

在线社区支持倾向对股市收益和波动的影响^①

钱宇, 李子饶, 李强, 袁华*

(电子科技大学经济与管理学院, 成都 611731)

摘要: 互联网技术的飞速发展使得非专业的个体投资者可以通过在线金融社区分享信息并表达倾向. 本文利用东方财富网的5 178 824条用户评论, 运用卷积神经网络的分类算法, 提取并测量了在线用户对未来市场看涨或看跌的支持倾向, 并从市场收益和波动两个方面, 检验了在线用户支持倾向及其一致性对股市的影响. 结果表明, 在线用户当期支持倾向对未来的股市收益具有显著的负向影响, 支持倾向的一致性则会放大市场的波动幅度. 进一步的证据表明, 用户支持倾向很大程度上是基于股票市场的历史表现而形成, 且具有一定的“羊群效应”.

关键词: 文本大数据; 在线社区; 支持倾向; 市场收益; 市场波动

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2020)02-0140-15

0 引言

有效市场假说指出, 资产价格能够反映资产基础价值的信息, 信息的不断揭示导致资产价格的变化^[1]. 早期大量的文献针对公司财务报告^[2]、并购重组等重大事项公告^[3]、分析师调研报告^[4]等的信息披露, 通过考察股票的短期市场反应, 对有效市场假说进行了实证检验. 然而, 这些相对标准化的报告或公告不仅是按照特定的信息披露规则进行披露, 而且通常是由上市公司官方或专业的金融机构进行发布. 近年来, 信息科技的飞速发展和深度融合把人们带入大数据时代, 数据呈现出体量大、多样性、密度低、速率高等属性特征^[5]. 特别是社交媒体体验分享的用户生成内容(user generated content, UGC), 对投资者尤其是非专业的“个体投资者”的信息获取与处理能力提出了巨大挑战.

越来越多的投资者通过在线金融社区(论坛、股吧)发表自己对于市场的评论, 海量评论中蕴含着情绪和支持倾向等丰富的信息, 并在以非专

业投资者为主的用户之间分享传递. 有证据^[6,7]表明, 人与人在社会互动过程中的信息获取与意见交换, 以及乐观或悲观的情绪传播, 对投资者的金融决策具有重要影响^[8]. Antweiler和Frank^[9]对股票留言板评论情绪的研究, 以及Chen等^[10]对金融社交媒体中的用户股评以及用户相互讨论形成的群体智慧的研究也支持了上述观点.

然而, 区别于年报、公告、新闻报道等具有相对正式规范的信息披露形式, 以及具有信息内容密度高、可靠性强等特点的传统信息来源, 在线金融社区的用户评论信息不仅形式上具有短文本、不规范、错误多、口语化的特点, 而且内容上也呈现出信息价值密度低、可靠性弱等特点^[11,12], 加之在线金融社区用户通常是整体专业水平较低的个体投资者. 为此, 如何充分利用在线金融社区的海量文本信息, 提取并测度个体投资者对未来市场走势的看法, 进而形成看涨或看跌的支持倾向? 以及非专业个体投资者整体的支持倾向能否揭示未来市场的收益和波动状况? 上述问题的解决,

① 收稿日期: 2019-06-04; 修订日期: 2019-12-29.

基金项目: 国家自然科学基金资助项目(71490723; 91846105; 71572029; 71671027; 71972027).

通讯作者: 袁华(1973—), 男, 四川达州人, 博士, 教授, 博士生导师. Email: yuanhua@uestc.edu.cn

可以帮助投资者构建相应的投资策略,不仅对于丰富大数据分析技术在金融领域的应用,而且对于投资者决策实践都具有重要的意义。

需要指出的是,本文主要讨论的是用户的支持倾向而不是朴素的用户情绪,前者是指用户对未来市场看涨看跌的预期,包含了更多的用户投资行为相关的信息;而后者是指一种对市场普遍的乐观或悲观的心态,是更为朴素的个人情感^[13,14]。本文将在控制用户情绪的条件下讨论用户支持倾向对股市收益和波动的影响,这与以往基于用户情绪的研究有所不同。

本文利用2017年1月1日~2018年9月30日期间东方财富网的5 178 824条用户评论数据,首先基于深度学习的卷积神经网络分类算法提取并测度在线用户评论的支持倾向,并将之与传统分类算法效果进行比较。然后,在控制用户情绪的条件下,利用分布滞后模型、GJR(1,1)模型,考察了用户支持倾向对沪深300指数成分股的收益和波动的影响。进一步,本文对日间和日内用户支持倾向的成因进行了探讨,结果表明历史市场表现是用户形成支持倾向的重要原因。本文的创新点主要有三点:第一,利用卷积神经网络的分类算法,识别测度了在线用户评论中的用户支持倾向,较之以前的传统分类算法具有更高的精确度和召回率;第二,考察用户支持倾向对市场收益率和波动性的影响时引入了市场非对称效应的影响,并讨论了用户支持倾向的成因,提供了用户支持倾向具有“羊群效应”的证据;第三,将用户支持倾向和用户情绪区分开来,弥补了以往研究中将二者混淆的不足。

1 文献综述

1.1 金融文本挖掘相关研究

早期的金融文本挖掘的分析方法较为简单,主要是基于单词级别的词频统计。例如,Li使用“lingua:en:fathom”软件计算公司年报中相关字符、单词、句子和段落的频率,并构建可读性指标FOG(迷雾指数),实证发现市场参与者乐观或悲观的情绪对股票价格具有重要影响^[2]。此外,在

文本向量表征上,TF-IDF向量模型被广泛用于金融文本挖掘^[15,16]。

金融文本信息提取的研究对象主要集中在新闻媒体、专业机构发布的各种形式的报告等高质量文本。这些高质量的文本在形式上具有表达正式、规范化的特点,其信息密度高,内容可靠性高。相比之下,本文的研究对象是在线金融社区,其主要用户是以“散户”为主的个体投资者,整体的信息解读、处理和分析能力较弱进而专业素质水平较低。此外,加上网络社区的开放性和缺乏监督性,用户评论往往具有短文本、不规范、文字错误、口语化、信息密度低、不可靠等特点^[11,12],这些都使得从文本中提取出用户支持倾向更加困难。

