

# 网络情感因子对股票收益率的影响研究

王 晴

广东财经大学金融学院, 广东 广州

收稿日期: 2023年5月30日; 录用日期: 2023年6月13日; 发布日期: 2023年7月26日

## 摘 要

随着网络技术的发展,越来越多的个人投资者通过股吧等网络平台,就股市或某只股票的走势发表观点,本文利用网络平台上的投资者的主观情绪,构建了一个网络情感因子。首先,利用爬虫程序对东方财富股吧从2021年7月3日到2022年7月3日的沪深300成分股股评文本进行采集,对股评进行分类后计算出网络情感因子值,并用线性回归模型实证网络情感因子显著正向影响本期股票收益,显著负向影响下期股票收益。论证了文本投资者情绪对股票价格的重要影响,有助于投资者了解股票价格的投资者行为风险,从而降低投资风险,也有助于监管部门管理投资者情绪风险,稳定股票市场。

## 关键词

投资者情绪, 情绪分析, 股票收益率

# Research on the Influence of Network Emotional Factors on Stock Returns

Qing Wang

School of Finance, Guangdong University of Finance and Economics, Guangzhou Guangdong

Received: May 30<sup>th</sup>, 2023; accepted: Jun. 13<sup>th</sup>, 2023; published: Jul. 26<sup>th</sup>, 2023

## Abstract

With the development of network technology, more and more individual investors are expressing their opinions on the trend of the stock market or a certain stock through online platforms such as stock bars. This article uses the subjective emotions of investors on online platforms to construct a network sentiment factor. Firstly, a crawler program was used to collect the stock review texts of the Shanghai and Shenzhen 300 component stocks of Dongfang Wealth from July 3, 2021 to July 3, 2022. After classifying the stock reviews, the online sentiment factor values were calculated, and a linear regression model was used to demonstrate that the online sentiment factor significantly

**positively affects the current stock returns and significantly negatively affects the next stock returns. The important impact of investor sentiment on stock prices is demonstrated, which helps investors understand the behavioral risks of stock prices and reduce investment risks. It also helps regulatory authorities manage investor sentiment risks and stabilize the stock market.**

## Keywords

**Investor Sentiment, Emotional Analysis, Stock Returns**

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着政府的大力扶持，我国互联网逐渐普及，网民数量快速增长。根据中国互联网信息中心官方数据显示，截至 2022 年 12 月，我国的网民规模达到了 10.67 亿，互联网普及率高达 75.6%。从 2010 年-2022 年我国网民规模和互联网普及率迅速发展，2022 年的网民规模和互联网普及率是 2010 年的两倍以上。这都为研究投资者主观情绪提供了条件。越来越多的个人投资者会在股评类平台上浏览帖子或自己发布帖子，帖子内容大多是投资者对股票的主观情绪，这些主观性情绪无形中又影响了投资者的投资决策。股吧、股票论坛等已经成为当前投资者获得股票资讯、发表股票看法必不可少的渠道，同时也是学者获取投资者网络情绪的有效平台。由于互联网上存在大量的互联网用户，其情绪在股市中的作用日益重要。因此，需要从投资者网络情感角度分析其对股票收益率的影响。

