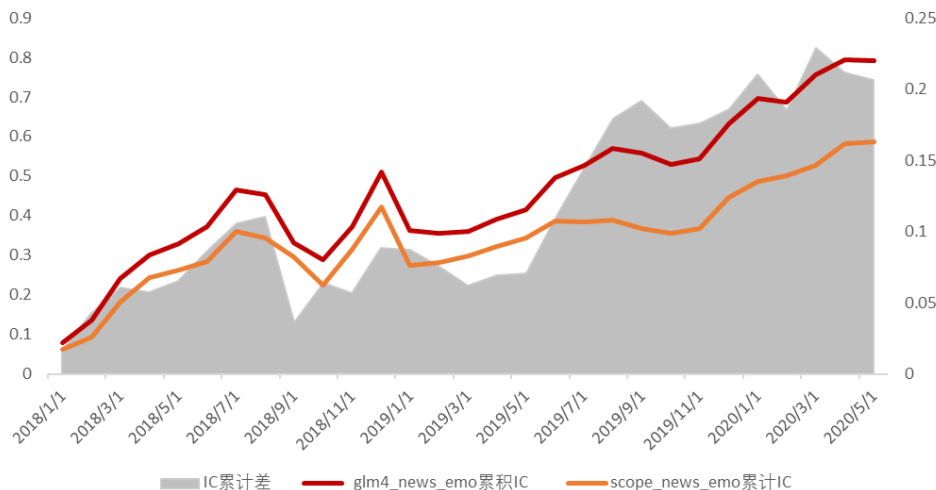
图表9: glm4_news_emo 因子部分时间段 IC 表现(月频、2018-2020.5)

因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
1.20%	0.07	1.30	1.92	0.63	15.50%	0.16	-6.00%	0.95

资料来源: Wind, 数库科技, 方正证券研究所

我们注意到，在回测期间，glm4_news_emo 因子表现出色，具有良好的单调性，且没有出现显著的回撤或效果减弱的情况，周度 IC 达到 1.2%，年化 ICIR 达到 1.3，多空年化收益率达到 15.5%，信息比率 0.95。然而，我们也必须客观地认识到，这段时间内新闻数据所展现的动量效应可能更为显著，这可能与当时的市场风格和其他相关因素有较强的关联。因此，我们不能直接得出 glm4_news_emo 在识别对股价有显著影响的信息方面，相比于数库科技提供的原始情感打分具有明显优势的结论。

图表10: glm4_news_emo 和 scope_news_emo 的累积 IC 对比



资料来源: Wind, 数库科技, 方正证券研究所

图表 10 显示，在回测期间，大模型打分生成的因子 glm4_news_emo 相较于 scope_news_emo 在 IC 表现上长期保持领先，展现出较高的稳定性及因子表现的高度相关性，这表明信息主要源自新闻文本，且未出现明显回撤。

图表11:glm4 大模型和数库科技原始情绪打分对比

人工判断	股票代码	股票名称	新闻简介	GLM4_Score	news_score	gap
数库科技打分合理	002237_SZ_EQ	恒邦股份	恒邦股份股东恒邦集团质押760万股 用于融资	1	0.00	1.00
数库科技打分合理	002155_SZ_EQ	湖南黄金	湖南黄金2019年盈利1.59亿元下滑41% 钨价格下降	1	0.04	0.96
数库科技打分合理	002450_SZ_EQ	*ST康得	乐视再亏百亿 *ST康得亏超50亿 谁是“退市种子选手”？	1	0.08	0.92
GLM4合理	000937_SZ_EQ	冀中能源	冀中能源微震技术已在冀晋蒙皖14个煤矿推广应用	0.9	0.00	0.90
GLM4合理	000710_SZ_EQ	贝瑞基因	贝瑞基因推出三代测序试剂金阻击新冠病毒变异，降低假阴率	0.9	0.01	0.89
GLM4合理	002534_SZ_EQ	杭锅股份	杭锅股份2019年净利3.6亿增长48% 减值损失同比减少	0.9	0.02	0.88
GLM4合理	688111_SH_EQ	金山办公	两亿人在家开会！马云钉钉赶超微信 云办公集体被挤爆	0.9	0.02	0.88
GLM4合理	603697_SH_EQ	有友食品	休闲食品概念股午后大涨 煌上煌股价上涨超8%	0.85	0.01	0.84
GLM4合理	200550_SZ_EQ	江铃B	江西银行向江铃汽车发放5亿元专项防疫信贷资金	0.85	0.15	0.70
GLM4合理	000333_SZ_EQ	美的集团	美的董事长方洪波：今年不裁员不降薪	0.72	0.02	0.70
GLM4合理	002409_SZ_EQ	雅克科技	雅克科技：子公司拟3.3亿元购买LG化学下属彩色光刻胶资产	0.7	0.00	0.70
GLM4合理	600887_SH_EQ	伊利股份	伊利跨界推0糖气泡水！	0.85	0.15	0.70
GLM4合理	002161_SZ_EQ	远望谷	远望谷回购条件并不充分	0.32	0.88	-0.56
GLM4合理	603605_SH_EQ	珀莱雅	上海家化总市值连续被后辈超越	0.32	0.88	-0.56
GLM4合理	688026_SH_EQ	洁特生物	洁特生物(688026.SH)2月3日股价涨108.54%提示风险	0.32	0.88	-0.56
GLM4合理	002086_SZ_EQ	ST东海洋	ST东海洋(002086.SZ)业绩快报 2019年度亏损收窄至5.53亿元	0.38	0.93	-0.55
GLM4合理	300293_SZ_EQ	蓝英装备	蓝英装备2019年净利1711万增长6% 各子公司运营效率有进一步提高	0.42	0.97	-0.55
GLM4合理	002502_SZ_EQ	*ST鼎龙	鼎龙文化预计2019年亏损收窄为7.5亿元至9亿元	0.32	0.87	-0.55
数库科技打分合理	300497_SZ_EQ	富祥药业	富祥股份2019营收同比增长16.62%至13.57亿元	0	0.99	-0.99
数库科技打分合理	002105_SZ_EQ	信隆健康	信隆健康2019年净利润4086.71万元增长274.07% 共享产品收入增长	0	0.91	-0.91

资料来源：Wind，数库科技，方正证券研究所

通过图表 11 的分析，我们进行了一项对 GLM4_Score 打分与原始情感打分差异较大实例的随机抽样检查，采用人工核查方法来探索两种打分系统之间的差异。观察结果显示，尽管两个模型在某些情况下各有其优势，但总体而言，GLM4 的打分似乎更为合理。这项发现帮助我们回答了关于原始情感打分准确性的疑问，表明为了提高新闻文本打分的精确度，采用更先进的大语言模型是构建有效因子和策略的关键步骤之一。

2.4 新闻发布时间点的市场反馈

随着新闻数量的爆炸性增长，一个不可避免的疑问浮现：实际上有多少新闻携带着真正有价值的信息，即那些可能对股价产生影响而市场尚未充分消化的信息。识别这类新闻至关重要，因为在我们的日常观察中，市场经常呈现出“利空消息已出尽，利好消息兑现即下跌”的局面。这要求我们不仅要精准地分析和理解新闻内容，还要能够评估其对股价的潜在影响，从而在众多信息中筛选出那些真正能够为投资决策带来价值的新闻。