1.2 金融社区、金融社交媒体的相关研究

Tumarkin和Whitelaw对RagingBull.com论坛的研究发现,在留言板发生积极的情绪变化之前,股价出现小幅异常上涨,事件日的留言板意见和异常回报相关,且交易量在事件日当天大幅增加,并在之后一个交易日保持高位。然而,该研究认为很少有证据表明用户意见能预测未来市场的回报^[17]。Antweiler等对Yahoo! Finance和RagingBull.com股票留言板上的信息进行研究,他们发现留言信息的分歧有助于预测交易量,发帖活动也有助于预测股价的波动率^[9]。Das和Chen基于五种分类器开发了一种从股票留言板中提取投资者情绪的算法,研究发现从留言板评论提取出的情绪信息与个股股价之间并没有很强的关系^[18]。Chen对美国最大的投资类社交媒体网站Seeking Alpha的文章和评论区进行分析,在控制了金融分析报告、新闻报道等传统信息源后,发现社交媒体网站的信息仍然能够提供价值可观的信息,能够有效地预测投资回报率和惊喜收益^[10]。Renault对股评社区(stocktwits.com)的消息进行情感分析,利用开盘后30min的在线投资者情绪成功预测了收盘前30min的标普500股指收益,并认为这种预测力来源于新手交易者^[19]。Eli等对Twitter中个人账号发表的意见进行了正负情感极性分类,在控制了其他决定收益的因素之后,发现Twitter的综合意见有助于预测季度收益以及收益

公告发布前后的异常回报^[20]。然而,以往的研究在进行信息提取时往往关注的是用户情绪,甚至将情绪与支持倾向混淆对待,在本文后续章节中将看到情绪与支持倾向的不同之处。

国内也有一些针对股票社区/社交媒体的研究工作,但大多数研究都基于简单的变量且高度依赖人工处理数据。刘海飞等基于关注度、信赖度和更新频率来衡量微博信息质量,研究了微博平台披露信息质量与股价的同步性^[21]。丁慧等研究了社交媒体“上证e互动”中问答互动的数量与股价崩盘事件的关联^[22]。俞庆进和张兵基于创业板股票的百度指数来构建投资者关注度指标,研究了其对市场收益率、换手率和成交量的影响^[23]。何贤杰等利用上市公司在新浪微博上发布的信息,讨论了上市公司在网络新媒体上披露的信息内容对股价同步性的影响^[24]。孙书娜等利用雪球社区用户的自选股信息构建了日度超额雪球关注度指标,对投资者关注的行为特征、投资者关注与市场表现之间的相关关系和因果关系进行了研究^[25]。与本文研究内容最相近的是部慧等的研究,他们用朴素贝叶斯分类算法提取了东方财富网股吧贴文中的投资者情绪,研究了投资者情绪对市场收益率、波动率和交易量的影响^[26]。然而,与他们的不同之处在于,本文采用卷积神经网络的分类算法分析在线金融社区用户评论,提取出看涨或看跌的用户支持倾向,在控制用户情绪的条件下利用分布滞后模型和GJR(1,1)模型,考察在线用户评论信息反映的支持倾向对股市波动和收益的影响,并进一步讨论了用户支持倾向的成因,提供了用户支持倾向具有“羊群效应”的证据。

2 数据、算法与指标

2.1 数据描述

艾瑞咨询发布了基于日均覆盖人数的《2017年6月PC网站榜单top1000》^[27],在财经类网站中,东方财富网以1071万的日均覆盖人数高居第一,是第二名同花顺的近5倍,遥遥领先于同行,是中国目前最大的在线股票社区。为了保证实验

数据具有代表性,本文选取2017年1月1日~2018年9月30日期间东方财富网中沪深300成分股的股票评论。剔除部分因停市、网站维护等原因导致用户评论不完整或无法获取的股票后,最终得到269只股票共5178824条评论数据,每条评论的平均字数为39,每只股票平均的评论数为19252条。

本文使用jieba模块对评论进行分词。为了准确识别金融语境下的词汇,本文在分词模块中加入了搜狗细胞词典中的财经金融词汇大全、财会术语、股票基金词库大全以及从东方财富网爬取的所有股票名称、股票代码构成的股票信息词库。Mikolov等提出的Word2vec模型是将文本转换为数值向量的最流行和最有效的方法之一^[28]。本文使用python中的gensim模块训练Word2vec模型并生成词向量。经测试,语料库中的最佳模型训练参数是skip-gram模型,词表征维度为300,窗口大小为6,其余都是默认配置。

本文对投资者支持倾向的定义为投资者对某一标准的预期,当投资者预期回报高于“平均水平”时,支持倾向为看涨;当投资者预期回报低于“平均水平”时,支持倾向为看跌,且这个“平均水平”因人而异^[29]。将情绪定义为投资者普遍的乐观和悲观情绪^[13,14],这种定义更接近于朴素的个人情感。

在标注数据时,本文判断支持倾向和情绪的基准是:评论中是否包含对未来市场或涨或跌的信息。表1给出了一些实例。有的评论中只包含用户纯粹的情绪宣泄,并不包含对未来市场涨跌信息的判断,如表1中序号①~④所示,这种特征正是在线金融社区用户评论信息价值密度低的体现;有的评论在宣泄情绪的同时也表达了对未来市场看涨或看跌的个人看法,如表1中序号⑤~⑧所示。

在用户评论中,同时包含了用户朴素的乐观/悲观的情感和对未来行情看涨看跌的期望,本文更关注后者而不是前者。在后续分类任务中,将用户对未来市场的支持倾向分为看涨、看跌和无明确倾向三种,将用户情绪分为正向、负向和无情绪三种。

表 1 评论中用户支持倾向与情绪示例

Table 1 Examples of user support tendency and sentiment

序号	文本	情绪	支持倾向
①	买入长安汽车被套住已经一年多了,明天终于可以解套了谢谢谢谢!	正向	无
②	今天这条大阳线爽歪歪,给主力点赞,下班可以庆祝一下了~	正向	无
③	奔跌停我傻,人家让清仓没照办直接多亏了5万,一直亏一直哭	负向	无
④	哭哭哭在最低点割肉了唉!一割就涨什么意思?	负向	无
⑤	还好换了换雄安龙头了,不然就被要套牢了	正向	看跌
⑥	明日再来个跌停,本股神会接点筹码,主力们你们是最可爱的人了鼓掌!!	正向	看跌
⑦	这是要起飞了你妹的,昨天33块卖的,明天可能涨五个点,个人看法买卖自便删除自选	负向	看涨
⑧	唉,抛在了黎明前,自认技不如人退场,剩下的朋友们有福了	负向	看涨

