

投资者舆情与上市公司股价财务关联研究 ——基于 DeepSeek 大模型情感分析

陈云

神州能源集团股份有限公司，重庆 401329

DOI:10.61369/ER.2025060012

摘 要： 本研究构建了一个基于轻量级大语言模型（DeepSeek）的细粒度舆情量化框架，对东方财富中文评论进行连续型情感评分与关联度过滤，生成日度加权情绪指数。基于2024年10只A股日频数据，结合计量与机器学习方法检验情绪对股价的非线性影响及时效特征。结果表明，社交媒体情绪与股价变动呈显著非线性关联，影响集中于短期（领先1-3天），中长期预测性能下降但存在动态波动；个股情绪敏感性存在异质性，科技类及高关注公司反应更敏感。本研究在方法上实现了基于大模型的细粒度情绪量化，在视角上融合市场反应与时间效应，揭示了情绪与股价间的非线性短周期机制，为后续融合多源数据探索情绪传导路径提供基础。

关 键 词： 投资者情绪；股票财务指标预测；大语言模型；机器学习

Research on the Financial Correlation Between Investor Sentiment and Stock Prices of Listed Companies—Based on Sentiment Analysis Using the DeepSeek Large Model

Chen Yun

Shenzhou Energy Group Co., Ltd., Chongqing 401329

Abstract： This study constructs a fine-grained sentiment quantification framework based on a lightweight large language model (DeepSeek) to assign continuous sentiment scores and apply correlation filtering to Chinese comments on Dongfang Stock Bar, generating daily weighted sentiment indices. Using daily frequency data from 10 A-share stocks in 2024, it examines the nonlinear impact and temporal characteristics of sentiment on stock prices through econometric and machine learning methods. The results indicate a significant nonlinear relationship between social media sentiment and stock price movements, with effects concentrated in the short term (1–3 days ahead), while medium- to long-term predictive performance declines but exhibits dynamic fluctuations. There is heterogeneity in sentiment sensitivity across individual stocks, with technology-sector and highly followed companies showing more pronounced reactions. This study achieves fine-grained sentiment quantification based on large models methodologically and integrates market reactions with temporal effects from a perspective, revealing the nonlinear short-cycle mechanism between sentiment and stock prices. It provides a foundation for subsequent research exploring sentiment transmission pathways by incorporating multi-source data.

Keywords： investor sentiment; stock price financial indicator prediction; large language model; machine learnin

一、文献综述

（一）投资者情绪的理论基础与度量演进：从市场整体到社交媒体的深度挖掘

投资者情绪研究的理论起点建立在对经典“有效市场假说”（Malkiel, 1989）的反思之上，并根植于 Simon（1990）的“有限理性”框架，该框架为情绪影响资产定价提供了核心行为基础。早期研究主要依赖封闭式基金折价率等市场整体代理变量，其间接性难以精准捕捉情绪的微观变化。互联网社交媒体的兴起，推

动了情绪度量方法的革命性转变，即从宏观代理指标转向对海量个体文本的直接挖掘。例如，McGurk 等人（2020）利用社交媒体文本构建情绪指标，证实了其对股票异常收益的预测能力。^[1]在中国市场背景下，Sun 等人（2021）通过构建专用金融情感词典与分析东方财富股吧海量帖子，开发了高频、细粒度的个体投资者情绪指数“GubaSenti”，其有效性超越了传统指标^[2]。这一系列研究（如 Liu 等人，2023；Xie 等人，2023）共同揭示了股价与投资者情绪之间存在显著的协同互动关系，为基于情绪的实证研究奠定了方法论基础^[3]。

（二）投资者情绪对股市的影响机制：收益、波动与预测模型

大量实证研究从多维度揭示了投资者情绪对股票市场的复杂影响机制。在资产定价与收益方面，情绪已被证实为一个重要的系统性因子。Zheng（2022）的研究表明，将情绪因子纳入中国 A 股的五因子模型能显著提升其解释力；王春（2014）基于基金资金流的实证发现情绪效应存在显著的市场异质性；陈其安等人（2012）则从理论模型出发，论证了不同情绪状态如何通过改变投资者的最优资产配置决策，进而系统性影响市场的风险-收益关系。在波动性方面，Xie 等人（2023）利用混频数据模型揭示了中国 A 股市场中，高频投资者情绪是驱动股票波动的重要力量^[4]。近年来，融合情绪指标的预测模型成为前沿。于孝建等人（2024）将文本情绪分析融入 LSTM 神经网络，提升了股指预测性能。更进一步的探索开始利用多模态数据（如 Obaid & Pukthuanthong, 2022 对新闻图片的分析）及分析社交机器人的市场操纵机制（Yu 等人，2025）^[5]。同时，大语言模型的应用正拓展至投资者心理剖析与教育领域（Xie 等人，2025），展现了新的研究潜力。

