

投资者情绪对开放式股票型基金业绩的影响研究

陈泽时

上海理工大学管理学院, 上海

收稿日期: 2024年2月22日; 录用日期: 2024年3月13日; 发布日期: 2024年4月26日

摘要

随着我国开放式基金市场的发展, 其中以开放式股票型基金为主流, 其管理的资产规模也越来越大。同时, 参与到基金市场的个人投资者越来越多, 我国基金市场的投资者结构中散户占比较高, 使得投资者情绪更容易引起一定的非理性投资行为。以基金吧为主的投资相关网络论坛, 成为个人基金投资者发表意见、交流观点的重要平台。因此, 本文用爬虫技术获取东方财富天天基金网基金吧2017~2022年的帖子信息, 通过BERT模型对文本信息进行分类预测, 进而构建投资者情绪指数, 最后利用面板回归模型, 研究投资者情绪对基金业绩的影响。研究发现, 投资者情绪对开放式股票型基金业绩有显著负面影响。

关键词

投资者情绪, 基金业绩, 机器学习

Research on the Impact of Investor Sentiment on the Performance of Open-End Equity Funds

Zeshi Chen

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 22nd, 2024; accepted: Mar. 13th, 2024; published: Apr. 26th, 2024

Abstract

With the development of China's open-ended fund market, in which open-ended equity funds are the mainstream, the size of its assets under management is also getting bigger and bigger. At the

same time, there are more and more individual investors involved in the fund market, and the investor structure of China's fund market accounts for a relatively high proportion of retail investors, making it easier for investor sentiment to cause certain irrational investment behavior. Investment-related network forums, mainly fund bars, have become an important platform for individual fund investors to express their opinions and exchange views. Therefore, this paper uses crawler technology to obtain the post information of fund bar of Oriental Wealth TianTian Fund Network from 2017 to 2022, classifies and predicts the text information through BERT model, and then constructs the investor sentiment index, and finally utilizes the panel regression model to study the impact of investor sentiment on fund performance. It is found that investor sentiment has a significant negative impact on the performance of open-end equity funds.

Keywords

Investor Sentiment, Fund Performance, Machine Learning

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

根据中国证券业基金业协会数据显示,截止至 2023 年 8 月,我国公募基金市场上现存续基金 11,111 支。其中开放式股票型基金数量为 2125 支,占市场基金总量的 19.12%,管理总资产净值达到 28,452.29 亿元。可见,开放式股票型基金在我国证券市场发挥着重要作用。我国基金市场是以散户投资者为主导,与机构投资者相比,个人投资者在基金市场的参与更为活跃,同时也影响了市场的流动性和短期波动。随着互联网信息媒介的发展,互联网社交媒介平台,例如股吧、基金吧等成为个人投资者获取信息和发表个人观点的主要平台,这些平台能够形成一个整体的投资者情绪,而投资者情绪能在一定程度上解释或者预测市场收益,投资者通过因情绪导致的价差来获取超额收益[1]。本文通过爬虫获得基金吧的发帖文本,在文本情感分类方面利用 BERT 预训练语言模型进行训练词向量,能够获得比其他模型更好的分类效果[2],以此来构建投资者情绪指标。在我国基金市场上,投资者行为更为关注 Fama-French 五因子模型的业绩[3],本文将使用经过五因子模型调整后的超额收益作为基金业绩,实证研究投资者情绪指数对基金业绩的影响。

2. 投资者情绪指标

2.1. 文本数据来源

根据天天基金网基金吧的发帖总量作为活跃排名,选取最活跃的 100 支股票型基金,通过 Python3.7 实现 Scrapy 框架爬取 2017 年 1 月 1 日至 2022 年 12 月 31 日之间的帖子信息,具体抓取字段有帖子标题、帖子作者昵称及 ID、帖子时间。从帖子信息来看,个人投资者们倾向将帖子主要内容写在标题部分,而帖子正文多仅重复标题内容,或为“如题”等无效信息。主贴标题部分能够将个人的观点表达清楚,有利于进行情感识别。而且,主贴内的回帖大多千篇一律,重复内容较多,回复简略,灌水严重多为“顶”、“支持楼主”等信息,这些简略的回帖让人难以判断其情感倾向,所以在抓取评论内容时,本文只针对主贴标题部分的内容。