2.2 CNN 分类算法

卷积神经网络(convolutional neural network, CNN)是一类包含卷积计算且具有深度结构的前馈神经网络(feedforward neural networks),是深度学习的代表算法之一,广泛应用于图像识别、语音处理和自然语言处理^[30],并在情绪分类问题上实

现了较好的效果。

本文参考 Kim^[31] 的 CNN 情感分类模型,搭建了包含一层卷积层、一层池化层和全连接层的 CNN 框架,如图 1 所示. 卷积层的输入为词向量化后 $m \times n$ 的句子矩阵(行的大小 m 为句子长度、列的大小 n 为词向量的维度). 对于文本处理来说,卷积核的大小一般取 $x \times n$,以步长 i 从第一行向下滑动,每次滑动都会得到一个卷积结果,滑动到最后一行便得到一个 $m-x$ 维的向量(即特征图). 卷积核的大小参数取决于文本类型,若文本越长、语义信息越复杂,卷积核越大. 同时,为了捕获更为丰富的语义信息,需要设置几种不同大小的卷积核,且卷积核的数目应该足够多. 每一个卷积核在输入的 $m \times n$ 句子矩阵滑动后,得到的 $m-x$ 维的向量作为池化层的输入,池化层中预设的池化函数将对特征图进行特征选择和信息过滤. 文本处理中的池化层一般采用 k -max pooling 的方法,输出一个 k 维的特征向量. 将所有卷积核的池化层输出结果拼接起来,得到一个 $k \times num_filters$ 维的句特征向量. 经过卷积层和池化层,词向量被映射成高维的语义句向量,已能够很好地表征语义信息. 全连接层是一个浅层神经网络,通过标注的数据集和标注的结果就可以训练浅层神经网络分类器,输出支持倾向分类.

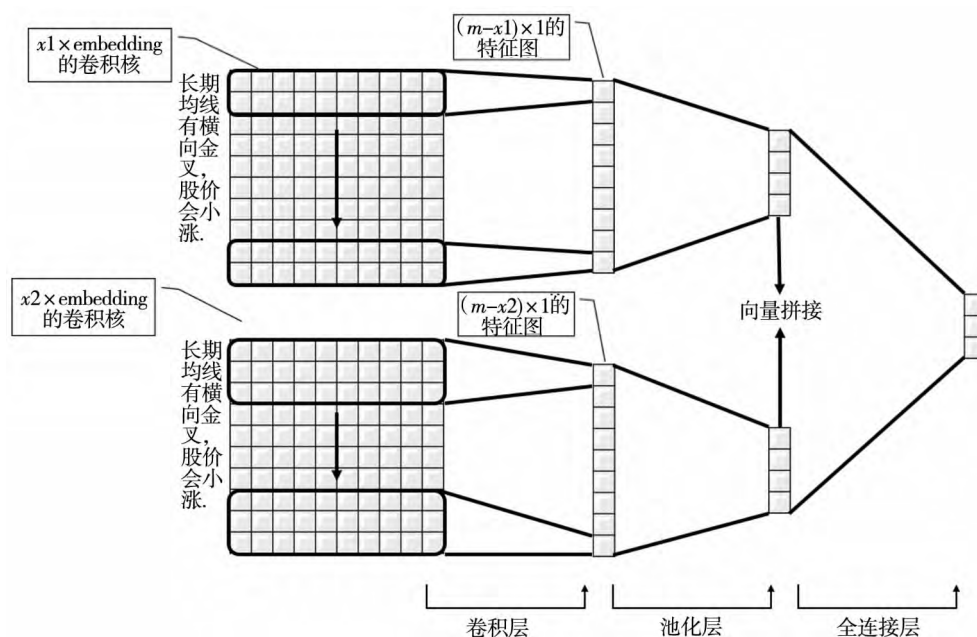


图 1 卷积神经网络结构图

Fig. 1 The structure of convolutional neural network

经测试,在本文语料库中表现最好的网络框架和参数为:卷积层的卷积核大小为 2×300 、 3×300 、 4×300 、 5×300 、 6×300 ,且每种卷积核的数目都使用了 100 个;卷积核以步长 1 在输入的矩阵滑动,从而生成特征图进入池化层;池化层采用 1-max pooling 的方法;所有卷积核经过池化层后的输出拼接成一个 500 维的高维句向量表征;500 维的高维句向量作为全连接层的输入,全连接层输出支持倾向分类。

为了训练模型,对于评论的标注按照自然语言处理标签标注的一般方法,由专家标注了 5 000 条用户评论中的支持倾向和情绪。为了说明标注的训练样本数量能够满足算法训练,本文通过逐步增加训练样本数的方式,来考察训练样本数的增加对训练效果的影响,图 2 展示了实验结果。可以看到,算法在训练样本为 3 000 左右就趋于收敛,说明标注的训练样本数量能够满足分类算法的训练。

为了保证模型的泛化能力和稳定性,实验对模型进行了 5 次 5 折交叉验证,结果表明 CNN 模型在用户支持倾向分类上能够达到 74.8% 的精

度和 74.7% 的召回率。进一步为了验证 CNN 分类算法的优越性,表 2 给出了 5 次 5 折交叉验证情形下该算法与其它常用分类器的对比结果。结果表明传统分类算法对于情绪分类效果较好,但对于支持倾向分类效果较差,这是因为支持倾向分类需要提取用户的潜在意图信息,比情绪分类任务更为复杂。相对于部慧等使用的贝叶斯算法^[26]、Cecchini 等^[16]使用的 SVM 或是其他经典分类算法^②,本文所提出的 CNN 算法都能实现更好的分类效果。

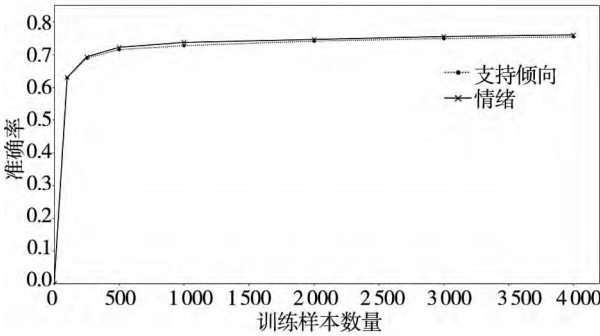


图 2 训练样本数量对训练效果的影响

Fig. 2 The impact of training samples' number on training effect

表 2 不同分类器模型对比结果

Table 2 Comparison results of different classifier models

分类模型	情绪			支持倾向		
	精度	召回率	F1 值	精度	召回率	F1 值
CNN	0.761	0.764	0.763	0.748	0.747	0.748
SVM	0.635	0.613	0.624	0.692	0.654	0.672
Naïve Bayes	0.636	0.515	0.569	0.642	0.536	0.584
Artificial Neural Network	0.707	0.715	0.711	0.675	0.682	0.678
Discriminant Analysis	0.672	0.691	0.681	0.654	0.669	0.661
Decision Tree	0.558	0.511	0.534	0.579	0.485	0.528
Random Forest	0.697	0.701	0.699	0.676	0.684	0.680
K-Neighbors Classifier	0.641	0.662	0.651	0.628	0.650	0.639

