

社交媒体投资者情绪对股票市场的影响研究

文 / 熊莉娜

社交媒体的蓬勃发展为投资者非理性表达提供了新的平台和渠道。投资者在网络平台留下的非结构化数据，可通过情感分析法获取有价值的信息。本文根据微博股评数据，构建网络投资者情绪与情绪分歧度指标，并借助面板向量自回归（PVAR）模型探究两者对股票价格和交易量的影响机制。研究发现：情绪指数对情绪分歧度存在负向影响，对股票价格和交易量存在正向影响；情绪分歧度对股市价格存在负向影响；情绪指数、情绪分歧度以及交易量之间存在双向影响。

关键词：社交媒体；投资者情绪；PVAR 模型；文本情感分析

一、文献综述

20 世纪 80 年代的传统金融学中，投资者完全理性的假设受到了质疑。研究过程中出现了许多传统金融学无法解释的金融异象，例如节假日效应、小公司效应、羊群行为和处置效应^{[1][2]}。我国股市投资者主要为散户，他们投机性强、缺乏金融理论知识并极易受情绪的影响，互联网的高速发展加剧了非理性情绪的崛起与传播。众多研究表明，股票对于舆论信息的反应极其敏感，有学者应用传统实证方法证了实股价的波动和投资者的情绪有密切的联系，亦有学者借助投资者情绪构建 LSTM 模型预测股价的变动，并取得了相比无情绪特征的模型更良好的效果^[3]。

现实中投资者情绪难以量化，目前情绪的度量 and 构建方法分为两种。一是间接度量法，采用市场调查指标、与股票市场有关的单一经济变量或组合变量，二是直接度量法，从网络中获取有价值的信息构建投资者情绪指数，这一方法主要通过词典法和机器学习法实现^[4]。词典法指运用情感词典衡量文本情绪及语调，机器学习法指使用向量机等机器对文本情感进行分类训练从而自动识别情感。

近年来采用文本情感分析法提取股票论坛、社交媒体、新闻报道等文本信息来构建投资者情绪已成为研究热点，但鲜有文献研究社交媒体投资者情绪、情绪分歧度与股票市场这三者间的相互关系。鉴于此，本文通过爬取社交媒体微博股评数据，基于中文金融情感词典^[5]对每条文本进行量化分析，构建社交媒体投资者情绪指数与情感分歧度，运用面板向量自回归（Panel Vector Autoregression, PVAR）模型对投资者情绪、情绪分歧度与股市收盘价、成交量之间的关系进行分析。

二、数据指标和模型

（一）数据来源与采集

本文按照行业分类抽取 2021—2022 年沪深 300 指数中的 30 支股票作为研究样本。本文的数据构成分为两部分。第一部分为股票的基本信息等结构化数据，该部分数据在一个提供股票、基金、期货等数据的 Python 财经数据接口 Tushare 中获取。第二部分为股评文本等非结构化数据，该部分数据通过编写爬虫程序获取各样本股的股评文本，最终得到了 1659494 条有效文本。

（二）模型的设定

本研究采用面板向量自回归模型深入剖析社交媒体情绪与股票市场的动态影响关系。该模型是一种动态的非结构化模型，在不确定变量之间的因果关系时可将所有变量

熊莉娜（1999—），女，汉族，四川南充人，硕士研究生在读。
单位：云南财经大学商学院（云南昆明，650000）。研究方向：行为金融、社交媒体、投资决策。

都视为内生变量,分析各变量及其滞后变量对模型中其他变量的影响,数学模型如下:

$$Y_t = C + \sum_{s=1}^m \beta_s Y_{t-s} + \varepsilon_t \quad (1)$$

Y_t 是时间序列向量,分别选取情绪指数、情绪分歧度、交易量、收盘价四个指标; C 是常数向量; β_s 是带估计系数矩阵, m 表示滞后阶数, ε_t 为模型的回归残差。

(三) 变量选取

1. 情绪指数 (Sentiment Index)

情绪指数代表每支股票的日整体情感倾向。本文采用姚加权学者专门针对中文建立的金融词典,编写程序自动抓取股评文本,清洗数据,删除广告、超链接等噪音文本,自动分词、判别词性情感得出单条博客的情感分数后计算日整体情绪值,指标构建公式如下:

$$SI = \frac{\sum_{i=1}^N score_i}{N} \quad (2)$$

式(2)中, $score_i$ 表示文本中积极或消极词汇个数与总词数的占比; N 代表日发帖数量。 SI 值越高,投资者情绪高涨,“积极”情绪占比较多。反之投资者情绪低迷,“消极”情绪占比较多。

2. 情绪分歧度 (Emotional Divergence)

情绪的分歧度是当日投资者情绪偏离平均值的程度,能够反映出投资者日内情绪变化,由投资者情绪指数方差得到,计算公式如下:

$$ED = \frac{\sum_{i=1}^N (score_i - SI)^2}{N} \quad (3)$$

式(3)中 ED 值越高,情绪异质性越高,市场中看涨、中立以及看跌等情绪的分散程度越大。反之,股民的市场情绪一致性高,微博中出现大量看涨文本或看跌文本。各变量描述性统计结果如表 1。