## 2. 文献综述

我国与投资者情绪有关的研究虽然起步晚，但是发展迅速，目前已有的关于投资者情绪的理论研究大概可以分为三种：一种是直接投资者情绪指数，一种是间接直接投资者情绪指数，还有一种是投资者的文本情绪指数。在建立投资者的直接情绪指数时，程坤和刘仁和(2005) [1]认为以好淡指数为代表情绪指数可以反映出股票市场的“牛熊”状态，并对投资者情绪的“短”、“中”指数与股票市场收益的关系进行了研究。王美今和孙建军(2004) [2]采用央视看盘指数构建了一个投资者情绪指标，发现沪深股市投资者情绪的改变不但会对股市的收益产生明显的影响，还会对股市的波动性产生明显的反向修正作用。在构建间接的投资者情绪指标时，研究者通常采用单一的或多个间接指标综合的方法。薛斐(2005) [3]对封闭式基金的折扣率与消费者信心指标进行了比较，实证发现消费者信心指标可以较好地测度投资者的情绪，是一个良好的代理指标。易志高与茅宁(2009) [4]对 BW 指数的构建方法进行了完善，通过对指标进行主成分分析，从而得到了可以更好地测度投资者的情绪的复合指标。韩立岩和伍燕然(2007) [5]以 IPO 首个交易日的股票市场指标构建投资者情绪指标，发现由于时间不同，投资者情绪对股票市场回报的作用呈现出“正向回报”和“逆向回报”两种特征。饶兰兰和涂裕荣(2016) [6]采用六个投资者代理情绪变量构造投资者情绪指标，通过实证分析发现，牛熊市不同，投资者情绪对股市收益率的作用呈现出不对称性。在过去十年中，国内学者在投资者文本情绪指标方面也进行了不少的研究，徐维军等(2022) [7]分别采用辞典方法和朴素贝叶斯法对东方财富网的发帖文字进行了情感分析，结果表明，使用朴素贝叶斯法构建的金融指标更有效。尹海员与杨庆松(2022) [8]通过建立双向长时记忆神经网络，实现了对东方财富股吧中发帖人的情绪辨别与分类，并建立了有效的日度投资者情绪指数。高扬等(2022) [9]在文本挖掘

技术的基础上,采用 Bi-LSTM 深度学习方法,对文本数据中的情绪进行识别,构建了一个投资者情绪指标,并研究该指标是如何通过影响科创板成交量从而对股市收益率产生作用。

### 3. 数据来源及说明

#### 3.1. 样本数据的选择

周度股票收益率数据来源于 CSMAR 数据库,构造周度网络情感因子值所需的股评数据来源于东方财富沪深 300 成分股股吧,时间范围为 2021 年 7 月 3 日~2022 年 7 月 3 日。

#### 3.2. 计算网络情感因子值

本文编写针对东方财富网股吧的爬虫程序,之后运行程序对沪深 300 成分股个股吧的股评进行爬取。获取股评文本之后,对数据进行分词等预处理之后将文本分为字数不少于 20 的长文本和字数少于 20 的短文本,长文本通过长短期记忆网络(LSTM)模型进行情感分类,短文本通过朴素贝叶斯(NBM)模型进行情感分类。最后统计每周股票的“negative”评论数和“positive”评论数,并以此来计算网络情感因子。本文借鉴 Antweiler 和 Frank (2004) [10]所提出的股市情感指标构建方法,周度的网络情感因子(NSF, Network Sentiment Factor)可以构建为:

$$NSF_{i,t} = \ln \frac{(1 + \text{每周“positive”评论数})}{(1 + \text{每周“negative”评论数})} \quad (1)$$

如公式(1)所示,当某只股票的周“positive”股评数大于“negative”股评数时,网络情感因子值为正,表示当周投资者整体上对股票 X 的表现持有积极看法,反之则为负,表示对该股票的表现持有消极看法。

### 4. 实证过程及结果

#### 4.1. 面板单位根检验

使用 IPS、ADF 两种检验方法对网络情感因子和股票收益率进行面板单位根检验,检验数据是否平稳。网络情感因子用 NSF (Network Sentiment Factor)表示,股票收益率使用 SR (Stock Return)表示,检验结果如表 1 所示。

Table 1. Panel unit root test results

表 1. 面板单位根检验结果

变量	IPS	ADF	平稳性
网络情感因子	-56.5613***	-42.9157***	平稳
股票收益率	-76.9157***	-86.4600***	平稳

注:\*\*\*表示在 1%的显著性水平下显著。

网络情感因子的 IPS 检验和 ADF 检验的显著性水平均为 1%,股票收益率的 IPS 检验和 ADF 检验的显著性水平也均为 1%,说明网络情感因子和股票收益率数据都是平稳的,可以进行格兰杰因果检验。