2.3 支持倾向序列构建

本文使用 CNN 算法识别了语料库中每条用户评论的支持倾向。在讨论这些支持倾向对市场的影响之前,需要将这些意见进行汇总,并构建一个支持倾向序列,因此借鉴了 Antweiler 和 Frank 的牛市信号指标(Bullishness Signal)^[9]。然而,牛市信号指标的不足在于,无法体现股票市场对信息反应的非对称效应^[34],即好消息和坏消息对股

价波动的影响存在差异。若坏消息对市场的影响比好消息对市场的影响更大,称之为正向杠杆效应;若好消息对市场的影响比坏消息对市场的影响更大,称之为反向杠杆效应^[35]。非对称效应已经从理论和实证研究中得以证实^[36,37]。与本文研究对象最为接近的是 Tetlock 的研究,他们的研究表明股票收益率对负面新闻的反应比对正面新闻的反应更加剧烈^[38]。从本文的研究对象——在线

② 算法详情参考 Tan 等的《数据挖掘导论》^[32] 和周志华的《机器学习》^[33]。

金融社区用户评论来看,相同程度的看涨倾向和看跌倾向对市场的影响是不一致的,而 Antweiler 和 Frank 的牛市信号指标认为相同程度的看涨倾向和看跌倾向对市场的影响是相同的,从而忽略了这种非对称效应. 因此,本文在结合 Antweiler 和 Frank 的研究基础上引入了股票市场中对信息的非对称效应的影响.

已经识别出每条评论中的看涨或看跌倾向,因此可以定义一段时间 $D(t)$ 内的看涨或看跌倾向为 $O_{t,c}$,如式(1)所示,其中 $c \in \{\text{看涨(bullish)}, \text{看跌(bearish)}, \text{无}\}$, $x_{i,c}$ 表示当第 i 条评论属于 c 时取值为 1,否则为 0; ω_i 表示第 i 条评论的权重. 本文更加关注的是表征看涨和看跌的支持倾向,可以利用一段时间内看涨看跌差值来表征这段时间内整体的支持倾向,此处加入市场非对称效应的调整因子 af (asymmetric factor),并进行归一化处理. 当 $af > 1$ 时,能够表征市场的正向杠杆效应,即人们对坏消息更敏感;当 $af < 1$ 时,能够表征市场的反向杠杆效应,即人们对好消息更敏感. 然而目前的指标还不够完善,它不能很好地表征支持倾向强度. 为此,本文考虑将看涨看跌的有效评论数加入支持倾向指标中^[9],得到式(2). 这样,支持倾向指标不仅能用其正负符号来表征该时期内整体的看涨或看跌的支持倾向,还能用其绝对值的大小表征该种支持倾向的强度

$$O_{t,c} = \sum_{i \in D(t)} \omega_i x_{i,c} \quad (1)$$

$$S_t = \frac{O_{t,bullish} - af \times O_{t,bearish}}{O_{t,bullish} + af \times O_{t,bearish}} \times \ln(1 + O_{t,bullish} + O_{t,bearish}) \quad (2)$$

式(1)中的 ω_i 表示该条评论可靠性程度或重要程

度,即具有较高专业素养的用户的意见会更加重要,因此其权重应该更大,而其他非专业个体投资者的权重应该更小. 然而,用户的个人专业信息很难被观测到,且在线金融社区的评论主体是专业素养较低的个体投资者,于是本文在实验中对所有的评论赋权为 1,即认为每条评论对市场的影响力相同^[9],这种赋权方式实质上捕捉的是广大个体投资者(散户)的支持倾向.

投资者的意见分歧被认为可能是交易的动机^[39,40],因此为了衡量这种意见分歧程度,本文直接引用了文献^[9]的一致性指数 A_t (agreement index),由式(3)给出. 一致性指数 A_t 是一个区间为 $[0,1]$ 的数,其值越大表明一致性程度越高. 为了后续说明用户支持倾向和情绪存在差异,以同样的方法计算了情绪和情绪一致性指数. 方便起见,用 $S1$ 和 $S2$ 分别表示支持倾向和情绪,用 $A1$ 和 $A2$ 分别表示支持倾向一致性和情绪一致性

$$A_t = 1 - \sqrt{1 - \left(\frac{O_{t,bullish} - O_{t,bearish}}{O_{t,bullish} + O_{t,bearish}} \right)^2} \quad (3)$$

结合构建的变量和沪深 300 指数的日收益率和日振幅数据,表 3(a) 给出了重要变量的描述性统计,从表中的单位根检验可知各重要变量均为平稳序列(在 5% 的显著性水平下). 表 3(b) 和表 3(c) 给出了支持倾向强度和支持倾向一致性与沪深 300 指数当期收益率、振幅的简单描述性统计. 从表中的结果可以看出,支持倾向与沪深 300 指数当期收益率正相关、与当期振幅负相关,支持倾向一致性与沪深 300 指数当期收益率负相关、与当期振幅正相关. 本文将在下一节进一步讨论支持倾向与市场表现的关系.