考虑到数据样本包含大量文本噪音,为保证数据的有效性,爬取天天基金网基金吧的帖子的同时,

进行以下数据清洗：剔除“公告帖”、“新闻贴”、“官方机构账号发帖”、帖子内容为“数字”、“符号”和“图片及表情”等帖子，同时剔除个人发帖量后四分位数的个人发帖用户，也即剔除掉发帖量小于等于三个的用户，来保证用户在基金吧是活跃积极的。尽可能得到的是相对活跃的个人投资者发布的纯文本信息。最终进行情感分析的帖子数量为 1,134,358 条。

2.2. 情感分类处理

从中随机抽取 10,000 条文本，参考文献[4]的标注方法，邀请 10 位金融从业人员进行情绪标注，将情感标签分为三类，积极情绪标注 1，中立情绪标注 0，消极情绪标注-1，每人标注 2000 条，确保每条发帖文本都有两位人员标注，对于标注结果相异的文本，请第三位人员标注，取三次标注结果中占多数标签作为该文本最终标签。

本文使用专门用于中文自然语言处理任务的预训练模型 `chinese-roberta-wwm-ext`，这是一个基于 RoBERTa 架构下开发，其中 `wwm` 代表 Whole Word Masking，即对整个词进行掩码处理，通过这种方式，模型能够更好地理解上下文和语义关联，提高中文文本处理的准确性和效果。模型评价指标包括准确率 (Accuracy)、F1 分数 (F1 score)。准确率表示模型输出的标签正好和测试集中的标签相等的个数占整个测试集样本数的比例。F1 分数是统计学中用于衡量二分类或多分类中模型精确度的指标，由于其同时兼顾了分类模型的查准率和召回率，往往被用于评价样本不平衡情况下的模型表现。由表 1 可以看出模型精度良好。

Table 1. BERT model assessment results

表 1. BERT 模型评估结果

Accuracy	Macro-F1	Weighted-F1
75.8%	72.4%	74.8%

2.3. 情绪指标构建

根据文献[5]的构建方法定义投资者情绪指数，将基金 i 在第 t 季度的投资者正向情绪的发帖量用 $Positive_{i,t}$ 来表示，基金 i 在第 t 季度的投资者负向情绪的发帖量用 $Negative_{i,t}$ 来表示。从而构造投资者情绪指数：

$$Sent_{i,t} = \ln \frac{1 + Positive_{i,t}}{1 + Negative_{i,t}} \quad (1)$$

当数值大于 0 时，表示当前季度投资者情绪为积极；当数值等于 0 时，表示当前季度投资者情绪为中立；当数值小于 0 时，表示当前季度投资者情绪为消极。

3. 实证结果与分析

3.1. 模型构建与变量选取

本文将衡量开放式股票型基金业绩($return$)作为被解释变量，将投资者情绪指数($sent$)作为解释变量建立了回归模型。

$$\begin{aligned} return_{i,t} = & \beta_1 * sent_{i,t-1} + \beta_2 * size_{i,t-1} + \beta_3 * familysize_{i,t-1} + \beta_4 * age_{i,t-1} \\ & + \beta_5 * shrindex_{i,t-1} + \beta_6 * std_{i,t-1} + \beta_7 * fee_{i,t-1} + \delta_1 \end{aligned} \quad (2)$$

其中被解释变量基金业绩计算方法[3]如下：

$$Return_{i,t}^{rew} - rf_t = \alpha_{i,t} + \beta_i (R_{m,t} - rf_t) + s_i^E SMB_T + h_i^E HML_T + r_i^E RAW_T + c_i^E CMA_T + \varepsilon_{i,t} \quad (3)$$

$$R_{i,t}^{FEfive} = \alpha_{i,t} + \varepsilon_{i,t} \quad (4)$$

$Return_{i,t}^{rew}$ 为基金 i 在第 t 月的原始收益率，指基金的复权单位净值增长率， rf_t 为该月的无风险收益率， $(R_{m,t} - rf_t)$ 为市场风险溢价因子，采取流通市值加权的方式， SMB_T 、 HML_T 、 RAW_T 、 CMA_T 为使用流通市值加权的办法计算的因子，以上五者的相关月度数据均可以在国泰安数据库中获取。之后通过对自变量与因变量进行多元统计回归，计算得出截距项 $\alpha_{i,t}$ 与残差项 $\varepsilon_{i,t}$ 的值，相加即可得出经 Fama-French 五因子模型调整后的超额收益。如式 4。

