

投资者情绪与原油价格的相关性分析

向虹蝶

贵州大学经济学院, 贵州 贵阳

收稿日期: 2024年7月19日; 录用日期: 2024年8月15日; 发布日期: 2024年11月4日

摘要

本文研究的时间段设为2018年1月1日至2023年12月21日, 运用Python爬取东方财富网原油吧标题, 再用SnowNLP对获得的文本信息进行情感评分, 从Wind上获取大庆原油价格和上证综合指数。相关性检验证明投资者情绪与上证综指都与原油价格呈正相关, 并且投资者情绪是原油价格的格兰杰原因, 原油价格可以被投资者情绪简单预测。本文为原油价格影响因素研究提供了一个新的视角。

关键词

原油价格, 投资者情绪, 上证综合指数

Analysis of the Correlation between Investor Sentiment and Crude Oil Prices

Hongdie Xiang

School of Economics, Guizhou University, Guiyang Guizhou

Received: Jul. 19th, 2024; accepted: Aug. 15th, 2024; published: Nov. 4th, 2024

Abstract

The research time period of this paper is set from January 1, 2018 to December 21, 2023, using Python to crawl the title of the crude oil bar of Oriental Fortune Network, and then using SnowNLP to score the sentiment of the obtained text information, and obtaining the Daqing crude oil price and the Shanghai Composite Index from Wind. The correlation test proves that investor sentiment and the Shanghai Composite Index are positively correlated with crude oil prices, and investor sentiment is the Granger reason for crude oil prices, and crude oil prices can be easily predicted by investor sentiment. This paper provides a new perspective for the study of the influencing factors of crude oil prices.

Keywords

Crude Oil Prices, Investor Sentiment, Shanghai Composite Index

Copyright © 2024 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

原油被称为“工业的血液”，不仅是生产、国防和交通运输的重要原料，还是重要的战略资源与大宗商品。近 20 年来，大宗商品价格存在异常大幅波动，有学者研究证明该波动与金融因素有关，原油价格也不例外，也就是说已经不只是供需、物流、政策等在影响原油价格，金融市场也在对其产生影响。有学者研究了股票指数、商品指数、黄金与美元指数等对原油价格的影响，但却忽视了金融市场中的投资者。市场参与者虽然不能直接买卖原油，但是可以买卖原油衍生品从而影响其价格，因此本文将研究投资者情绪与我国大庆原油价格的相关性，并对其进行预测，还会加入上证综合指数进行对比分析，并作为控制变量进行回归分析，为我国原油价格的影响因素研究提供新的视角。

2. 文献综述

投资者情绪指数是对投资者情绪高涨或是低迷的一种量化手段，行为金融学认为投资人理性的假设是错误的，在真实市场中投资者会受到许多方面的影响，从而使市场产品价格产生异常波动，因此越来越多的专家学者通过投资者情绪这一角度去研究对价格等的影响[1]。国外较多使用的是投资者智能指数，友好指数以及有效的反向指标个体投资者协会指数来预测 S&P500 的收益率[2]-[4]。国内显性市场情绪指标包括基于“央视看盘”所提供数据编制的 BSI 指数，由《股市动态分析周刊》杂志编制的好淡指数。陈晓红(2016)基于新浪微博 2014 年数据和百度的搜索指数，得到情绪指数对股票价格有预测作用，搜索指数对交易量的预测有显著作用[5]。张宗新等(2013)运用主成分分析法建立投资者情绪指数，认为投资者情绪对信念存在正面冲击，不同信息偏好将导致不同的情绪波动频率[6]。关筱谨等(2022)运用向量自回归模型(SV-TVP-VAR)得出投资者情绪和媒体关注度均是影响股市波动的重要因素[7]。曹雄飞等(2023)研究得出疫情股票增加时会引起股市的下行[8]。张天顶等(2023)通过推特文本，采用门限广义自回归条件异方差(TGARCH)模型揭示了投资者情绪对国际原油市场的影响[9]。

大宗商品作为经济基础中的工业原材料、能源以及生活必需品的来源，其价格走势与特征对于全球经济具有重要影响，因此一直受到各界广泛关注。近年来，大宗商品价格存在大幅波动现象，有学者认为“大宗商品金融化”是造成该现象的主要因素[10]。而原油不仅涉及工农业生产、国防和交通运输，还是很重要的战略资源与大宗商品。公共卫生安全事件爆发给国民健康、经济发展和社会稳定造成了前所未有的压力，国际原油价格出现持续暴跌，WTI 原油期货甚至首次出现了负油价。Shaikh (2021)和 Borgard (2021)等使用基于新闻报道的传染病股市波动指数，分别验证了传染病相关不确定信息对于原油投资者情绪和原油价格过度反应行为的重要影响[11][12]。周锦岚(2023)利用原油生产、需求、汇率等 195 个变量构建了原油价格不确定综合指数，很好地反映了国际原油市场不同时期的波动特征[13]。