表 3(a) 重要变量描述性统计

Table 3(a) Descriptive statistics of important variables

	收益率(%)	振幅	支持倾向	支持倾向一致性指数	情绪	情绪一致性指数
均值	0.013 4	1.213 0	0.042 6	0.006	-0.624 5	0.019
中位数	0.040 7	1.034 9	0.058 1	0.002 5	-0.619 3	0.015 2
最大值	3.031 9	4.054 5	1.385 7	0.062 3	0.471 8	0.083 2
最小值	-4.272	0.349 5	-1.001 4	1.28E-07	-1.492 5	7.03E-06
标准差	0.945 7	0.641 6	0.418	0.008 4	0.302 7	0.014 7
偏度	-0.36	1.555 8	0.096 4	2.509 9	0.316 6	1.080 9
峰度	5.064 6	5.923 4	3.039 7	11.674 7	3.303 8	4.307 4
Jarque-Bera	85.058 8	324.317 3	0.689 2	1 787.156	8.777	113.553 8
统计量	(0.000 0)	(0.000 0)	(0.708 5)	(0.000 0)	(0.012 4)	(0.000 0)
单位根检验	-20.167 6	-5.056 4	-3.216 5	-5.416 8	-3.389 5	-3.485 2
	(0.000 0)	(0.000 0)	(0.019 7)	(0.000 0)	(0.011 9)	(0.008 8)
样本数	427	427	427	427	427	427

表 3(b) 支持倾向强度与收益率、振幅基本统计
Table 3(b) Basic statistics of support tendency, returns and amplitude

支持倾向强度	看跌(1) ($S1 \leq -0.5$)	(2) ($-0.5 < S1 \leq -0.25$)	(3) ($-0.25 < S1 \leq 0$)	(4) ($0 < S1 \leq 0.25$)	(5) ($0.25 < S1 \leq 0.5$)	看涨(6) ($0.5 < S1$)	t 值 (6) - (1)
收益率 (%)	-1.245	-0.373	0.060	0.196	0.379	0.504	1.749*** (8.976)
振幅 (%)	3.666	3.101	2.836	2.629	2.609	2.627	-1.039*** (-7.719)
样本数	45	51	102	103	65	61	

注：*、**、*** 分别代表 10%、5% 和 1% 的显著性水平，括号中为 t 检验的统计值，下同。

表 3(c) 支持倾向一致性与收益率、振幅基本统计
Table 3(c) Basic statistics of agreement index of support tendency, returns and amplitude

支持倾向一致性	低(1) ($A1 \leq 0.0005$)	(2) ($0.0005 < A1 \leq 0.001$)	(3) ($0.001 < A1 \leq 0.003$)	(4) ($0.003 < A1 \leq 0.006$)	(5) ($0.006 < A1 \leq 0.015$)	高(6) ($0.015 < A1$)	t 值 (6) - (1)
收益率 (%)	0.142	0.037	0.194	-0.040	-0.096	-0.287	-0.429*** (-2.402)
振幅 (%)	2.746	2.655	2.818	2.829	2.915	3.094	0.348*** (2.790)
样本数	87	47	96	60	83	54	

3 用户支持倾向对市场的影响

3.1 支持倾向对市场收益率的影响分析

当市场参与者在金融社区中发表自己的看法时，他们的支持倾向可能会产生投资交易活动，推动资产价格变化从而影响市场收益率。本节将讨论用户评论中的支持倾向信息能否影响市场收益率。

以支持倾向发生前 10 日作为基准，统计分析支持倾向发生前后 10 日的累计收益率，如图 3 所示。其中，当 af 取不同值时，横坐标之上的线和横坐标之下的线分别表示了支持倾向为正和为负的情形。

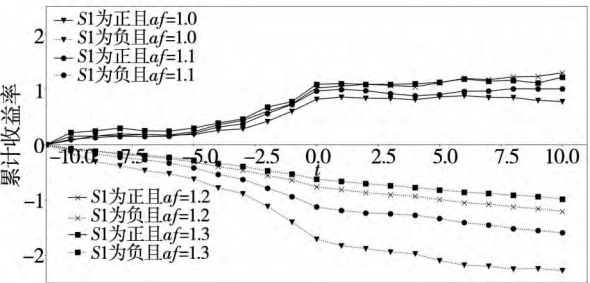


图 3 支持倾向发生前后 10 天的累计收益率
Fig. 3 The cumulative rate of return for 10 days before and after the support tendency

从图 3 中 $af = 1.0$ 的累积收益率可以看出，当看涨看跌倾向的程度相同时，市场对看跌的反

应程度更加剧烈，这验证了已有研究关于中国沪深股市存在正向杠杆效应的结论^[41]。因此，本文通过增大 af 来调整杠杆效应。通过式 (4) 单独讨论了支持倾向对市场收益率的非对称性效应的调节，其中 $S1_{t-i}$ 是支持倾向， r_{10} 是过去 10 日的累积市场收益率。表 4 给出了支持倾向对市场收益率的非对称性效应的回归结果。从结果可以看出，随着 af 值的增大，支持倾向对市场杠杆效应的刻画就越强，显著性水平和拟合优度也随之提升，说明市场确实存在杠杆效应，通过刻画杠杆效应而使得回归效果更优。在后续回归中 af 被设置为 1.8 (在 4.3 中进一步讨论了 af 的取值以及支持倾向对市场收益率的非对称效应)

$$r_t = c_0 + \sum_{i=0}^m \beta_i S1_{t-i} + \alpha r_{10} + \varepsilon_t \tag{4}$$

为了表征从支持倾向发生到市场发生反应这一滞后现象，与 Chen 等^[10] 类似，本文同样使用了分布滞后模型，其表达式由式 (5) 给出。其中， r_t 表示市场收益率， $S1_{t-i}$ 为支持倾向， $A1_{t-i}$ 为支持倾向一致性指标， $S2_t$ 为情绪， $A2_t$ 为情绪一致性指标， $message_{t-i}$ 为发帖量， X 为一组控制变量，其设定参考了已有研究^[42,43] 以及现实交易中的常用变量，包括：前 10 个交易日的累计收益率和平均波动率、交易量、换手率、市盈率、市净率、市销率、市现率、波动率， ε_t 是随机误差项。其中，市场支持倾向通过将个股等权相加整合而成，市场数

据则使用了沪深 300 指数相关数据(下同). 与 Antweiler 和 Frank 一样, 本文认为只有最近期的用户倾向信息与股票收益率相关. 为了更一般地探讨支持倾向在收益率回归中的影响, 设置的最大滞后期为 2. 回归以 2017 年 1 月 1 日 ~ 2018 年 9 月 30 日期间共 427 个交易日为样本. 表 5 给出了回归结果, 受篇幅限制, 此处只列出了最关心的变量结果(下同). 在表 5 中, 模型(1)给出了仅有简单控制变量的情形, 拟合优度较低; 模型(2) ~

模型(4)通过逐步加入支持倾向及其滞后期来观察支持倾向对收益率的影响, 支持倾向指标显著, 拟合优度逐渐提高, 说明支持倾向对于收益率有解释力; 模型(5)进一步加入了用户情绪信息作为控制变量, 结果依然稳健

$$r_t = c_0 + \sum_{i=0}^m \alpha_i S1_{t-i} + \sum_{i=0}^m \beta_i A1_{t-i} + \mu_1 S2_t + \mu_2 A2_t + \sum_{i=0}^m \gamma_i \ln(\text{message}_{t-i}) + X\delta + \varepsilon_t \quad (5)$$