图表12:2020年2月打分最高的新闻

股票代码	股票名称	新闻简介	GLM4_Score	news_score	gap
002973_SZ_EQ	侨银环保	一封慰问信，道出侨银“战疫情”	0.99	0.84	0.15285
600030_SH_EQ	中信证券	中信证券大集合公募化改造再下一城	0.97	0.93	0.04365
603486_SH_EQ	科沃斯	科沃斯机器人获德国iF大奖背后：中国“智”造从积淀到破局	0.97	0.85	0.1229
600690_SH_EQ	海尔智家	海尔日日顺急送物资、健康热水和饮水	0.97	0.66	0.3057
300232_SZ_EQ	洲明科技	13家机构蜂拥调研 洲明科技受机构扎堆关注	0.97	0.92	0.0513
900902_SH_EQ	市北B股	工商银行静安支行为抗疫药企提供保驾护航	0.97	0.68	0.29095
603236_SH_EQ	移远通信	移远通信正式成为沪港通标的股	0.97	0.90	0.06725
002368_SZ_EQ	太极股份	太极股份：阿里钉钉引领性云办公爆发 太极股份涨停	0.97	0.97	0.0011
002131_SZ_EQ	利欧股份	“宅经济”走红 利欧股份旗下电商直播迅猛增长	0.97	0.97	-0.00085
688186_SH_EQ	广大特材	广大特材登陆科创板 上市首日捐赠百余万驰援抗疫	0.97	0.87	0.1039
		0			
300138_SZ_EQ	晨光生物	晨光生物2019年净利1.93亿增长33% 全年辣椒红色素销量5900多吨	0.97	0.92	0.0527
603359_SH_EQ	东珠生态	【建筑-苏多永】东珠生态：战略调整业务模式，订单放量成长可期	0.97	0.79	0.17975
600604_SH_EQ	市北高新	工商银行静安支行为抗疫药企提供保驾护航	0.97	0.68	0.29095
002946_SZ_EQ	新乳业	新希望乳业：2019年营收超56亿元 同比增长14.13%	0.97	0.89	0.0845
688036_SH_EQ	传音控股	传音控股驰援武汉，捐赠500万元用于新型肺炎疫情防控工作	0.97	0.87	0.1025

资料来源：Wind，数库科技，方正证券研究所

如图表 12 所示，排名靠前的新闻大多与社会责任相关或仅描述市场行情。基于投资经验，社会责任相关新闻对股价的影响较为缓慢，而描述上涨或涨停股票的市场行情新闻在中短期内往往伴随反转效应。与此相反，能快速且持续推动股价上涨的通常是业绩预告或分析师点评等信息。但是，当所有信息等权重处理时，这些关键信息容易被噪音和不相关的信息淹没。因此，本章节通过分析新闻发布后的市场波动性和动量来捕捉市场对新闻的反馈，旨在构建一个新的因子，有效识别对股价具有实质性影响的新闻，以提高信息筛选的准确性和降低噪声影响。

为深入理解新闻发布后市场的即时反应，我们专注于短期市场表现，通过精细观察新闻后续的市场动态来揭示其对股票收益率的影响。具体而言，考虑到新闻的发布时间通常位于非交易时段，我们选择在新闻发布后的第一个交易日，计算开盘后首个 15 分钟内的股票表现，使用分钟级别频率数据来计算此时段内收益率的平均值与标准差。这种方法使我们能够准确捕捉到市场对新闻的短期反应，评估新闻影响力。

图表13:新闻发布后观察示例

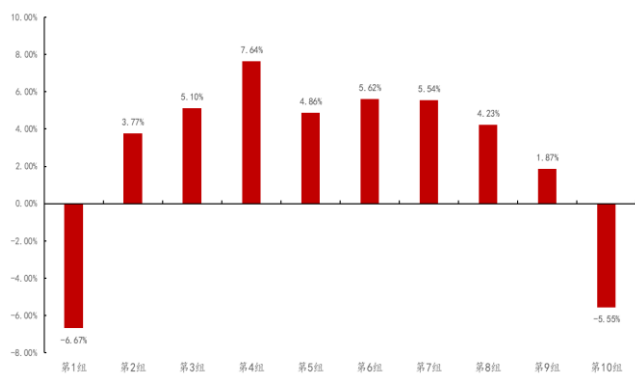
新闻ID	股票代码	新闻发生时间	新闻观察日期	观察开始时间	观察结束时间
	+08:00		2017-01-05	93000	94500
	+08:00		2017-01-05	93000	94500
	+08:00		2017-01-05	93000	94500
	+08:00		2017-01-17	93000	94500
	+08:00		2017-01-03	93000	94500
	+08:00		2017-01-03	93000	94500
	+08:00		2017-01-03	93000	94500
	+08:00		2017-01-06	93000	94500
	+08:00		2017-01-06	93000	94500
	+08:00		2017-01-06	93000	94500

资料来源：Wind，数库科技，方正证券研究所

在本研究中，我们根据以上所述，构建了 `news_reaction_mean` 与 `news_reaction_std`，旨在捕捉市场对新闻发布的短期反应。这两个因子分别代表了我们所关注时间段内股票收益率的平均值和标准差，为我们提供了对市场反应敏感度和波动性的直接测量。

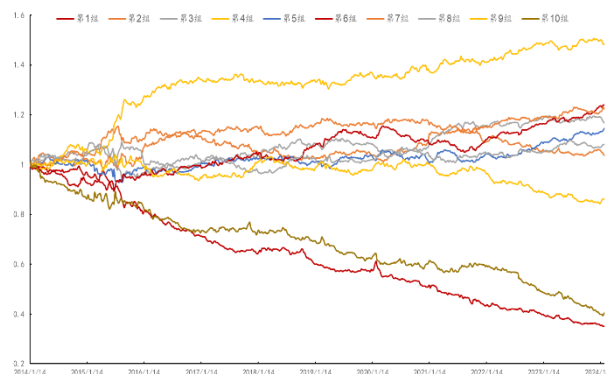
具体操作流程中，我们系统地回顾了过去 7 天内发布的所有新闻，对这些新闻进行了等权重的处理，计算得到每条新闻对应时间窗口内的收益率平均值和标准差，并将这些数据聚合到股票维度。这样的方法主要是为了获得市场对新闻即时反馈，解决“利空出尽”、“利好兑现”两种场景，新闻发布的时间点，及市场的反馈更重要，而不是新闻本身的情感打分。

图表14:news_reaction_mean 因子十分组年化超额收益



资料来源：Wind，数库科技，方正证券研究所

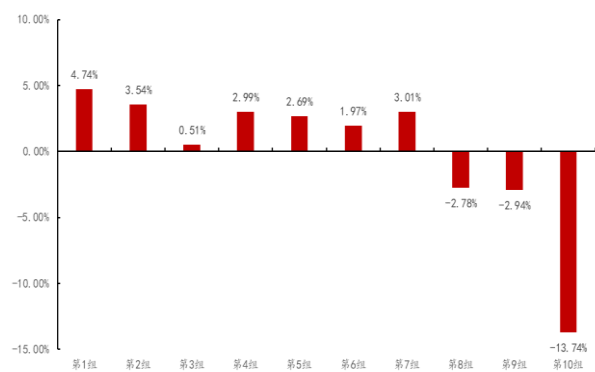
图表15: news_reaction_mean 因子十分组超额曲线



资料来源：Wind，数库科技，方正证券研究所

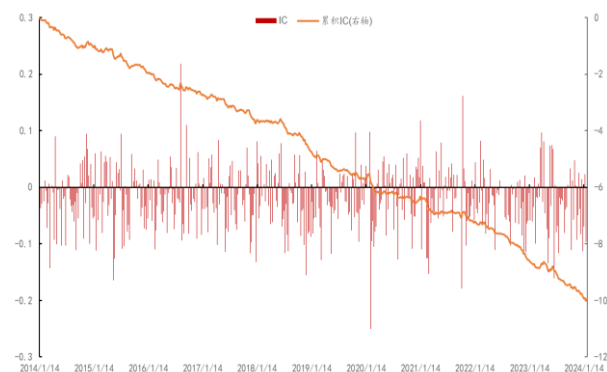
根据图表 14 和图表 15，我们可以观察到一个明显的现象，那些在新闻发布后表现出较大涨幅或跌幅的股票，其后续表现往往会进一步下跌。相反，那些在初期涨跌幅适中的股票，在后续可能会展现出更佳的表现。基于这一发现，我们决定对 `news_reaction_mean` 因子进行截面均值距离化处理，即将因子值分布的两端均纳入空头组合，称为 `news_reaction` 因子，以探索因子在预测股价未来走势中的有效性。