目前对原油价格的研究，较多为对国际原油价格的研究，并且是从供需、政策、时政等角度进行研究预测，可是学者研究发现原油价格越来越大程度上受金融因素影响，而研究股市时，根据行为金融学将投资者情绪纳入影响因子，随着市场异象情况频发，投资者情绪越来越受到广大学者关注。原油不仅

作为经济基础中的工业原材料、能源，还是重要的战略资源与大宗商品，研究其价格走势、影响因素是非常重要的，因此本文将结合行为金融学从投资者情绪角度出发，研究其与大庆原油价格的相关性，以及对大庆原油价格的可预测性，丰富影响原油价格因素的研究。

3. 投资者情绪和数据分析

3.1. 构建投资者情绪评分

3.1.1. 文本挖掘

已有投资者情绪的研究大多是从百度、东方财富网、新浪财经、微博等平台爬取文本信息[14]-[16]。本文为保证获取文本信息的有效性，提高研究的严谨程度。在经过各平台的搜索后，发现东方财富网的用户关注度高，原油吧的信息充足，最早可获得 2011 年的数据，且在技术上也更容易获得数据，因此选择东方财富网的原油吧作为文本挖掘的来源。

3.1.2. 研究数据的选取

本文将研究区间设置在 2018 年 1 月 1 日至 2023 年 12 月 21 日，通过 wind 获得日度大庆原油价格、上证综合指数和深证成分指数。为了快速、方便获得数据，运用 Python 爬取东方财富网原油吧的发帖标题，获得数据 38,470 条。最后通过对获得数据的整理和清洗，得到 1359 组数据。

3.1.3. 情感评分

本文将使用 Python 中的 SnowNLP 对东方财富原油吧的文本数据进行情感分析并打分。SnowNLP 对于评价性的文本段落而言会更加准确，契合于本文想要得到的投资者情绪。SnowNLP 不只有情感分析，还有提取摘要、关键字等功能，方便我们检查情感评分是否有误并进行调整，表 1 是抽取的部分文本信息以及对应的情绪评分。

Table 1. Sentiment analysis results of Oriental Fortune Crude Oil Bar
表 1. 东方财富原油吧情感分析结果表

序号	文本信息	情感评分
1	原油的行情不要预测，只要跟着就赚	9
2	国际油价一路飙升，新潮能源在美国产油企业获利丰厚！	8
3	在龙门超级空单的轰炸下，原油及原油系多头丢盔弃甲，哭爹喊娘	2
4	多原油回撤 83.7 直接多，止损 83，目标 85~85.7	4
5	86.4 已多，给到就是干	2
6	原油，逼空结束牛不动了？9~20	4
7	热烈庆祝原油成功突破 700！！！！	9
8	原油到了阶段性高点，需注意风险！	3
9	原油你说高点也算是高点，我感觉还会有冲高一波，多头比较猛美	6

由于在同一天内会有相同或不同的几个文本数据，将会采用平均数得到日度的情绪评分。SnowNLP 得出的评分在 0~1 之间，为了更好的拟合模型，本文将对情绪分数扩大十倍，4 分以下的为消极情绪占比约为 1.8%，其余为积极情绪占比约为 98.2%，可以看出投资者对原油的价格十分乐观，也符合原油为重要大宗商品，价格处于较高位置的现实情况。

3.2. 分析数据

3.2.1. 描述性统计

从表 2 中能够看到一共有 1359 组数据，原油价格的均值在 66.642 元左右，最低值为 12.56 元，最高值为 124.18 元，极差和标准差都较大数据可能较为离散。投资者情绪评分均值为 6.439，极差仅为 7，可以预测该数据较为集中。

Table 2. Descriptive analysis table
表 2. 描述性分析表

变量	Obs	Mean	Std. Dev.	Min	Max
原油价格	1359	66.642	19.809	12.56	124.18
情绪评分	1359	6.439	0.957	2	9
上证综指	1359	3156.44	279.027	2464.36	3715.37