表 4 支持倾向对市场收益率的非对称性效应的调节

Table 4 The adjustment of asymmetric effect of support tendency on market returns

	$af=1.0$	$af=1.2$	$af=1.4$	$af=1.6$	$af=1.8$
$S1_t$	1.371 *** (8.155)	1.401 *** (8.182)	1.444 *** (8.198)	1.496 *** (8.211)	1.555 *** (8.226)
$S1_{t-1}$	-0.586 *** (-4.632)	-0.611 *** (-4.76)	-0.64 *** (-4.866)	-0.673 *** (-4.957)	-0.709 *** (-5.039)
$S1_{t-2}$	-0.744 *** (-6.52)	-0.742 *** (-6.484)	-0.749 *** (-6.452)	-0.762 *** (-6.423)	-0.778 *** (-6.399)
r_{10}	0.098 *** (5.421)	0.098 *** (5.459)	0.098 *** (5.497)	0.097 *** (5.535)	0.097 *** (5.573)
adjusted R^2	0.320	0.323	0.325	0.327	0.328
样本数	427	427	427	427	427

表 5 用户支持倾向对市场收益率影响的回归结果

Table 5 The regression results of the effect of user support tendency on market returns

	(1)	(2)	(3)	(4)	(5)
$S1_t$		1.413 *** (8.388)	2.07 *** (10.699)	2.088 *** (10.997)	2.646 *** (12.132)
$S1_{t-1}$			-1.503 *** (-9.106)	-1.21 *** (-6.934)	-1.241 *** (-7.632)
$S1_{t-2}$				-0.67 *** (-5.216)	-0.681 *** (-5.312)
$A2_t$					2.88 (0.333)
$S2_t$					-0.998 ** (-2.243)
$A1_t$		-22.526 *** (-3.141)	-21.545 *** (-3.6)	-22.789 *** (-3.956)	-22.805 *** (-3.186)
$A1_{t-1}$		-9.163 (-1.575)	4.227 (0.802)	5.371 (1.057)	6.317 (1.309)
$A1_{t-2}$		11.115 * (1.789)	10.805 ** (1.979)	16.799 *** (3.3)	16.89 *** (3.639)
$\ln(\text{message}_t)$	-0.615 *** (-2.922)	-0.609 ** (-2.575)	-1.283 *** (-5.314)	-1.213 *** (-4.779)	-0.985 *** (-4.194)
$\ln(\text{message}_{t-1})$	0.709 *** (3.089)	1.042 *** (4.653)	1.474 *** (6.434)	1.149 *** (5.039)	0.956 *** (4.529)
$\ln(\text{message}_{t-2})$	-0.207 (-1.2)	-0.264 (-1.351)	-0.268 (-1.538)	-0.088 (-0.507)	-0.167 (-1.02)
其他控制变量	控制	控制	控制	控制	控制
adjusted R^2	0.137	0.319	0.454	0.480	0.524
样本数	427	427	427	427	427

表 5 的收益率回归结果中,支持倾向高度显著,当期符号为正,滞后期为负,即当期看涨导致当期的正收益,而滞后期的看涨将导致当期的负收益.已有的基于用户情绪的研究表明,情绪对当期收益率有正向影响,但对未来的收益率无预测能力^[26,29].本文基于用户支持倾向的研究结果与之不同,当期支持倾向不仅对当期收益率有正向影响,而且与未来收益率相反,这说明股吧用户支持倾向并不能正确预测未来的市场收益.若用户支持倾向存在“羊群效应”,则可能导致这种支持倾向与收益率方向相反的结果.本文在 4.1 节和 4.2 节中进一步讨论了用户支持倾向的“羊群效应”.在加入情绪信息作为控制变量后,回归结果依旧成立,这进一步说明用户支持倾向确实存在着不同于用户情绪的信息可以用于解释市场收益率.

3.2 支持倾向对市场波动性的影响分析

本节将讨论用户评论中的支持倾向是否会影响市场波动性.

衡量市场波动性的指标有很多,本文用振幅来代表单日市场波动,计算公式为

$$Amplitude_t = (\frac{max_price_t}{min_price_t} - 1) \times 100\% \quad (6)$$

为了刻画从支持倾向发生到市场发生反应这一滞后现象,本文再次采用了分布滞后模型,其表达式由式(7)给出.其中, $A1_{t-i}$ 为支持倾向一致性指数, $S1_{t-i}$ 为支持倾向, $A2_t$ 为情绪一致性指数, $S2_t$ 为情绪, $message_{t-i}$ 为发帖量, X 为一组控制变量,其设定与式(5)中一致, ε_t 是随机误差项.在表 6 中,模型(1)给出了只有简单控制变量的结果;模型(2)~模型(4)通过逐步加入支持倾向一致性指标及其滞后项来探究其对振幅的影响,当加入滞后 2 期的一致性指标时,指标不显著且拟合优度有所下降,说明只有当期和滞后 1 期的支持倾向一致性指标具有解释力;模型(5)加入了用户情绪作为控制变量

$$Amplitude_t = c_0 + \sum_{i=0}^m \alpha_i A1_{t-i} + \sum_{i=0}^m \beta_i S1_{t-i} + \mu_1 A2_t + \mu_2 S2_t + \sum_{i=0}^m \gamma_i \ln(message_{t-i}) + X\delta + \varepsilon_t \quad (7)$$

表 6 用户支持倾向一致性对市场波动性影响的回归结果

Table 6 The regression results of the effect of user agreement of support tendency on market volatility

	(1)	(2)	(3)	(4)	(5)
$A1_t$		8.373 *** (2.825)	6.255 ** (2.089)	6.22 ** (2.095)	4.477 (1.234)
$A1_{t-1}$			8.223 *** (2.669)	8.112 *** (2.543)	8.507 *** (2.661)
$A1_{t-2}$				0.458 (0.177)	0.368 (0.142)
$A2_t$					-0.314 *** (-2.9)
$S2_t$					-0.488 *** (-4.516)
$S1_t$		-0.314 *** (-3.295)	-0.303 *** (-3.197)	-0.304 *** (-3.188)	0.065 (0.746)
$S1_{t-1}$		-0.442 *** (-4.463)	-0.493 *** (-4.738)	-0.492 *** (-4.628)	4.817 (0.983)
$S1_{t-2}$		0.092 (1.11)	0.064 (0.802)	0.061 (0.7)	0.247 (0.903)
$\ln(message_t)$	0.103 (0.883)	-0.101 (-0.883)	-0.061 (-0.528)	-0.059 (-0.511)	-0.047 (-0.39)
$\ln(message_{t-1})$	0.091 (1.06)	0.251 ** (2.201)	0.131 (1.15)	0.132 (1.151)	0.1319 (1.146)
$\ln(message_{t-2})$	0.093 (-0.913)	-0.2* (-1.933)	-0.154 (-1.489)	-0.159 (-1.464)	-0.151 (-1.37)
其他控制变量	控制	控制	控制	控制	控制
adjusted R^2	0.574	0.64	0.647	0.646	0.645
样本数	427	427	427	427	427