图表16:news_reaction 因子十分组年化超额收益率



资料来源：Wind，数库科技，方正证券研究所

图表17:news_reaction 因子 IC 统计



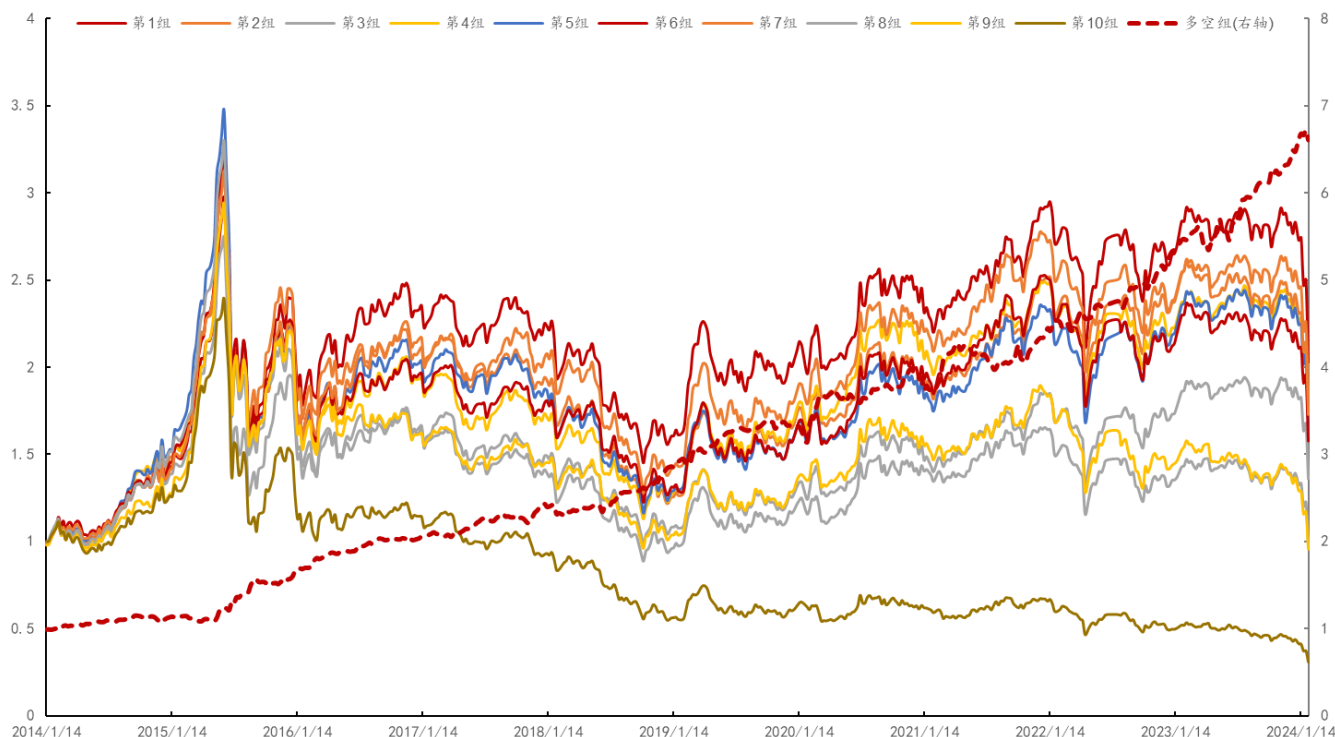
资料来源：Wind，数库科技，方正证券研究所

图表18:news_reaction 因子全时间段 IC 表现 (2014-2024.2.5, 周频)

因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
-2.04%	0.05	-2.68	-8.20	0.63	20.78%	0.10	-7.03%	2.14

资料来源：Wind，数库科技，方正证券研究所

图表19:news_reaction 因子分组收益曲线 (2014-2024.2.5, 周频)

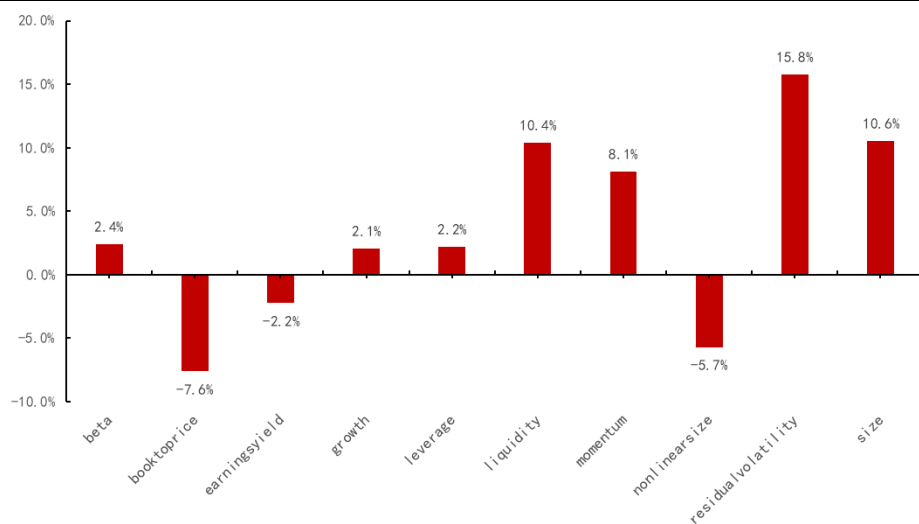


资料来源: Wind, 数库科技, 方正证券研究所

根据图表 16, 我们注意到通过对因子距离化处理后, 因子的单调性显著增强。我们将原因子的多头端和空头端都通过均值距离化后纳入空头, 这种做法基于市场中存在的一种持续性利空效应的观察, 如盈余公告效应中的空头端现象, 以及在多头端, 对于大多数普通投资者而言, 新闻发布后股票的快速上涨往往预示着随之而来的快速下跌。通过这种去极化的处理, 我们的因子更能捕捉到市场中这种持续的利空延续存在的效应, 从而提高了因子预测市场走势的准确性。

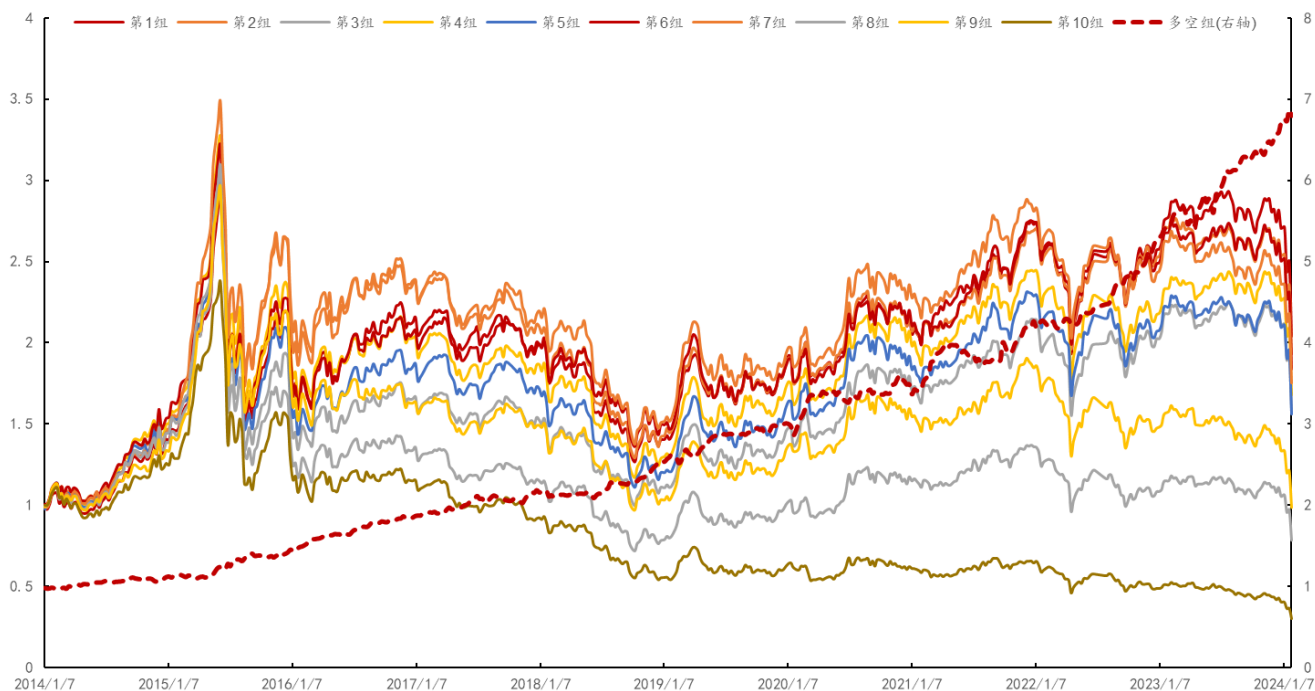
根据图表 17 和 18, 我们可以看到在周频调仓下的因子表现十分稳定, 周 IC 均值在-2.04%, 年化 ICIR -2.68, IC 胜率 63%。因子多空年化收益率 20.78%, 信息比率 2.14, 最大回撤 7%; 几乎和各个大类因子相关性较低, 剥离之后仍然有效。