3.2.2. 趋势对比图

从图 1 中可以看到投资者情绪评分是一个较为平稳的序列，而原油价格有明显的上升趋势，因此在后文需对其进行平稳性处理。除此之外，投资者情绪评分与原油价格的变化趋势似乎一致性不高，可能的原因有很多。可能是因为两者的量级不同，投资者情绪评分所展示出的波动范围较小，绝大部分评分差距存在于小数点后，在图中难以展现。也许是因为存在一定的时滞性，投资者发帖后并不能立马反映在价格上，又或者有一些投资者在价格低位时故意散播向好消息，希望以此让价格回升。到底这两者之间是否存在关系还需在后文用模型检验。

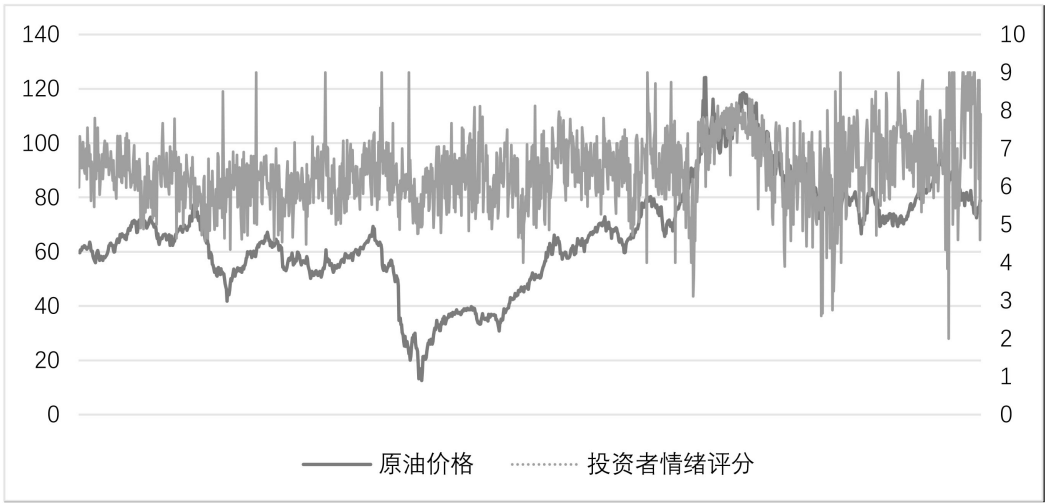


Figure 1. Trend chart of crude oil prices and investor sentiment scores
图 1. 原油价格与投资者情绪评分趋势图

从图 2 可以看到尽管整体来看上证综合指数的波动幅度小于原油价格的波动幅度，但在很多时间段内两者的变动趋势较为一致，也符合学者们研究得到的原油价格与股票指数存在相关性，因此后文将用上证综指作为回归模型的控制变量。

从图 3 可以看出上证综合指数与原油吧投资者情绪评分趋势几乎没有相关性，这符合现实逻辑，也

排除了后文回归时可能存在的多重共线性问题。



Figure 2. Trend chart of the Shanghai Composite Index and crude oil prices

图 2. 上证综合指数与原油价格趋势图

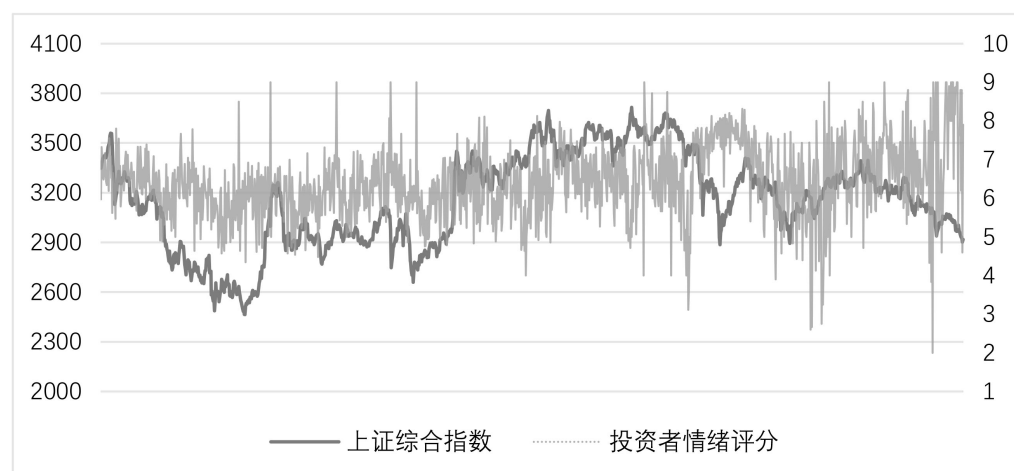


Figure 3. Trend chart of the Shanghai Composite Index and investor sentiment scores