从表 6 的振幅回归结果中,支持倾向一致性指数显著且符号为正,即用户的分歧越小(一致性越高),市场的波动性越大.已有的理论认为投资者的分歧将导致交易,从而产生更大的市场波动.然而,本文的样本仅仅是股吧论坛中的广大用户,大部分都是散户投资者,仅能代表在线金融社区用户群体的看涨看跌倾向,而其他投资群体如机构券商等可能有相反的看涨看跌倾向,从而促成具有不同看涨看跌倾向的群体之间的交易.如果股吧论坛中的用户支持倾向存在“羊群效应”,用户达成一致性就可能是一种羊群行为,进而导致更大的市场波动性.为此,本文在 4.1 节和 4.2 节进一步讨论了用户支持倾向的“羊群效应”.

4 进一步的检验

4.1 日间市场表现与支持倾向的关系

在讨论用户支持倾向对市场收益和波动的影响时,可能存在反向因果的问题.用户的支持倾向可能来源于历史市场表现,即历史市场表现低迷时倾向于看跌,历史市场表现较好时倾向于看涨,从而利用用户支持倾向对未来收益进行预测时产生了不符合直觉的结果.图 3 显示,无论是看涨还是看跌的支持倾向,在支持倾向发生前,市场就已经有所反应,且在支持倾向发生后市场累计收益率几乎保持不变.这说明,0 时刻的支持倾向可能更多地反映了投资者对过去一段时间内市场走势的态度,并认为这种趋势将得以延续.在收益率回归模型中,滞后期的支持倾向对紧随其后的股市收益率都具有显著的负向影响.

为了说明支持倾向更多的是投资者对过去一段时间市场行情的趋势延续,本文以支持倾向/情绪为因变量 $S1_t/S2_t$,以历史支持倾向/情绪 $S1_{t-i}/S2_{t-i}$ 、历史市场收益率 r_{t-i} 、历史支持倾向、情绪的一致性指数 $A1_{t-i}$ 和 $A2_{t-i}$ 、历史发帖数 $message_{t-i}$ 以及控制变量 X (与式(5)中一致)为自变量,进行回归检验,如式(8)和式(9)所示.在表 7 中,列(1)~列(3)展示了式(8)的回归结果;列(4)展示了式(9)的回归结果.

表 7 的回归结果表明,用户的支持倾向与过去一段时间内的市场表现具有显著的正向关系,

即当过去一段时间内的市场表现较好,用户倾向于发表看涨评论;当过去一段时间内的市场表现较低迷,用户倾向于发表看跌评论.这说明用户支持倾向信息更有可能来源于市场趋势.此外,表 7 显示支持倾向信息来源于过去多期的市场收益率,且重要程度随滞后期的增加而递减,而情绪信息仅与当期和滞后一期的市场收益率相关,且符号不稳定,这进一步说明了支持倾向信息不同于情绪信息,即支持倾向信息是一种相对更加有效的、与用户投资行为更相关的信息

$$S1_t = c_0 + \sum_{i=0}^m \beta_i r_{t-i} + \sum_{i=0}^m \beta_i S2_{t-i} + \sum_{i=0}^m \alpha_i A1_{t-i} + \sum_{i=0}^m \alpha_i A2_{t-i} + \sum_{i=0}^m \gamma_i \ln(message_{t-i}) + X\delta + \varepsilon_t \quad (8)$$

$$S2_t = c_0 + \sum_{i=0}^m \beta_i r_{t-i} + \sum_{i=0}^m \beta_i S1_{t-i} + \sum_{i=0}^m \alpha_i A1_{t-i} + \sum_{i=0}^m \alpha_i A2_{t-i} + \sum_{i=0}^m \gamma_i \ln(message_{t-i}) + X\delta + \varepsilon_t \quad (9)$$

4.2 日内市场表现与支持倾向的关系

行为金融学认为投资者群体具有“羊群效应”^[44],为了讨论从在线金融社区用户评论中提取出的支持倾向信息是否存在“羊群效应”,本节以 30min 为周期,检验日内市场表现与支持倾向的关系.

以 30min 为一个周期,将一个交易日的交易时间拆分为 8 个时段,样本数量变为了交易日数 $\times 8$.在回归方程的设定上,参考了 Zhang 和 Liu 的设定^[45],如式(10)所示.其中, $s_{i,t}$ 为第 i 个交易日第 t 个时段的支持倾向/情绪, $R_{i,t-1}$ 为第 i 个交易日第 $t-1$ 个时段的累计收益率, $S1_{i,t-1}$ 和 $S2_{i,t-1}$ 为第 i 个交易日第 $t-1$ 个时段的累计支持倾向和情绪, $A1_{i,t-1}$ 和 $A2_{i,t-1}$ 为第 i 个交易日第 $t-1$ 个时段的累计支持倾向一致性指数和情绪一致性指数, $Y_{i,t}$ 为时变控制变量,包括第 i 个交易日第 t 个时段的发帖数、交易量、振幅, Z_i 为时不变控制变量,包括第 i 个交易日的市盈率、市净率、市销率、市现率, $\varepsilon_{i,t}$ 是误差项.由于高频数据获取的限制,本文采用的时变市场数据是基于沪深 300 指数的,且时间为从 2018 年 1 月 2 日~2018 年 9 月 28 日共 183 个交易日.表 8 给出了模型的回

归结果. 其中,模型(1)~模型(3)以用户支持倾向 $s1_{i,t}$ 为因变量;模型(4)以用户情绪 $s2_{i,t}$ 为因变量

$$s1_{i,t} = c + R_{i,t-1} + \beta_0 S1_{i,t-1} + \beta_1 S2_{i,t-1} + \beta_2 A1_{i,t-1} + \beta_3 A2_{i,t-1} + \alpha Y_{i,t} + \delta Z_i + \varepsilon_{i,t} \tag{10}$$