图表20:news_reaction 因子和大类风格因子相关性



资料来源: Wind, 朝阳永续, 方正证券研究所

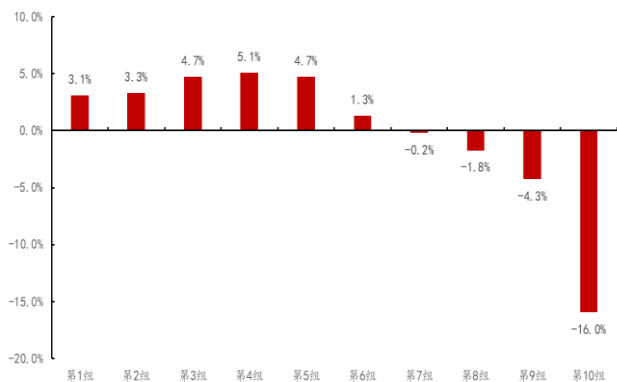
图表21:news_reaction 因子剥离主要风格因子分组收益曲线 (2014-2024. 2. 5)



资料来源: Wind, 数库科技, 方正证券研究所

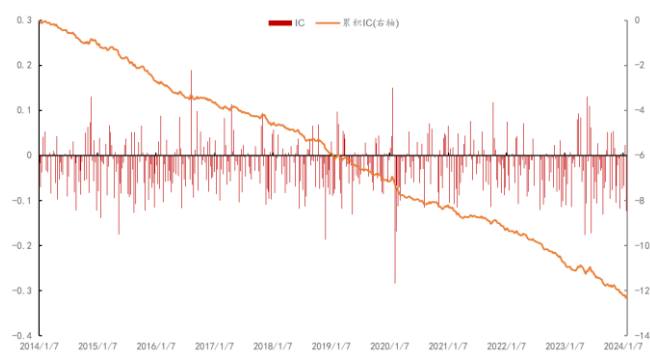
进一步我们考察新闻发布后 15 分钟内收益率的波动率, 考察市场对新闻反应的一致性, 波动率越大代表市场分歧越大, 反之代表市场分歧越小, 同样回看过一周的新闻, 等权合成每个股票的波动率, 得到 news_reaction_std 因子。

图表22:news_reaction_std 因子十分组年化超额收益率



资料来源: Wind, 数库科技, 方正证券研究所

图表23:news_reaction_std 因子 IC 统计



资料来源: Wind, 数库科技, 方正证券研究所

图表24:news_reaction_std 因子全时间段 IC 表现 (2014-2024. 2. 5, 周频)

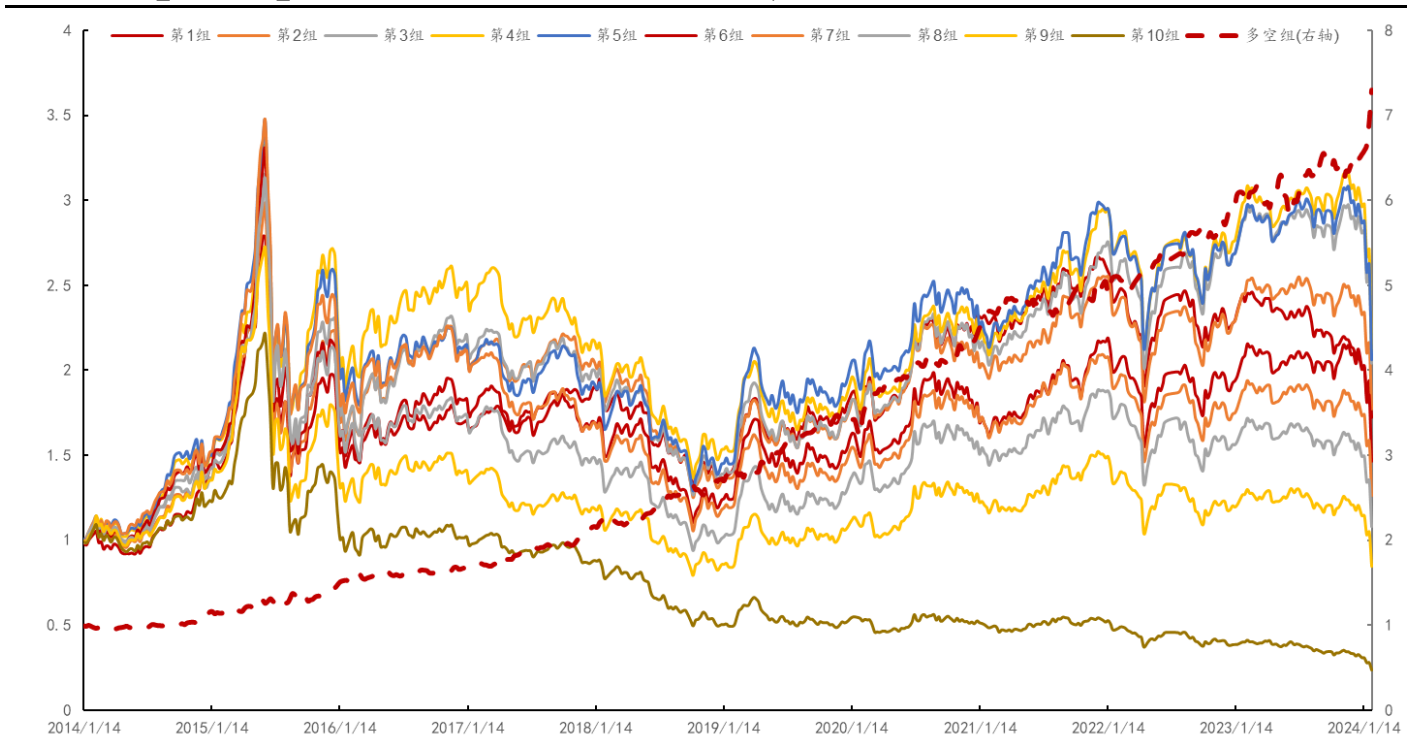
因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
-2.51%	0.06	-3.18	-9.72	0.31	21.95%	0.10	-8.75%	2.14

资料来源: Wind, 数库科技, 方正证券研究所

通过图表 21 我们观察到 news_reaction_std 因子显示出了良好的十分组收益单调性。这意味着, 如果在新闻发布后时间为 15 分钟内股票收益率的波动率较高, 这通常表明市场对某一消息的解读存在较大的分歧, 导致未来收益显著为负。这种现象可能反映了市场在短期内对信息反应的不确定性和投资者预期的分歧。另一方面, 较低的波动率并不直接预示着较高的收益。但基于这样时点的波动