表 1 描述性统计结果

变量	样本数	平均值	标准差	最小值	最大值
SI	13949	0.630	0.200	0	1
ED	13949	0.440	0.200	0	1
$Close$	13949	0.570	0.170	0	1
$High$	13949	0.500	0.0500	0	1
Vol	13949	0.240	0.140	0	1

三、实证分析

(一) 平稳性检验

时间序列平稳是回归分析赖以实施的基本假设。为防止出现伪回归,本文在前人的基础上采用了三种方法对变量进行平稳性检验,详见表 2。结果表明,模型中 SI 、 ED 、 $Close$

和 Vol 变量在 1% 的水平上显著,均是平稳性序列。为确保研究的谨慎性,本文又采用协整检验方法,结果仍显著,故各序列均为平稳序列。

表 2 各序列的平稳性检验

变量	IPS 检验	LLC 检验	HT 检验	检验结论
SI	-74.841***	-83.840***	-360***	平稳
ED	-78.178***	-88.180***	-390***	平稳
$Close$	-91.682***	-120***	-520***	平稳
Vol	-55.868***	-56.864***	-210***	平稳

注:*** 表示在 1% 的水平上显著。

(二) GMM 估计结果分析

本文通过格兰杰因果检验发现 SI 、 ED 和 Vol 之间均存在双向格兰杰因果关系, $Close$ 是 Vol 的单向格兰杰因果关系。进一步对变量进行 GMM 估计和脉冲响应分析探索各变量间具体的影响机制,参考 AIC、BIC、HQIC 信息原则,将该模型的最优滞后阶数设为 1 阶,GMM 估计结果如表 3 所示。

表 3 GMM 估计结果

变量	SI	ED	$Close$	Vol
$SI(-1)$	-0.033	0.268***	0.006	0.158***
$ED(-1)$	-0.459***	0.614***	-0.012	0.166***
$Close(-1)$	-0.540***	0.413***	-0.058***	0.159***
$Vol(-1)$	-0.113***	0.072***	-0.003	0.640***

注:*** 表示在 1% 的水平上显著。

滞后一期的 SI 与 ED 正相关,估计系数为 0.268,当期投资者的社交媒体情绪分歧度将受到前期市场情绪的影响。滞后一期 SI 对 Vol 产生 15.8% 的正向影响,即投资者的决策行为会受到前一期的情绪影响。滞后一期的 $Close$ 对 SI 具有强烈的抑制作用,对 ED 、 Vol 具有促进作用,投资者情绪受到前期股票价格的影响,前一期收盘价攀高引发消极的投资者情绪,相反社交媒体中的情绪异质性增加,投资者的决策行为分歧度高将作用于股票市场导致交易量上涨。滞后一期的 Vol 与 SI 呈负相关,与 ED 呈正相关。

(三) 脉冲响应分析

本文仅对具有格兰杰因果关系的变量进行脉冲响应分析,进行 1000 次蒙特卡洛模拟,时间跨度为 5 期,实验结果如图 1 所示。

由图 1 可知:(1) SI 、 ED 在受到自身冲击后当期达到正向最大值,说明两者均存在正反馈效应,但随着滞后期数的增加该效应逐渐减弱。(2) SI 受到 ED 和 Vol 的冲击后在第 1 期达到负向峰值,直至第 5 期仍是负影响,表明 ED 、 Vol 对