（三）现有研究不足与本研究的创新

既有研究存在三方面局限：一是情绪量化粒度不足，多依赖离散分类（如 McGurk et al., 2020; Sun et al., 2021），缺乏连续强度度量与噪声过滤；二是时间效应检验单一，鲜有系统分析情绪影响随领先期演变的非线性动态（如 Liu et al., 2023）；三是方法融合不深，计量与机器学习模型多独立应用^[6]。

针对上述不足，本研究创新在于：

- 方法上，应用 DeepSeek 大模型，通过提示词工程实现中文股吧评论的连续型情感强度评分与高精度关联度过滤，达成细粒度、低噪声的情绪度量^[7]。
- 视角上，系统检验多领先窗口（1 至 30 天），揭示情绪影响的短期主导、中期衰减与长期微弱回升的复杂时效异质性^[8]。
- 技术融合上，整合 OLS、随机森林与 XGBoost 进行交叉验证，增强情绪与股价间非线性、短周期关系结论的稳健性。

二、数据来源

本文的数据基础来源于对 Github 开源项目 weiyu，该项目

本身提供了一个灵活的爬虫框架与舆情采集基础代码，本文在此基础上，针对东方财富吧这一特定金融社交平台的数据结构与反爬机制，专门设计并实现了一个自动化 Agent 进行东方财富吧数据收集，最终收集到 167322 条数据^[9]。

三、情感打分模型

本文基于大语言模型构建了一套双层次量化分析框架，旨在对上市公司股票的网络评论进行精准的情感与关联度评估。

首先，情感分析层面，本文设计了一个包含 11 个独立维度的综合评价体系。其中，乐观性、信心度、满意度、希望感、兴奋感 5 个正向维度（评分 1-5 分）用于量化积极情绪强度；中立性维度（0 分）标识客观陈述；悲观性、担忧感、失望感、愤怒感、恐惧感 5 个负向维度（评分 -1 至 -5 分）则系统性地衡量各类消极情绪。本文通过专门调优的 Prompt 指令，引导大模型基于每条评论的整体语义，综合考量这 11 个维度，输出 -5 到 5 之间的综合情感分数^[10]。其次，为确保分析对象的纯净性与相关性，本文还设计了独立的关联度打分 Prompt 与从 1 分（垃圾 / 机器人评论）到 10 分（极强关联）的评分标准。

最后，在数据融合与指标构建阶段，本文以日为单元进行聚合计算当日情绪值平均值，以当日情绪除以平均值进行加权，最终乘以关联度指标计算得到能够代表整体评论情绪的每日加权情绪指标。

四、实验结果

（一）变量描述性统计

本研究选取了如下 10 家上市公司作为样本，包括科技、金融、医药、新能源等不同行业，数据时间范围为 2024 年全年，涵盖日评论情绪指数和日股价数据等关键指标，先将所有样本变量合并进行分析^[11]。

表 1 10 支上市公司股票名称及代码

立讯精密	牧原股份	东方财富	阳股华泰	宁德时代
002475	002714	300059	300121	300750
ST 新潮	渤海化学	首创证券	寒武纪	上纬新材
600777	600800	601136	688256	688585

表 2 变量描述性统计

变量	均值	标准差	最小值	最大值
日评论情绪指数	0.0118	0.5907	-4.3920	4.7900
日评论数量	67.6066	94.3323	3.0000	1275.0000
日股价收益率 (%)	0.1699	3.4331	-17.5121	20.4064
成交量	17608769.3337	39163211.8205	0.0000	527247103.0000
换手率 (%)	2.4639	4.0485	0.0000	34.6600

注：因为有非交易日，所以成交量和换手率可能为 0

为深入探究各变量之间的相互关系，本研究采用 Pearson 相关系数进行相关性分析，并通过热力图直观展示变量间的关联程度。相关性分析不仅有助于理解变量间的线性关系，还能为后续的回归分析提供重要参考^[12]。

从相关性热力图可以看出：

- 日评论情绪指数与股价收益率存在一定的正相关性，相关系数在 0.05-0.10 之间，表明投资者情绪对股价有一定的正向影响，但影响程度有限。
- 评论数量与日评论情绪指数有较高的相关性，相关系数约