图 3. 上证综合指数与投资者情绪评分趋势图

观察图 1~3 可以看到公共卫生安全事件爆发时原油价格和上证综合指数都有下降趋势，尤其是原油价格大幅下降，达到了这五年内的最低点，而上证综指对比起来有小幅下降，这也表明我国股票市场有抵御突发情况的能力。在公共卫生安全事件后这两者都开始会回升，原油价格也较快地回到之前的水平并且持续上升到更高点。

4. 实证设计及结果

4.1. 相关性检验

基于 2018 年 1 月 2 日到 2023 年 12 月 21 日间的每个周一到周五，剔除节假日和缺失数据，共得到 1369 组数据，对原油价格与投资者情绪评分和上证综合指数分别进行相关性检验。选取了五年时间，包含了疫情以及经济复苏后的情况，在一定程度上能为本文的研究提供一个较好的样本环境。结果如表 3 所示：

Table 3. Correlation test results**表 3.** 相关性检验结果

	原油价格
情绪评分	0.309
上证指数	0.197

根据表 3 结果, 可以得出投资者情绪评分与原油价格收盘价以及上证综合指数均呈正相关, 尽管相关性只有 0.309 和 0.197 不是非常显著, 即投资者的情绪越积极, 上证指数越高, 原油的价格就会更高, 并且投资者情绪与原油价格的相关性要强于上证指数。

4.2. 格兰杰因果检验

在进行格兰杰因果检验之前, 由于时间序列固有的特性, 要先对原始数据进行平稳性检验。根据前文的趋势图可以看出原油价格是典型的非平稳序列, 通过 EViews 单位根检验得出原油价格有单位根, 投资者情绪评分和上证综合指数无单位根, 因此只需要对原油价格进行平稳化处理。经过一阶差分后原油价格序列的原始假设单位根被拒绝, 则序列可以认为是平稳的。

格兰杰因果检验是用于检验一组时间序列是否为另一组时间序列的原油。如果说 A 为 B 的格兰杰原因, 则说明 A 的变化是引起 B 变化的原因之一。若在包含了变量 X、Y 的过去信息条件下, 对变量 Y 的预测效果只要优于仅单独有 Y 过去信息对 Y 进行预测的结果, 也就是说变量 X 有助于解释变量 Y 的将来变化, 则认为变量 X 是变量 Y 的格兰杰原因。需要注意的一点是格兰杰因果检验是检验统计上的时间先后顺序, 并不表示二者之间存在真正的因果关系。

表 4 的第一列表示特定的假设条件。比如, “情绪评分 - 原油价格” 表示原始假设为 “情绪评分不是原油价格的格兰杰原因”。通过表 4 中列出的各个 P 值, 这个原始假设在滞后期为 1, 2, 3 被强烈拒绝, 即原假设不成立, 情绪评分是原油价格的格兰杰原因。而反过来的假设在滞后期为 1, 2, 3 都没有被拒绝, 即假设成立, 原油价格却不是情绪评分的格兰杰因果关系。

Table 4. Results of Granger causality test**表 4.** 格兰杰因果检验结果

	1	2	3
情绪评分 - 原油价格	0.000***	0.000***	0.000***
原油价格 - 情绪评分	0.693	0.37	0.041*

4.3. 基于投资者情绪的预测

前文已经验证了变量之间的相关性, 因此这部分将用回归模型来验证投资者情绪对原油价格的预测。

$$Y_t = c_1 + a_1 X_t + \varepsilon_{t1}$$

$$Y_t = c_2 + a_2 X_t + b_2 Z_t + \varepsilon_{t2}$$

其中 Y 代表原油价格, X 代表情绪评分, Z 代表上证指数。

根据表 5 的结果可以看到, 两个回归方程中投资者情绪评分与原油价格在置信度为 99% 的情况下是呈现正相关关系, 也就是说当原油投资者对于原油市场的态度越积极, 原油价格就会越高, 因此可以用该方法简单预测原油价格。加入上证综指后 R² 有小幅上升, 说明拟合程度更高了一些。