下一步将计算出的超额收益率数据频度调整为季度数据，采取累计超额收益率的计算方法来计算，具体见公式 5：

$$Return_{i,t} = (1 + R_{i,3t-2})(1 + R_{i,3t-1})(1 + R_{i,3t}) - 1 \quad (5)$$

其中， $Return_{i,t}$ 为基金 i 在第季度的超额收益率， $R_{i,3t}$ 为基金 i 在第第 $3t$ 月的超额收益率，因此，采取此方法可以分别得出经不同风险因子调整后的超额收益率，即本文的被解释变量基金业绩。

模型中的变量及其定义见表 2。

Table 2. Variables and definitions

表 2. 变量及其定义

名称	指标	定义
基金业绩	$return_{i,t}$	五因子模型调整后的超额收益率
投资者情绪	$sent_{i,t-1}$	基金吧发帖文本构建
市场环境	$shrindex_t$	第 t 季度与 $t-1$ 季度上证指数的变化率
基金规模	$lnsize_{i,t-1}$	基金 i 在 $t-1$ 季末的净资产值规模自然对数
基金家族	$lnfamilysize_{i,t-1}$	基金 i 所属管理公司在 $t-1$ 季末净资产值自然对数
波动率	$std_{i,t}$	基金 i 在 t 季度日收益率标准差
基金费率	$fee_{i,t}$	基金 i 在 t 季末的费用总额与基金季末份额的比值
基金年龄	$lnage_{i,t-1}$	基金 i 成立至 $t-1$ 季末的天数自然对数

3.2. 回归分析

本文使用面板数据，根据 F 检验和 Hausman 检验结果，选择固定效应模型作为本文的模型。使用 STATA16 获得如表 3 回归结果。从回归结果来看，模型的决定系数分别为 43.1%，说明本文构造的投资者情绪指数对开放式股票型基金业绩有一定的解释能力，模型具有较好的拟合效果。

表 3 回归实证结果表明投资者情绪对开放式股票型基金的业绩有一定的负向影响，投资者情绪指数的系数为 -0.002，意味着当投资者情绪指数提高 1 个单位时，开放式股票型基金的业绩可能会下降 0.002 个单位。这可能是由于当投资者投资市场的情绪高涨时，羊群效应导致投资者盲目跟随市场上主流决策，可能会导致市场资金增多，加大基金市场的波动，导致基金的成分股波动受到影响，从而影响到基金的策略。

3.3. 稳健性检验

在稳健性检验方面，本文考虑更换被解释变量基金业绩的衡量指标，测算相对收益率。

$$Return_{i,t}^{relative} = Return_{i,t}^{raw} - eastindex_t \quad (6)$$

其中 $Return_{i,t}^{raw}$ 为基金 i 在 t 季度的原始收益率， $eastindex_t$ 为 t 季度东财基金普通股股票型基金指数收益率。

Table 3. Regression results of investor sentiment on fund performance

表 3. 投资者情绪对基金业绩回归结果

解释变量	被解释变量 return
sent	-0.002*** (-2.87)
shindex	0.013** (2.14)
size	-0.011*** (-4.98)
familysize	-0.006 (-0.58)
std	-0.034*** (-4.28)
fee	-0.001* (-1.95)
age	0.020*** (2.63)
Constant	-0.054 (-0.55)
固定效应	YES
Observations	1796
R-squared	0.431

其中*** $p < 0.01$ ，** $p < 0.05$ ，* $p < 0.1$ 。

Table 4. Robustness test regression results

表 4. 稳健性检验回归结果

解释变量	被解释变量 relative return
sent	-0.010*** (-6.16)
shindex	0.480 (1.23)
size	0.006* (1.81)

续表

family size	-0.006 (-0.32)
std	0.023*** (4.46)
fee	0.006* (1.66)
age	0.023** (1.78)
Constant	-0.117 (-0.66)
固定效应	YES
Observations	1796
R-squared	0.504