为0.3–0.5，说明评论量多的日期情绪信号更为强烈，评论量可作为情绪强度的辅助指标^[13]。

• 成交量与股价收益率的相关性较高，相关系数约为0.2–0.3，符合市场预期，表明交易量是反映市场活跃度和价格变动的重要指标。

• 换手率与成交量呈现高度正相关，相关系数接近0.9，这是由于两者均反映了股票的交易活跃度，存在较强的共线性。

• 日评论情绪指数与成交量、换手率的相关性较弱，相关系数均低于0.1，说明情绪指标与交易活跃度指标之间的直接关联较小。

• 所有变量之间的相关性均通过了统计显著性检验（ $p<0.05$ ），表明这些关联关系在统计上是可靠的。

日评论情绪指数与股价收益率的正相关性支持了我们进一步探讨情绪对股价预测能力的研究假设，最后我们选择成交量的归一化数据乘以股价作为股票财务价值关联指标。

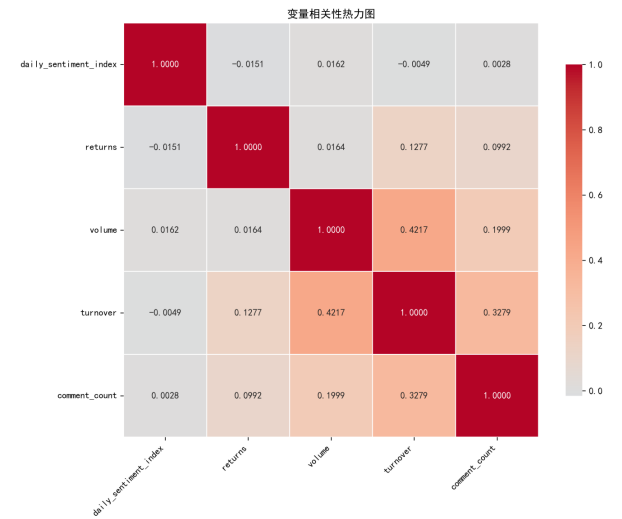


图1 变量相关性热力图

（二）变量时间序列检验与变量共线性分析

对主要变量进行单位根检验，结果显示所有变量均通过平稳性检验（ p 值 <0.05 ），满足时间序列分析的基本假设。通过方差

膨胀因子（VIF）检验，各变量的 VIF 值均小于5，说明变量间不存在严重的多重共线性问题。

（三）OLS回归分析与机器学习分析

表3 组合实验分析结果（不同领先期）

模型类型	OLS	随机森林	梯度提升
领先1天平均 R ²	-6.9987	-0.2776	-0.5190
领先1天平均 MSE	260.9583	12.1380	18.1436
领先3天平均 R ²	-8.6142	-0.2648	-0.5032
领先3天平均 MSE	317.0534	17.5602	22.3869
领先7天平均 R ²	-10.1277	-0.2098	-0.4792
领先7天平均 MSE	377.8069	24.7032	30.3879
领先15天平均 R ²	-10.2387	-0.1998	-0.3998
领先15天平均 MSE	374.6223	24.0012	27.0019
领先30天平均 R ²	-10.1219	-0.1702	-0.3974
领先30天平均 MSE	369.2288	22.3234	26.1918

注：表格展示了不同模型在不同领先期的平均表现，R² 值表示平均解释力，MSE值表示平均预测误差。"- "表示数据不可用。

从模型效果分析，R 方数据结果显示 OLS 回归线性关系不明显，而机器学习取得了较好的 MSE，说明情绪指标可能通过非线性关系作用于股票财务价值关联指标。

从时效性角度分别考察情绪指数领先1天、3天、7天、15天、20天的影响，结果显示：

• 情绪指数领先1天和3天时，MSE 更小，能一定程度说明模型预测效果相对较好，投资者情绪对股价的影响构成较强的关联关系。

• 领先7天以上时，模型效果可以见到显著变差，第15天和第30天 MSE 开始变小模型预测效果逐渐变优。

（四）个股实验分析

本研究继续对10家上市公司样本进行个股对比实验分析，通过对比不同公司的分析结果，可以发现舆情与股价关系的行业差异和个体特征。

结果显示个股在舆情敏感性上存在显著差异，这种异质性与公司所属行业、市值规模、流动性及市场关注度等因素密切相关。高关注度、高流动性的股票多表现出更强的情绪传导效应。