率因子还是与其他普通时点测算出来的波动率因子表现类似，说明在新闻发布时点上可能存在大量噪音，即新闻存在重复发布问题。

图表25:news_reaction_std 因子分组收益曲线（2014-2024.2.5, 周频）



资料来源：Wind，数库科技，方正证券研究所

如图表 24 和 25 所示，在回测期间，news_reaction_std 因子实现了 21.95% 的多空组合收益，同时最大回撤为-8.75%。该因子主要收益来源于空头组合，与 news_reaction 因子相似，表现出多头超额收益的不显著性。此外，与主要市场风格因子的相关性低。剥离主要风格因子影响后，news_reaction_std 因子表现仍然稳定。从剥离主要风格因子相关性后，我们发现新闻发布后波动率处于中等水平的股票在多头组合中获得显著且稳定的超额收益，而空头持续有效，表明在这样的因子构建上，可能为一个非线性的因子模型。

3 新闻的注意力机制和“凸显效应”

3.1 “凸显效应”因子

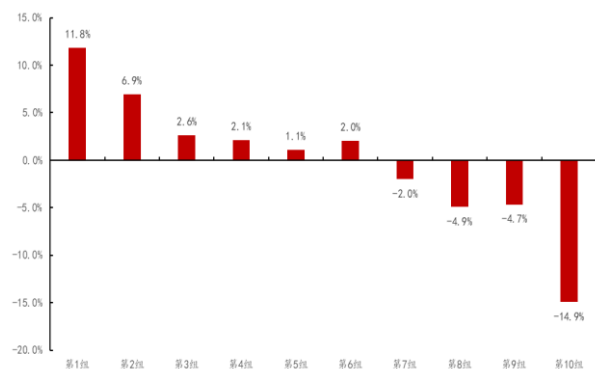
在本研究的第二节中，我们分析了目前使用新闻舆情数据时可能面临的问题、采取了多种方法尝试解决。首先，我们确保能够及时获取并处理数据，选择了能够快速提供新闻舆情数据的供应商，这方面数库科技能够较好完成基础任务。另一方面，利用先进的大型模型进一步提升了新闻文本情感打分的准确性，尝试增强选股策略表现。

我们发现一个重要的现象：市场的关注度总是有限的，且注意力总是集中在热度极高的股票：无论这种热度是由正面还是负面新闻驱动——它们的未来表现往往会展现出显著的负向收益。这一现象揭示了市场投资者的有限注意力特征，即市场倾向于对利好消息过度定价，而在利空消息出现时，机构投资者的持股情况及其减仓节奏将影响股票的未来价格走势。这强调了避免随波逐流、谨慎处理市场的热点和噪音，更多把注意力放在基本面改善加低关注度股票上。

在进一步的因子探索中，我们选用了同花顺提供的股票热度数据，一种融合了该平台新闻动态、用户论坛讨论等多元信息源的指标数据。该数据的更新频率

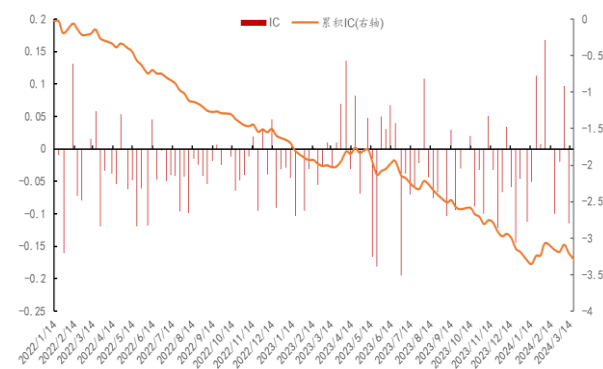
为 15 分钟，自 2022 年开始提供，目的是为了即时映射出市场情绪和投资者的关注度变化。用户互动、论坛讨论或新闻报道的任何显著变动，均能影响到热度指数的波动。基于这一数据，我们构建了“凸显效应”因子，其构建方法是将 15 分钟频率的数据等权重合成为日频数据。通过这种方式，我们旨在捕捉到股票的新闻热度水平：选择那些热度较低的股票，同时避免投资于热度较高的股票。这样的策略基于一种假设，即市场的注意力是有限的，投资者往往会被热门话题所吸引，而忽略掉那些可能具有潜在价值但当前并不受到广泛关注的股票。“凸显效应”因子正是试图利用这种市场心理现象，通过筛选出热度较低的股票来寻找未被充分发掘的投资机会。

图表26：“凸显效应”因子十分组超额收益（周频）



资料来源：Wind, Ifind, 方正证券研究所

图表27：“凸显效应”因子 IC 统计（周频）



资料来源：Wind, Ifind, 方正证券研究所

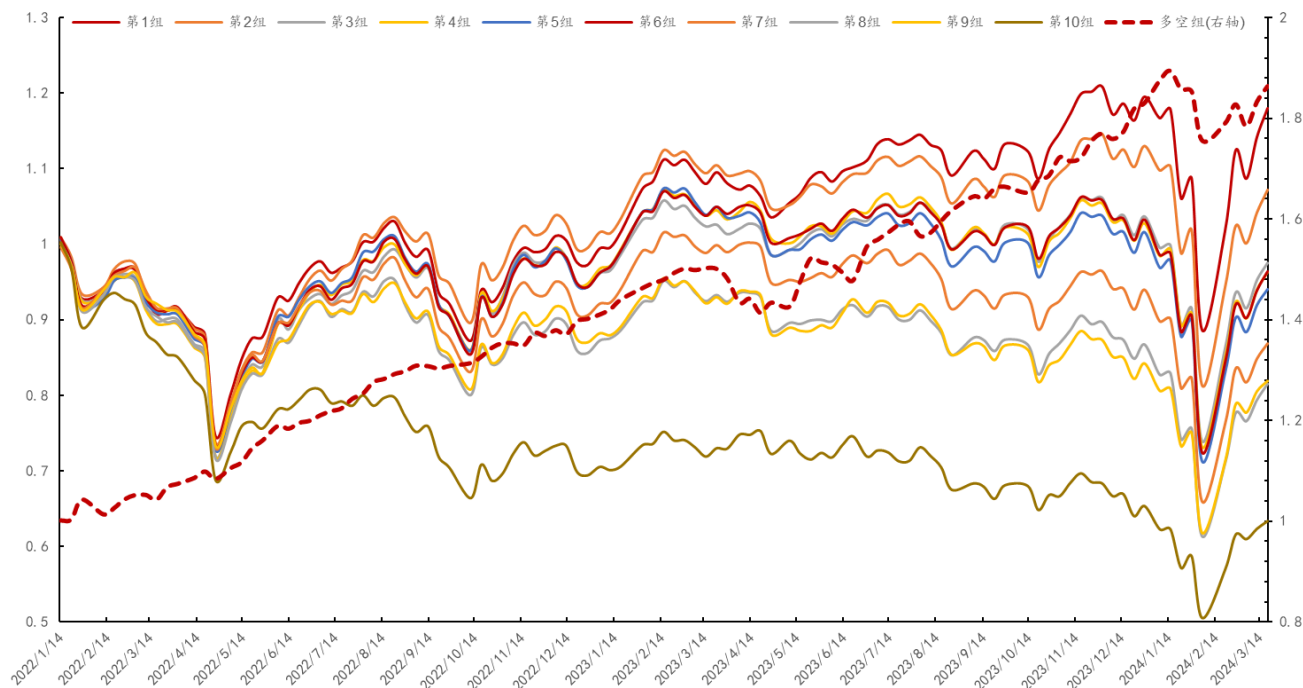
图表28：“凸显效应”因子全时间段 IC 表现（2022-2024.3.18，周频）

因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
-3.13%	0.07	-3.32	-4.71	0.72	33.02%	0.10	-7.31%	3.19