#### 4.2. 格兰杰因果检验

面板数据通过单位根检验之后,将网络情感因子和股票收益率之间进行格兰杰因果关系检验,使用 Stata 软件的 xtgcause 命令对面板数据进行格兰杰因果检验,格兰杰检验提供了两个检验统计量 Z-bar 和

Z-bar tilde 和对应的 P 值。对于一个 N 足够大而 T 相对较小的面板数据而言, 应该考虑使用 Z-bar tilde 统计量[11]。本文的面板数据 N 较大, 而 T 较小, 是一个短面板数据, 因此使用 Z-bar tilde 统计量。表 2 为滞后两期的格兰杰因果检验结果。

**Table 2.** Granger causality test results

**表 2.** 格兰杰因果检验结果

H0	Z-bar tilde	P
NSF does not Granger-cause SR	-2.2768	0.0228
SR does not Granger-cause NSF	4.3933	0.0000

从表 2 中可以看出, 对于原假设: NSF 不是 SR 的格兰杰原因, Z-bar tilde 统计量的为-2.2768, 对应的 P 值为 0.0228, 可以在 5% 的显著性水平下拒绝原假设, 网络情感因子是股票收益率变化的格兰杰原因, 网络情感因子对股票收益率的变化具有预测作用。对于原假设: SR 不是 NSF 的格兰杰原因, Z-bar tilde 统计量的值为 4.3933, 统计量对应的 p 值为 0.0000, 说明在显著性水平 1% 的水平下拒绝原假设, 股票收益率是网络情感因子变化的格兰杰原因, 股票收益率对网络情感因子的变化具有预测作用。因此, 网络情感因子和股票收益率之间互为因果关系。

### 4.3. 豪斯曼检验

通过格兰杰因果关系检验可知, 网络情感因子和股票收益率互为因果关系, 接下来用回归分析法进一步研究滞后一期网络情感因子与股票收益率的线性关系。

对面板数据进行处理时, 一般需要通过豪斯曼检验确定对面板数据采用固定效应模型还是随机效应模型。豪斯曼检验的原假设  $H_0$ : 随机效应模型有效, 豪斯曼检验结果如表 3 所示。

**Table 3.** Hausman test results

**表 3.** 豪斯曼检验结果

	FE	RE	difference
滞后一期网络情感因子	-0.5306	-0.0346	-0.4961

Chi2(1) = 23.67\*\*\*, Prob > chi2 = 0.0000.

由于豪斯曼检验中 P 值为 0.0000, 强烈拒绝原假设, 所以应该使用固定效应模型来对面板数据进行处理。

### 4.4. 网络情感因子对股票收益率的影响

使用双固定效应模型对网络情感因子和股票收益率进行回归分析, 结果如表 4 所示。

**Table 4.** Regression results of network sentiment factors and stock returns

**表 4.** 网络情感因子与股票收益率回归结果

	变量	Coef	std.Error	t 值	p 值
滞后一期网络情感因子	$NSF_{i,t-1}$	-0.5306	0.1446	-3.6685	0.0002
	C	-0.3506	0.0995	-3.5249	0.0004
当期网络情感因子	$NSF_{i,t}$	7.0868	0.1482	47.8336	0.0000
	C	4.2790	0.1019	41.9938	0.0000

从回归结果可以看出，滞后一期的网络情感因子和当期网络情感因子都可以预测股票收益率，两者都在 1% 的水平下显著。

通过一元线性回归模型得出，滞后一期的网络情感因子和股票收益率的方程系数为负值，滞后一期的网络情感因子负向影响股票收益率，当滞后一期的网络情感因子增加一个单位，股票收益率就会减少 0.5306%，方程的 P 值为 0.0002，说明方程整体是显著的。滞后一期的网络情感因子和股票收益率显著相关。