表4 个股实验分析结果汇总（情绪指数领先3天）

股票名称	OLS R ²	OLS MSE	随机森林 R ²	随机森林 MSE	梯度提升 R ²	梯度提升 MSE
立讯精密	0.0399	7.8888	-0.1461	9.4176	-0.3201	10.8477
牧原股份	0.0001	6.2333	-0.6869	10.5156	-1.0538	12.8031
东方财富	-1.0862	79.2777	-0.1968	45.4817	-0.3961	53.0560
阳股华泰	-83.9777	2947.1157	-0.3525	46.9058	-0.8730	64.9591
宁德时代	0.0926	15.7594	-0.0194	17.7043	-0.3103	22.7570
ST 新潮	-0.0973	4.7561	-0.2802	5.5489	-0.4715	6.3782
渤海化学	-1.1401	47.2650	-0.4222	31.4101	-0.6607	36.6764
首创证券	0.0495	13.1840	0.1258	12.1250	-0.0152	14.0817
寒武纪	-0.0465	34.4682	-0.4790	48.7162	-0.5851	52.2105
上纬新材	0.0234	14.5861	-0.1903	17.7772	-0.3458	20.0993

注：R² 值表示模型解释力，MSE值表示预测误差。样本期为2024全年日度数据。

五、深度讨论

通过对上诉实验结果进行分析，能一定程度发掘投资者舆情

与股价变动之间的复杂关系，本文主要发现如下：

1. 投资者情绪作用途径：机器学习模型（随机森林、梯度提升）在预测精度上显著优于传统 OLS 回归，投资者情绪与股市价格波动不是简单线性关系，而存在复杂非线性作用，需要高维方法捕捉多维路径来进行分析和解释。

2. 影响的时效性：所有模型均在情绪指数领先 1 - 3 天时预测效果最好，表明舆情对股价的影响具有短期集中特征。随着领先期延长至 7 天，模型性能显著下降，反映出市场信息消化迅速，长期预测易受其他因素干扰，15-30 天模型效果有一定程度变好，这与舆情作用的时效性、投资者恢复理性有高度关联。

3. 个股异质性：不同股票的舆情敏感性差异显著，这与公司

行业属性、市值规模及市场关注度密切相关。例如，科技类公司及高流动性、高关注度个股的股价对舆情变化反应更为明显。

六、启示与展望

本研究认为未来需结合非线性模型重新审视情绪在资产定价中的作用；可将行为金融学框架与机器学习方法进一步融合，研究 15 天后投资者情绪对股市影响的非线性作用路径，揭示可能存在的理性变化关系。通过整合新闻、社交媒体、研报等多源信息，构建更全面的舆情指标；开展细粒度情感分析，并区分机构与个人投资者进行差异研究。

参考文献

- [1] 陈其安，朱敏，赖琴云. 基于投资者情绪的投资组合模型研究 [J]. 中国管理科学, 2012, 20(3): 47-56.
- [2] 王春. 投资者情绪对股票市场收益和波动的影响 - 基于开放式股票型基金资金净流入的实证研究 [J]. 中国管理科学, 2014, 22(9): 49-56.
- [3] 于孝建，刘国鹏，刘建林，等. 基于 LSTM 网络和文本情感分析的股票指数预测 [J]. 中国管理科学, 2024, 32(8): 25-35.
- [4] Liu, Q., Lee, W.S., Huang, M., & Wu, Q. (2023). Synergy between stock prices and investor sentiment in social media. *Borsa Istanbul Review*, 23(1), 76-92.
- [5] Malkiel, B.G. (1989). Efficient market hypothesis. In *Finance* (pp. 127-134). Palgrave Macmillan, London.
- [6] McGurk, Z., Nowak, A., & Hall, J.C. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance*, 44(3), 458-485.
- [7] Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, 144(1), 273-297.
- [8] Simon, H.A. (1990). Bounded rationality. In *Utility and probability* (pp. 15-18). Palgrave Macmillan, London.
- [9] Sun, Y., Zeng, X., Zhou, S., Zhao, H., Thomas, P., & Sun, L. (2021). What investors say is what the market says: Measuring China's real investor sentiment. *Personal and Ubiquitous Computing*, 25(3), 587-599.
- [10] Xie, D., Cui, Y., & Liu, Y. (2023). How does investor sentiment impact stock volatility? New evidence from Shanghai A-shares market. *China Finance Review International*, 13(1), 102-120.
- [11] Xie, X., Yang, X., & Cui, R. (2025, March). Research on Investor Education in China Based on Large Language Models. In *2025 14th International Conference on Educational and Information Technology (ICEIT)* (pp. 484-489). IEEE.
- [12] Yu, Y., Zhou, R., Jiang, R., & Liu, F. (2025). Exploring the role of social bots in cryptocurrency manipulation: Machine learning insights from the LUNA crash. *Electronic Markets*, 35(1), 110.
- [13] Zheng, B. (2022). Investor Sentiment, Five - factor Model and Portfolio Returns: Empirical Research on China's A - Share Markets. *World Scientific Research Journal*, 8(3), 485-500.