资料来源：Wind, Ifind, 方正证券研究所

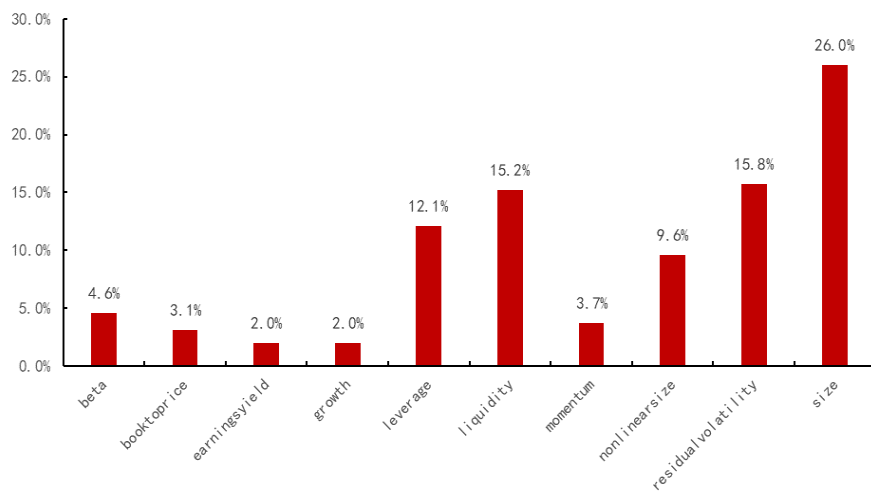
“凸显效应”因子在回测期间的表现相当出色，多空年化收益率为 33.02%，信息比率为 3.19。周度 IC 均值为-3.13%，年化 ICIR 为-3.32。这一表现突显了该因子在捕捉市场未充分反映的价值方面的有效性。然而，我们也必须承认，低热度股票往往在市值风格上存在一定的偏好，倾向于选择小市值股票。正如图表 30 所示，与主要市场风格因子的相关性分析中，与市值因子的相关性高达 26%。这一现象尤其在 2024 年春节前后的小微盘股回撤期间显得尤为明显，因子空头配置中大市值股票的倾向性增强，对收益的影响较为显著。

图表29：“凸显效应”因子分组收益曲线（2022-2024.3.18，周频）



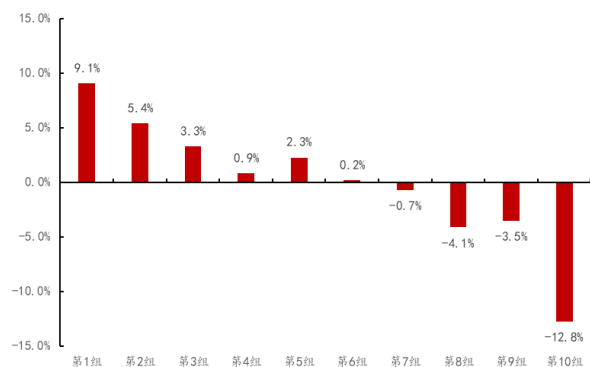
资料来源：Wind，Ifind，方正证券研究所

图表30：“凸显效应”因子和大类风格因子相关性



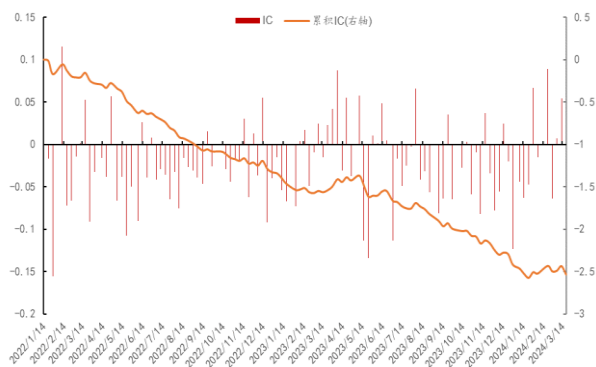
资料来源：Wind，Ifind，方正证券研究所

图表31：纯“凸显效应”因子十分组超额收益（周频）



资料来源：Wind，Ifind，方正证券研究所

图表32：纯“凸显效应”因子 IC 统计（周频）



资料来源：Wind，Ifind，方正证券研究所

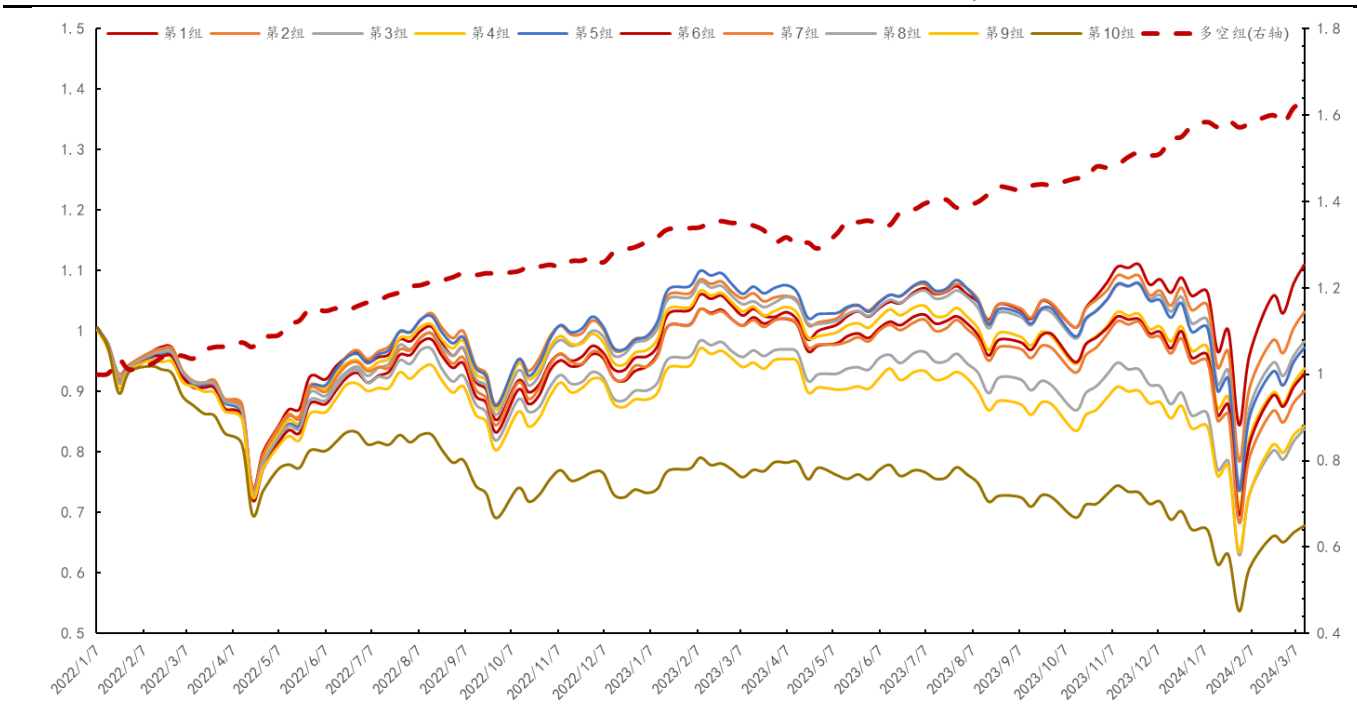
图表33: 纯“凸显效应”因子全时间段 IC 表现 (2022-2024. 3. 18, 周频)

因子绩效					十分组多空组合表现			
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率
-2.44%	0.05	-3.49	-4.99	69.8%	26.15%	0.07	-4.72%	3.76

资料来源: Wind, Ifind, 方正证券研究所

我们尝试将“凸显效应”因子剥离所有常见风格因子, 尝试提纯一些新闻关注度、热度的收益来源, 以此避免受到风格因子表现波动影响。回归主要风格因子后的“凸显效应”因子表现优异, 年化多空收益率为 26.15%, 最大回撤仅为-4.72%, 信息比率高达 3.76。特别在 2024 年春节前后, 面对小微盘股回撤, 该因子表现稳健, 多空净值创新高。周度 IC 均值-2.44%, 年化 ICIR-3.49, 周胜率 69.8%, 证明了纯“凸显效应”因子的有效性和稳定性。

图表34: 纯“凸显效应”因子剥离常见风格因子分组收益曲线 (2022-2024. 3. 18, 周频)