通过一元线性回归模型得出，当期网络情感因子和股票收益率的方程系数为正值，也就是说当期的网络情感因子对于个股收益率正向影响，当期的网络情感因子增加一个单位，股票收益率会增加 7.0868%，方程的 P 值为 0.0000，说明方程整体是显著的。即当期网络情感因子对于当期的股票收益率是有显著相关性的。

网络情感因子和当期股票收益率正相关，而和下期股票收益率负相关。本文认为，当期网络情感因子情感倾向在短期内导致股票收益率显著同向变化，可以正向预测当期股票收益率。而等到下期，市场对于网络情感因子的过度反应将反转，市场修正价格回归。网络情感因子就和下一期的股票收益率负相关，可以反向预测股票的收益情况。

## 5. 结语与建议

随着互联网的普及，网络已经成为了许多投资者交流信息的重要渠道。通过网络平台，投资者不仅能够分享自己的投资观点，还能从他人的投资经验中获益。因此，必须重视网络平台上的投资者情绪，并将其作为一个重要的指标与股票收益率进行检验。本文利用东方财富网沪深 300 成分股吧的股评数据构建网络情感因子指标，并利用回归模型实证网络情感因子显著影响股票收益率。

证券市场近年来飞速发展，可于此相对应的是证券市场制度的不完善，投资者投资决策的不理性。我国股票市场还没有完全符合有效市场的假设条件，因此本文构建了能体现投资者行为的行为金融学因子，即网络情感因子，并实证网络情感因子显著影响股票收益率具有重要意义。首先，对于投资者来说，网络情感因子能够提高投资者对于证券市场的认知和分析能力，同时也可以将网络情感因子作为一个有效的市场指标帮助投资者构建投资组合，做出更加理智的投资决定。其次，对于监管者来说，网络情感因子可以让证券监管机构更加了解投资者情绪，明确投资者投资行为的理性程度，同时警惕投资者主观情绪对于股票价格涨跌的影响。有利于监管机构更好的监管证券市场，减少证券市场情绪性风险的产生。

## 参考文献

- [1] 程昆, 刘仁和. 投资者情绪与股市的互动研究[J]. 上海经济研究, 2005(11): 88-95.
- [2] 王美今, 孙建军. 中国股市收益、收益波动与投资者情绪[J]. 经济研究, 2004(10): 75-83.
- [3] 薛斐. 我国投资者情绪指数选择的实证检验[J]. 世界经济情况, 2005(14): 14-17.
- [4] 易志高, 茅宁. 中国股市投资者情绪测量研究: CICSI 的构建[J]. 金融研究, 2009, 353(11): 174-184.
- [5] 韩立岩, 伍燕然. 投资者情绪与 IPOs 之谜——抑价或者溢价[J]. 管理世界, 2007(3): 51-61.
- [6] 饶兰兰, 涂裕荣. 牛熊市下投资者情绪对股票收益影响的非对称性分析[J]. 商业经济, 2016(8): 141-143.
- [7] 徐维军, 黄静龙, 付志能, 张卫国. 基于金融文本情绪挖掘的 Black-Litterman 投资组合模型研究——以东方财富股吧发帖文本和我国 A 股市场为例[J]. 运筹学学报, 2022, 26(4): 1-14.
- [8] 尹海员, 杨庆松. 基于 Bi-LSTM 模型挖掘的股吧投资者情绪对股价泡沫的影响[J]. 管理学报, 2022, 19(12): 1874-1885.
- [9] 高扬, 申怡然, 徐嘉熙. 投资者情绪对科创板市场收益率的影响——基于文本数据挖掘视角[J]. 运筹与管理, 2022, 31(12): 1-10.

2022, 31(2): 184-190.

- [10] Antweiler, W. and Frank, M.Z. (2004) Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, **59**, 1259-1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
- [11] Lopez, L. and Weber, S. (2017) Testing for Granger Causality in Panel Data. *The Stata Journal*, **17**, 972-984. <https://doi.org/10.1177/1536867X1801700412>