Table 5. Regression results 1
表 5. 回归结果 1

	(1)	(2)
	原油价格	原油价格
情绪评分	6.399*** (11.97)	5.859*** (10.91)
上证综指		0.0104*** (5.67)
_cons	25.44*** (7.31)	-4.031 (-0.65)
N	1359	1359
adj. R ²	0.095	0.115
r ²	0.0955	0.116
F	143.2	89.31

t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.4. 稳健性检验

为使模型结果更有说服力，需要进行稳健性检验。将数据进行 1%~99%的缩尾处理后再次进行回归，回归结果如表 6 所示。

Table 6. Regression results 2
表 6. 回归结果 2

	(1)	(2)
	原油价格	原油价格
情绪评分	6.325*** (11.93)	5.794*** (10.88)
上证综指		0.0103*** (5.63)
_cons	25.91*** (7.51)	-3.093 (-0.50)
N	1359	1359
adj. R ²	0.094	0.114
r ²	0.0949	0.116
F	142.4	88.62

t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

可以看到替换变量后，模型结果产生细微变化但依然显著，可以认为模型是稳健的。

5. 总结

本文通过 Python 对东方财富网原油吧进行文本挖掘，运用 SnowNLP 库将抓取的文本按照关键情绪词进行打分，4 分以下为消极情绪，其余为积极情绪。再以此评分为基础量化投资者情绪，设置变量为情

感评分。分析投资者情绪与原油价格的相关性及预测能力,实证结果表明,原油吧投资者情绪与原油价格存在相关性,尽管相关性不强,情绪评分是原油价格的格兰杰因果。由于原油作为重要的战略资源和大宗商品,其价格的影响因素是十分复杂的,因此投资者情绪仅可简单预测原油价格。而上证综合指数与原油价格也有较小的相关性,加入这一变量后模型的预测精度有细微的提升,作用不太明显。综上所述,网络社交平台中蕴含的信息可以有助于预测原油价格,投资者在进行投资决策时可以综合参考网络上的各类信息。

目前国内对原油价格影响因素的研究大多从供需、物流、政策和实证等角度,在全球金融化的背景下,原油也越来越“金融化”[17]。因此本文只是做一个初步的尝试,研究投资者情绪与原油价格是否存在相关性,是否可以纳入影响原油价格的因素,为相关研究领域提供一个新的视角,希望可以补充完善影响原油价格因素的研究。

参考文献

- [1] 饶育蕾,刘达锋.行为金融学[M].上海:上海财经大学出版社,2003:143-149.
- [2] Brown, G.W. and Cliff, M.T. (2004) Investor Sentiment and the Near-Term Stock Market. *Journal of Empirical Finance*, **11**, 1-27. <https://doi.org/10.1016/j.jempfin.2002.12.001>
- [3] Fisher, K.L. and Statman, M. (2003) Consumer Confidence and Stock Returns. *The Journal of Portfolio Management*, **30**, 115-127. <https://doi.org/10.3905/jpm.2003.319925>
- [4] Brown, G.W. and Cliff, M.T. (2005) Investor Sentiment and Asset Valuation. *The Journal of Business*, **78**, 405-440. <https://doi.org/10.1086/427633>
- [5] 陈晓红,彭宛露,田美玉.基于投资者情绪的股票价格及成交量预测研究[J].系统科学与数学,2016,36(12):2294-2306.
- [6] 张宗新,王海亮.投资者情绪、主观信念调整与市场波动[J].金融研究,2013(4):142-155.
- [7] 关筱谨,张骏,刘彦迪.媒体关注度、投资者情绪与股票市场波动[J].统计与决策,2022,38(24):143-148.
- [8] 曹雄飞,杨双会,范炜彬.基于文本挖掘的疫情情绪对股票市场的影响[J].金融理论与教学,2023(5):44-52.
- [9] 张天顶,杜天文.社交平台投资者情绪对国际原油价格的影响研究[J].国际石油经济,2023,31(11):74-82.
- [10] 李沛然,李奇霖,宋佳馨.经济不确定性、金融化与大宗商品价格协同性[J].金融评论,2019,11(2):61-80,124.
- [11] Shaikh, I. (2021) On the Relation between Pandemic Disease Outbreak News and Crude Oil, Gold, Gold Mining, Silver and Energy Markets. *Resources Policy*, **72**, Article 102025. <https://doi.org/10.1016/j.resourpol.2021.102025>
- [12] Borgards, O., Czudaj, R.L. and Hoang, T.H.V. (2021) Price Overreactions in the Commodity Futures Market: An Intraday Analysis of the COVID-19 Pandemic Impact. *Resources Policy*, **71**, Article 101966. <https://doi.org/10.1016/j.resourpol.2020.101966>
- [13] 周锦岚.原油价格不确定性:传导路径、经济效应及货币政策应对[D]:[博士学位论文].长春:吉林大学,2023.
- [14] 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,2004,39(10):75-83.
- [15] 王朝晖,李心丹.我国投资者情绪波动性与股市收益[J].宁波大学学报(人文科学版),2008,21(6):89-93,98.
- [16] 赖琴云,朱敏,陈其安.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.
- [17] 张天顶,施展.国际大宗商品价格冲击对中国股票市场的影响[J].南开经济研究,2022(11):59-74,114.