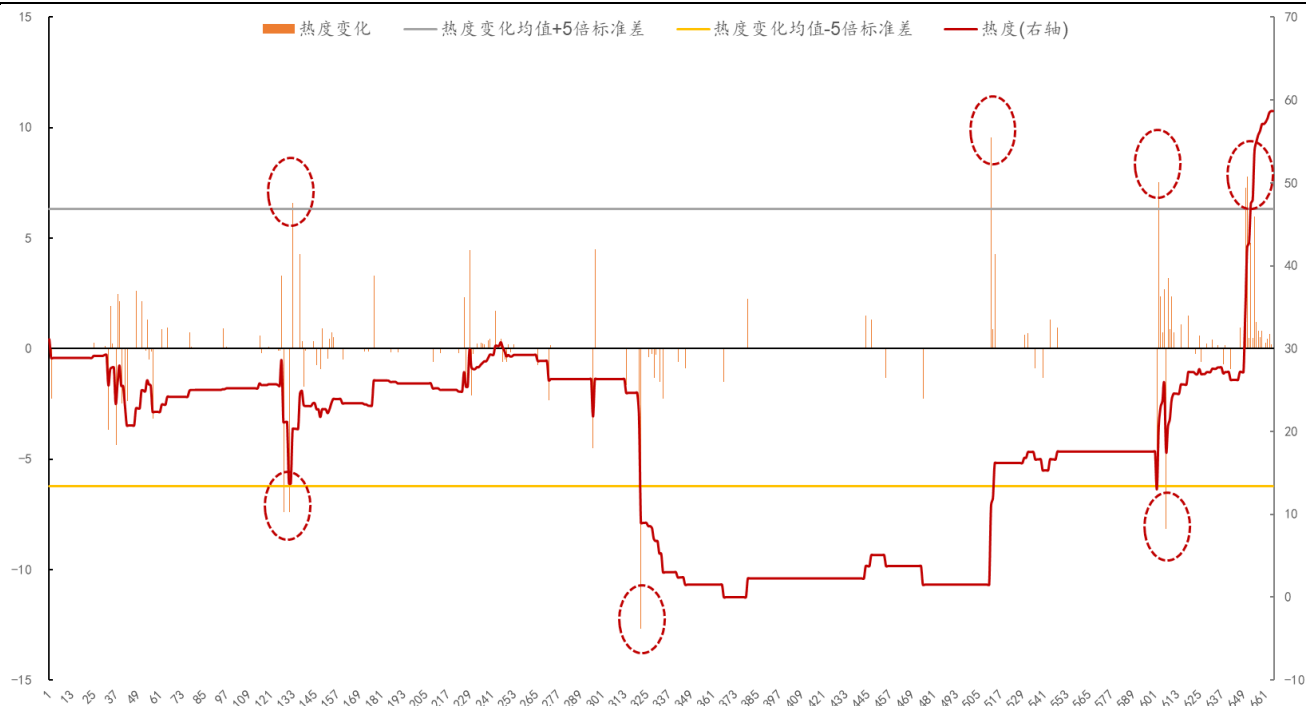


资料来源: Wind, Ifind, 方正证券研究所

3.2 热度异常变化的动量效应

在本研究中我们转向探讨股票热度数据的另一种应用, 即通过捕捉热度的显著变化点来识别潜在的市场机会。与“凸显效应”因子构建方法不同, 我们希望找到一些热度突然提升、下降的时刻, 结合量价提取特征。具体而言, 通过计算每只股票过去一周内热度数据的 15 分钟频数据变化差值及其统计特征。基于此, 我们定义了异常热度变化的阈值为均值加减 5 倍标准差, 以此来界定热度的显著升高或降低时刻。核心逻辑在于识别那些可能因为特定新闻事件或市场情绪变化而引起热度异常波动的股票。

图表35: 热度异常变化时间点识别

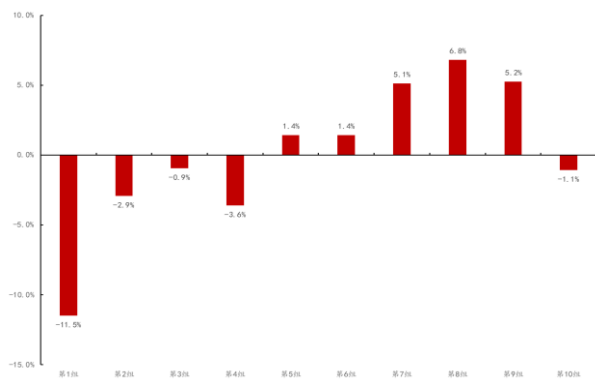


资料来源: Wind, Ifind, 方正证券研究所

新闻、热度数据可以表征为个人投资者对股票关注度的指标。根据有限注意力机制, 个人投资者倾向于将其注意力聚焦于极端利好或利空的消息上。因此, 当股票的涨跌方向与热度变化方向不一致时, 应引起投资者的高度警觉。特别是在热度上升而股票价格下跌的情况下, 这往往预示着可能的利好消息已经被市场充分消化。

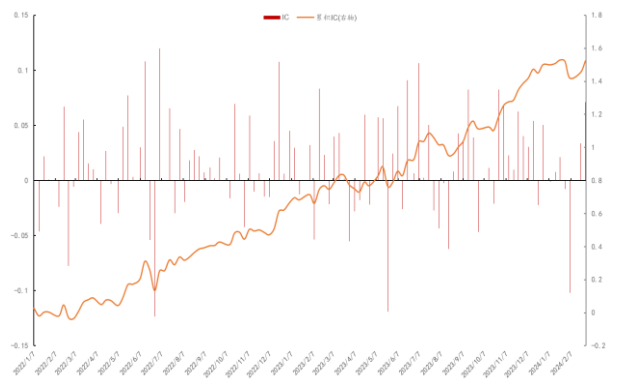
在分析热度异常变化的时间段中, 进一步筛选出那些市场涨跌幅与股票热度变化方向不一致的时间段。我们根据这些特定时间段内的股票涨跌情况, 计算出相关时间段的涨跌幅。依此为基础, 通过回顾过去一周的数据, 我们计算出这些特殊时间段涨跌幅的平均值, 从而构建了“热度异常”因子。

图表36: “热度异常”因子十分组超额收益 (周频)



资料来源: Wind, Ifind, 方正证券研究所

图表37: “热度异常”因子 IC 统计 (周频)



资料来源: Wind, Ifind, 方正证券研究所

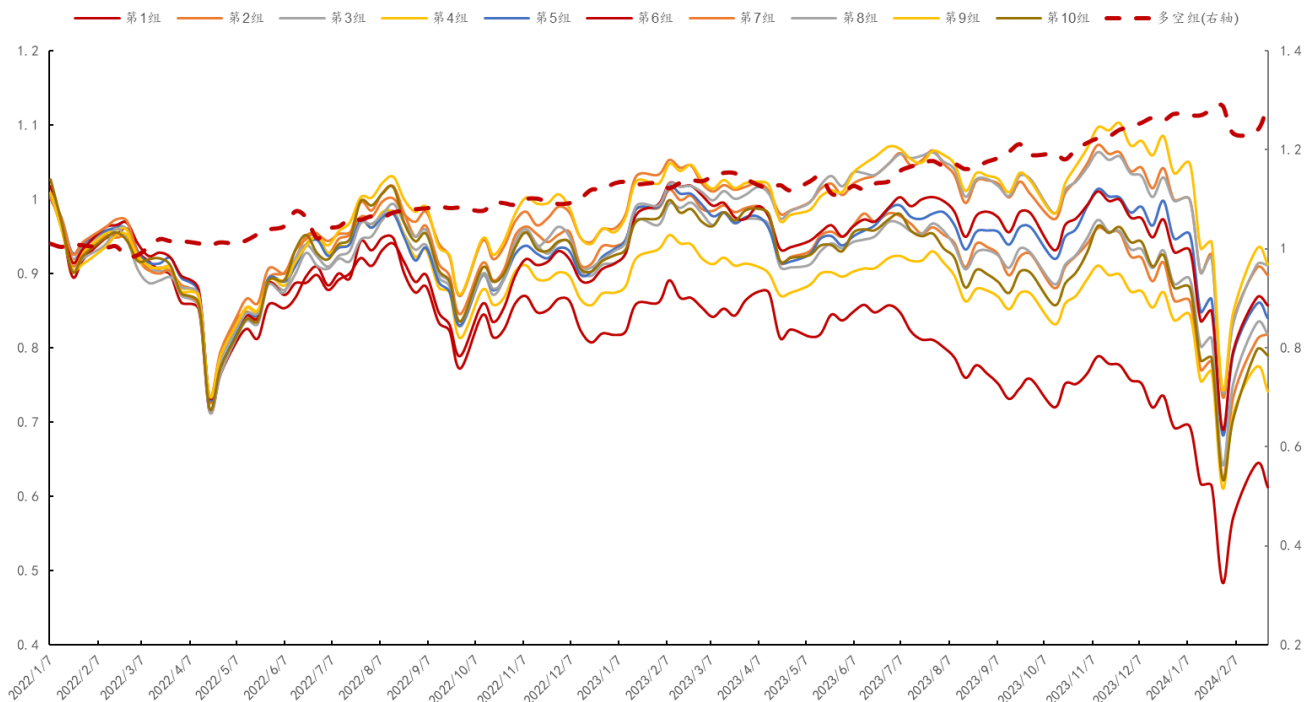
图表38: “热度异常”因子全时间段 IC 表现 (2022-2024. 3. 18, 周频)