其中*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

使用相对收益率替换被解释变量对模型进行稳健性检验, 结果见表 4, 显示投资者情绪系数显著性水平并未大幅变化, 并且其系数方向并未发生改变, 其他控制变量显著性和系数方向与原模型基本一致, 该模型具有稳健性。

3.4. 异质性分析

3.4.1. 不同情绪状态

考虑到两种不同的情绪状态对基金市场的影响可能具有非对称性, 本文进一步从积极和消极两种情绪状态下投资者情绪对基金业绩影响差异进行研究。

$$\begin{aligned} return_{i,t} = & \beta_1 * sent_{i,t-1} * (1 - D_{i,t}) + \beta_2 * sent_{i,t-1} * D_{i,t} + \beta_3 * size_{i,t-1} + \beta_4 * familysize_{i,t-1} \\ & + \beta_5 * age_{i,t-1} + \beta_6 * shrindex_{i,t-1} + \beta_7 * std_{i,t-1} + \beta_8 * fee_{i,t-1} + \delta_1 \end{aligned} \quad (7)$$

其中 $D_{i,t}$ 为虚拟变量, 当 $sent_{i,t-1} > 0$ 时, $D_{i,t} = 1$, 否则 $D_{i,t} = 0$ 。

Table 5. Heterogeneity test (different emotional states)

表 5. 异质性检验(不同情绪状态)

解释变量	被解释变量 return
Sent * D	0.0072*** (4.74)
Sent * (1-D)	-0.0044*** (-3.07)
shr index	0.0223** (2.11)
size	-0.0110*** (-4.98)

续表

family size	-0.0061 (-0.57)
std	-0.0334*** (-2.79)
fee	-0.0005 (-1.64)
age	0.0199 (1.50)
Constant	-0.0542 (-0.55)
固定效应	YES
Observations	1796
R-squared	0.0972

其中*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

从表 5 回归结果来看, 在看涨情绪中, 投资者情绪指数与基金业绩之间存在显著正向影响, 投资者情绪指数的回归系数为 0.0072, 而在看跌情绪中, 投资者情绪指数与基金业绩之间则存在显著负向影响, 投资者情绪指数的回归系数为 0.0044, 表明看涨情绪对基金业绩的影响为积极影响且影响能力更大。原因可能在于, 中国基金散户投资者的投机行为较多, 当投资者出于看涨情绪中, 投资者的非理性因素加大, 盲目跟风现象明显, 短期内大幅提升基金认购份额, 拉升基金业绩, 产生积极影响, 同时对行情小道消息更为敏感, 会随时准备高位套现自己手中的基金, 从而加大了看涨情绪对基金业绩的作用。当在看跌情绪中, 由于处置效应可能影响投资者的交易决策行为, 使他们在市场上更倾向于出售盈利的投资而持有亏损的投资, 投资者更倾向于风险喜好, 不愿意认识到自己的亏损, 进而拒绝实现亏损, 同时会采取反向投资策略, 这可能导致投资者的投资行为过热, 市场出现过度卖出行为, 进而影响市场价格的波动。

3.4.2. 不同股东性质

根据东方财富网基金所属基金公司股东和资本金数据信息, 将基金分为中资基金和合资基金, 其中中资基金 31 支, 所属基金公司包括汇添富、易方达、创金合信、前海开源、南方、广发等; 合资基金 69 支, 所属基金公司包括富国、工银瑞信、海富通、摩根、汇丰晋信、中欧等。分别对中资基金和合资基金的样本进行回归。

从表 6 回归结果可以看出, 投资者情绪对中资基金和合资基金的显著性方向一致, 均为负向影响。投资者情绪和对中资基金的回归系数分别为 0.0044, 大于对合资基金的 0.0014, 说明相比于合资基金, 国内基金市场的个人投资者情绪对中资基金的影响程度更大。原因可能在于锚定效应扩大了基金的品牌效应, 投资者对于历史表现业绩突出和管理团队优异的基金更加有信心, 再之后的投资决策中也更容易受到初始印象的影响。从结果来看, 投资者对于持有中资基金的稳定性和投资信心更弱, 更容易受到市场噪音的影响, 增加非理性投资行为, 导致市场波动, 从而对中资基金业绩产生较大的影响。