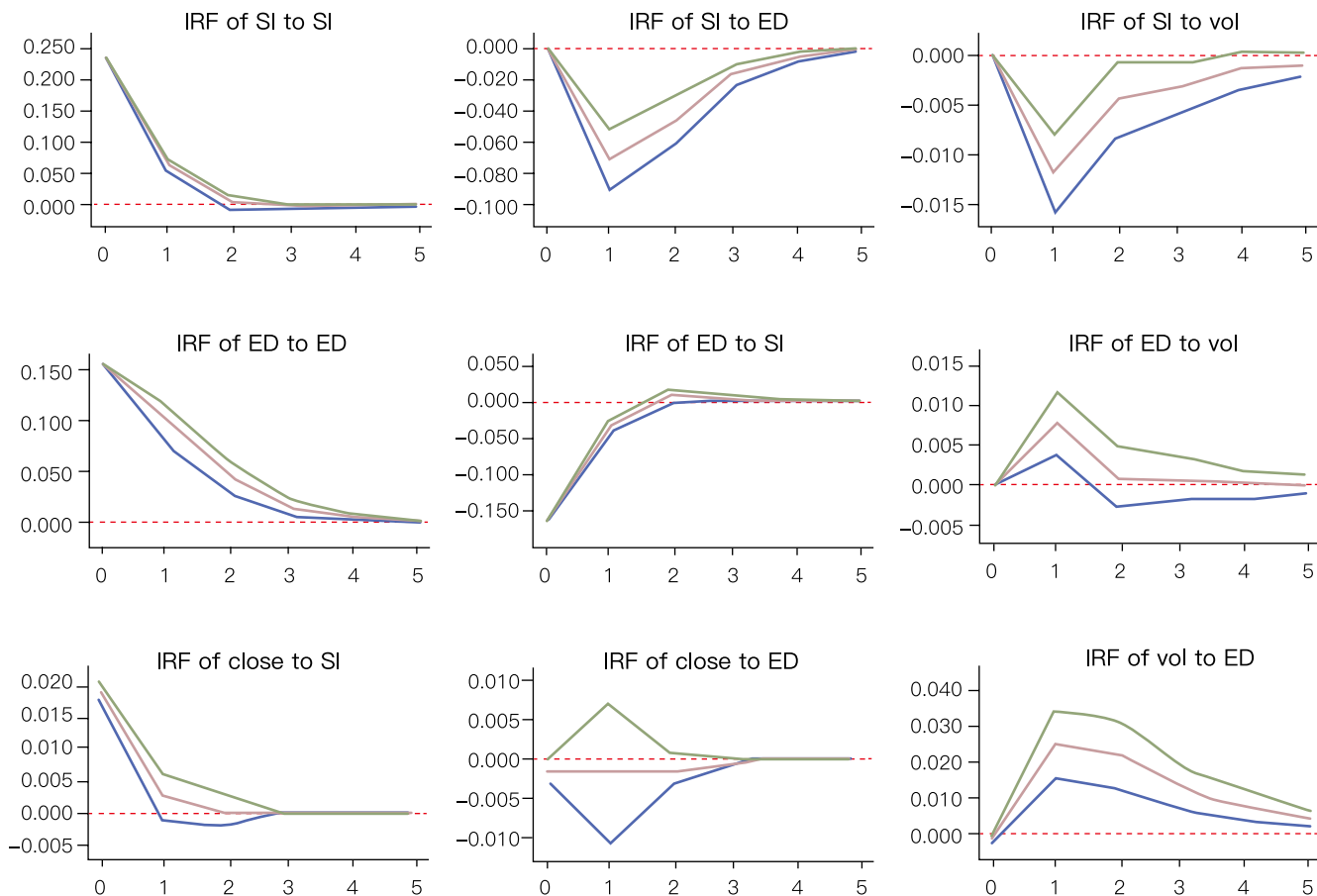


图1 脉冲响应

注：横轴为滞后期数，纵轴表示冲击大小，双侧实线构成的区域为在5%置信水平上的两个标准误差置信区间，中间实线表示脉冲响应函数曲线。

SI 具有抑制作用，情绪分歧度越高，越容易在后期产生消极情绪，可见中国投资者在面对不确定事件时，为了规避风险往往采取消极应对方式。(3) *ED* 受 *SI* 的一个标准差冲击后，在前两期对 *ED* 产生明显的负向影响，第3期开始产生正向影响，说明极端情绪对情绪的异质性具有短期的抑制作用，情绪值越大，决策意见越统一，投资者的情绪分歧度越小。

(4) *Close* 在受到 *SI* 冲击后产生正向影响，第1期达到峰值，随后影响减弱，收敛至0，表明投资者非理性情绪在社交媒体交互功能的影响下，投资者大面积跟风，盲目自信地决策交易引发股价的暴涨暴跌。(5) *ED* 受到冲击后对 *Vol* 产生正向影响，表明 *ED* 受到冲击后会导致股票交易量迅速上升。

四、结语

本研究基于2021—2022年沪深300指数中30只股票的数据建立PVAR模型，研究了社交媒体投资者情绪、情绪分歧度与股票市场之间的关系。研究发现：第一，情绪指数、情绪分歧度、股市交易量之间存在双向影响。第二，投资者情绪对情绪分歧度存在负向影响，对股市价格和交易量存在正向影响。第三，情绪分歧度与股市价格呈负相关影响。本

研究对投资者而言，有利于增强自身识别有效信息的能力，减少情绪化交易行为，形成理性、正确的投资观。对监管部门而言，能够合理监管社交媒体中的投资情绪，治理因非理性情绪导致的市场乱象。

参考文献：

- [1] 刘新争，高闯．基于复杂网络的私有信息共享对机构投资者羊群行为的影响[J]．管理学报，2022(19)：506-515.
- [2] Su Chi-Wei, Cai Xu-Yu, Tao Ran. Can stock investor sentiment be contagious in China? [J]. Sustainability, 2020(12):1571.
- [3] 欧阳资生，李虹宣．网络舆情对金融市场的影响研究：一个文献综述[J]．统计与信息论坛，2019(34)：122-128.
- [4] 张同明，张宁．股票市场投资者情绪指数研究综述[J]．计算机科学，2021(48):143-150.
- [5] 姚加权，冯绪，王赞钧，等．语调、情绪及市场影响：基于金融情绪词典[J]．管理科学学报，2021(24)：26-46.