因子绩效					十分组多空组合表现				
IC均值	IC标准差	ICIR	IC T值	IC胜率	因子年化多空收益率	波动率	最大回撤	信息比率	
1.37%	0.05	1.97	2.79	0.66	10.89%	0.08	5.00%	1.37	

资料来源: Wind, Ifind, 方正证券研究所

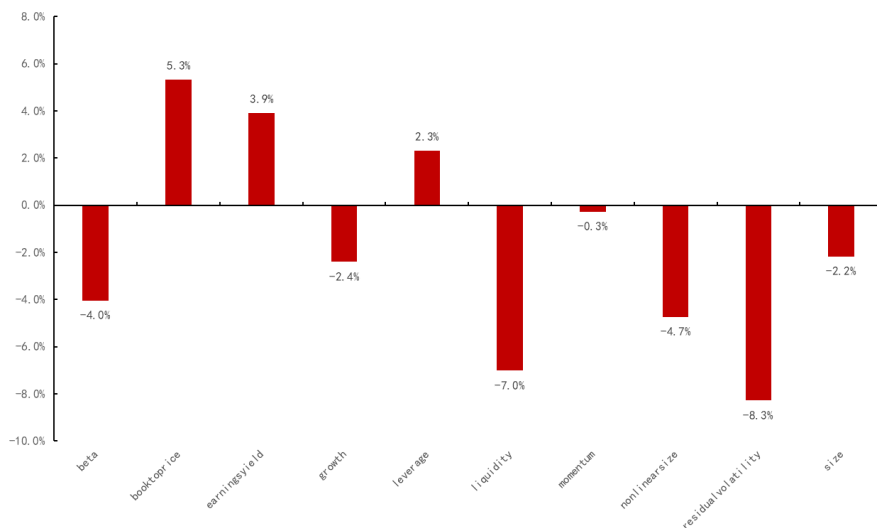
“热度异常”因子在回测期间展现了 10.89% 的多空年化收益率，信息比率达到 1.37，主因在于该因子并非呈完全单调线性，第十分组的收益表现一般。但从总体上看，因子仍呈现出一定的动量趋势。周度 IC 为 1.37%，年化 ICIR 达到 1.97，显示出较好的表现。此外，该因子与主要风格因子的相关性较低，展现了较强的特异性。

图表39：“热度异常”因子剥离常见风格因子分组收益曲线（2022-2024.3.18，周频）



资料来源：Wind，Iifind，方正证券研究所

图表40：“热度异常”因子和大类风格因子相关性



资料来源：Wind，Iifind，方正证券研究所

4 总结

一直以来，投资者尝试利用新闻数据构建有效因子，但构建有效策略的难度较大。简单来说，策略过程中遇到了几个主要挑战，主要涉及到数据量、新闻质量、打分准确性及市场定价问题。

基于智普 GLM4 模型和新闻文本进行情感识别打分，获取大语言模型对每条新闻的情感评分，构建 glm4_news_emo 因子。回测区间内周频调仓的 glm4_news_emo 因子 IC 达到 1.2%，年化 ICIR 达到 1.3，多空年化收益率 15.5%，信息比率 0.95。大语言模型能够较大提升原有情感打分模型表现，在处理文本信息上具有显著基于新闻发布后市场的涨跌幅，截面均值距离化后构建 news_reaction 因子，我们可以看到在周频调仓下的因子表现十分稳定，周 IC 均值在-2.04%，年化 ICIR -2.68，IC 胜率 63%。因子多空年化收益率 20.78%，信息比率 2.14，最大回撤-7.03%。

市场的注意力始终在极度利好、极度利空两类信息上，导致利好信息被过度定价，而利空很多时候会延续跌势；市场注意力越高、热度越高的股票后市都大概率下跌。基于 IFind 股票热度数据，等权合成到日频维度数据，周度回测下的“凸显效应”因子多空年化收益率为 33.02%，信息比率为 3.19。周度 IC 均值为-3.13%，年化 ICIR 为-3.32。风格剥离后得到纯“凸显效应”因子，因子多空年化收益率为 26.15%，最大回撤仅为-4.72%，信息比率高达 3.76，周度 IC 均值-2.44%，年化 ICIR-3.49。

通过识别热度异常波动及其与涨跌幅方向不一致的时间点，这些时间点的涨跌幅通常具有动量效应，因此以此构建“热度异常”因子。“热度异常”因子周度 IC 为 1.37%，年化 ICIR 达到 1.97，多空年化收益 10.89%，信息比率 1.37，与主要风格因子相关性低，异质性强。

5 风险提示

本报告基于历史数据分析，历史规律未来可能存在失效的风险；市场可能发生超预期变化；各驱动因子受环境影响可能存在阶段性失效的风险。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

本研究报告由方正证券制作及在中国（香港和澳门特别行政区、台湾省除外）发布。根据《证券期货投资者适当性管理办法》，本报告内容仅供我公司适当性评级为 C3 及以上等级的投资者使用，本公司不会因接收人收到本报告而视其为本公司的当然客户。若您并非前述等级的投资者，为保证服务质量、控制风险，请勿订阅本报告中的信息，本资料难以设置访问权限，若给您造成不便，敬请谅解。

在任何情况下，本报告的内容不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求，方正证券不对任何人因使用本报告所载任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告版权仅为方正证券所有，本公司对本报告保留一切法律权利。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。如需引用、刊发或转载本报告，需注明出处且不得进行任何有悖原意的引用、删节和修改。

评级说明：

类别	评级	说明
公司评级	强烈推荐	分析师预测未来12个月内相对同期基准指数有20%以上的涨幅。
	推荐	分析师预测未来12个月内相对同期基准指数有10%以上的涨幅。
	中性	分析师预测未来12个月内相对同期基准指数在-10%和10%之间波动。
	减持	分析师预测未来12个月内相对同期基准指数有10%以上的跌幅。
行业评级	推荐	分析师预测未来12个月内行业表现强于同期基准指数。
	中性	分析师预测未来12个月内行业表现与同期基准指数持平。
	减持	分析师预测未来12个月内行业表现弱于同期基准指数。
基准指数说明		A股市场以沪深300 指数为基准；香港市场以恒生指数为基准，美股市场以标普500指数为基准。

方正证券研究所联系方式：

北京：西城区展览馆路 48 号新联写字楼 6 层

上海：静安区延平路71号延平大厦2楼

深圳：福田区竹子林紫竹七道光太银行大厦31层

广州：天河区兴盛路12号楼隼峰苑2期3层方正证券

长沙：天心区湘江中路二段36号华远国际中心37层

网址：<https://www.foundersc.com>

E-mail：yjzx@foundersc.com