Table 6. Heterogeneity test (different shareholders)
表 6. 异质性检验(不同股东性质)

解释变量	被解释变量 return	
	中资	合资
sent	-0.0044*** (-9.90)	-0.0014*** (-4.37)
shrindex	0.5579* (2.18)	-0.3852 (-1.41)
size	-0.0178*** (-5.79)	-0.0074** (-2.14)
familysize	-0.0276* (-1.73)	-0.0078 (-0.53)
std	-0.0113*** (-2.54)	-0.0501*** (-4.49)
fee	-0.0001* (-0.54)	0.0114 (-0.0115)
age	0.0618*** (3.72)	-0.0089 (-0.50)
Constant	-0.1375 (-1.11)	0.1496 (0.99)
固定效应	YES	YES
Observations	601	1,195
R-squared	0.1204	0.1382

其中*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

4. 结论与建议

4.1. 结论

本文得出了以下几点结论。

1) 投资者情绪对基金业绩有显著负向影响。在季度阶段, 投资者情绪指数与开放式股票基金业绩之间存在反向关系, 过度自信、反应过度 and 羊群效应带来的非理性投资者情绪高涨可能会导致开放式股票型基金业绩出现负的收益变化。

2) 在看涨情绪中, 投资者情绪对基金业绩显著正相关, 看跌情绪中, 投资者情绪对基金业绩显著负相关。原因可能在于市场普遍出于看跌环境, 在市场反应积极的情况下, 投资者处于看涨会大量申购基金, 增加基金流量, 在季度范围内对基金业绩产生一定的正向影响。在看跌情绪下, 处置效应可能导致大量基金份额卖出, 资金抽回也可能导致基金业绩进一步下降。

3) 投资者情绪对中资基金产生更大的影响。相比与合资基金公司旗下的基金, 国内投资者更看重中资基金的长期稳定性和品牌规模, 并且在市场波动下对中资基金的投资信心下降和情绪不乐观, 可能出现大量买卖行为影响中资基金的业绩。

4.2. 建议

监管部门应及时建立基金市场情绪波动监测机制，关注投资者情绪影响基金市场的机制和途径，从源头机理上杜绝大涨大跌的市场现象，适时采取相应措施来平抑市场波动，降低市场风险，预防非理性情绪对基金市场造成的冲击。

在基金的流动性风险角度上，开放式股票型基金的管理团队应该注重投资者情绪，因为投资者情绪的波动将会导致基金过度的申购和赎回波动，并进而影响基金的资金流量，影响基金的投资稳定性。基金管理团队应该加强风险防范的研究，提前建立合理的预案，防范基金资金流动性风险。在有预案的基础上合理进行配置资产组合，从而提高基金对抗流动性风险的可能，有效应对基金流动性风险的变化。

对于个人投资者，应该主动合理地转变投资观念。在投资基金市场的过程中避免盲目追求投机，频繁改变资产配置，以期赚取短期利益，应当多方面分析信息渠道和信息质量，同时丰富增强自身投资相关知识，合理做出理性行为，以一种健康的投资理念进入基金投资市场。健康的投资理念应该是将各类金融市场当作个人闲置资金资产配置的市场，对冲通胀风险与资产折旧风险，从而能够合理地应对市场波动。对于机构投资者，应该加强对投资者的金融知识普及，与投资者建立长期互惠的委托关系。

参考文献

- [1] 蒋玉梅, 王明照. 投资者情绪与股票收益: 总体效应与横截面效应的实证研究[J]. 南开管理评论, 2010, 13(3): 150-160.
- [2] 方英兰, 孙吉祥, 韩兵. 基于 BERT 的文本情感分析方法的研究[J]. 信息技术与信息化, 2020(2): 108-111.
- [3] 李志冰, 刘晓宇. 基金业绩归因与投资者行为[J]. 金融研究, 2019(2): 188-206.
- [4] 部慧, 解峥, 李佳鸿, 吴俊杰. 基于股评的投资者情绪对股票市场的影响[J]. 管理科学学报, 2018, 21(4): 86-101
- [5] 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>