表 7 日间收益率对支持倾向、情绪的影响的回归结果

Table 7 The regression results of the effect of daily stock returns on user support tendency and sentiment

	(1)	(2)	(3)	(4)
r_t	0.159 *** (12.733)		0.152 *** (12.969)	-0.021 *** (-2.609)
r_{t-1}	0.071 *** (5.544)		0.061 *** (4.649)	0.019 *** (3.121)
r_{t-2}	0.047 *** (3.449)		0.047 *** (3.436)	-0.002 (-0.408)
r_{t-3}	0.037 *** (3.122)		0.038 *** (3.277)	-0.001 (-0.196)
$S2_t$		0.828 *** (4.128)	0.625 *** (4.269)	
$S2_{t-1}$		0.206 ** (1.77)	0.149 (1.317)	
$S2_{t-2}$		-0.103 (-0.729)	-0.067 (-0.662)	
$S2_{t-3}$		-0.145 (-1.0)	0.015 (0.119)	
$S1_t$				0.186 *** (5.8)
$S1_{t-1}$				-0.133 *** (-5.189)
$S1_{t-2}$				0.013 (0.573)
$S1_{t-3}$				-0.026 (-1.21)
控制变量	控制	控制	控制	控制
adjusted R^2	0.755	0.674	0.77	0.934
样本数	427	427	427	427

表 8 日内收益率对支持倾向、情绪的影响的回归结果

Table 8 The regression results of the effect of intraday stock returns on user support tendency and sentiment

	(1)	(2)	(3)	(4)
$R_{i,t-1}$	6.841 *** (10.225)		5.687 *** (7.932)	1.256 ** (2.123)
$S1_{i,t-1}$		0.189 *** (7.675)	0.112 *** (4.304)	-0.084 *** (-3.944)
$S2_{i,t-1}$	0.09 *** (2.984)	-0.07 * (-1.909)	-0.00015 (-0.004)	0.161 *** (5.294)
控制变量	控制	控制	控制	控制
adjusted R^2	0.560	0.547	0.566	0.543
样本量	1464	1464	1464	1464

从日内市场表现与支持倾向、情绪的回归结果可以看出,日内支持倾向 / 情绪与日内市场累计收益率正相关,即用户在跟随市场累计收益率并发表自己的支持倾向,这说明在线金融社区用户存在“羊群效应”. 进一步,日内当期支持倾向与累计支持倾向正相关,日内当期情绪与累计情绪正相关,表明用户日内已经发表评论的支持倾向和情绪会影响后面的用户的支持倾向和情绪,这进一步加强了在线金融社区用户评论的“羊群效应”. 在线金融社区用户评论的这种“羊群效应”可能导致出现在 3.1 节和 3.2 节中部分不符合直觉的结果.

4.3 股市波动性的非对称效应

本节将使用金融时间序列分析中的 GJR(1, 1) 模型以检验第 3 节中回归结果的稳健性,同时讨论在构造支持倾向 S_t 时取不同的非对称效应调整因子 (af) 所带来的影响.

GARCH 类模型是金融时间序列分析中经典的模型簇,其中在市场非对称效应下表现最好的模型

是 GJR(1, 1) 模型^[35]. 因此,本节运用 GJR(1, 1) 模型来检验前文得出结论的稳健性.

GJR(1, 1) 模型由均值方程和方差方程两部分组成,本文参照了 Antweiler 等的方程设定^[9], 并进一步加入了控制变量. 式(11) 给出了均值方程,其中, $S_{1,t-i}$ 是支持倾向, Y 是控制变量,包括情绪和过去 10 日的累计收益率. 式(12) 给出了方差方程. 其中, $A_{1,t-i}$ 、 $A_{2,t-i}$ 和 $message_{t-i}$ 分别是支持倾向一致性指数、情绪一致性指数和发帖量, X 是控制变量,与 3.2 节中的定义一致. $I(\varepsilon_{t-1} < 0)$ 在 $\varepsilon_{t-1} < 0$ 时取 1, 否则取 0. 当存在正向杠杆时, $\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)$ 项显著,有一个负的冲击被加入模型中用以刻画非对称性. 表 9 给出了 GJR(1, 1) 模型的结果

$$r_t = c_0 + \sum_{i=0}^m \beta_i S_{1,t-i} + \gamma Y + \varepsilon_t \quad (11)$$

$$\sigma_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 \sigma_{t-1}^2 + a_3 \varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0) + \sum_{i=0}^m \alpha_i A_{1,t-i} + \sum_{i=0}^m \gamma_i \ln(message_{t-i}) + X\delta \quad (12)$$

表 9 GJR(1,1) 模型结果

Table 9 The results of GJR(1, 1) model

	$af = 1.0$	$af = 1.4$	$af = 1.8$	$af = 2.0$
$S_{1,t}$	1.915 *** (12.364)	2.028 *** (12.933)	2.44 *** (14.348)	2.558 *** (14.309)
$S_{1,t-1}$	-0.681 *** (-5.249)	-0.72 *** (-5.552)	-0.828 *** (-5.997)	-0.883 *** (-5.99)
$S_{1,t-2}$	-0.452 *** (-3.88)	-0.461 *** (-3.784)	-0.572 *** (-4.8)	-0.594 *** (-4.361)
$S_{2,t}$	-1.048 *** (-7.021)	-1.009 *** (-6.896)	-1.255 *** (-8.593)	-1.261 *** (-8.54)
r_{-10}	0.076 *** (4.972)	0.073 *** (4.897)	0.077 *** (5.361)	0.074 *** (5.104)
ε_{t-1}^2	-0.097 *** (-2.95)	-0.092 *** (-2.628)	-0.076 ** (-2.194)	-0.069 ** (-1.988)
$\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)$	0.113 * (1.656)	0.147 * (1.851)	0.129 (1.612)	0.129 (1.59)
σ_{t-1}^2	0.503 *** (3.183)	0.529 *** (3.504)	0.483 *** (2.726)	0.524 *** (2.703)
控制变量	控制	控制	控制	控制
adjusted R^2	0.395	0.406	0.424	0.429
样本数	427	427	427	427

从表 9 可以看出,支持倾向和情绪的符号方向均与第 3 节的回归结果保持一致. 当 $af = 1.0$

时, $\varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)$ 项显著,说明确实存在正向杠杆效应,于是通过逐渐增大 af 来调整正向杠杆效

应,当 af 增大到1.8时, $\varepsilon_{i-1}^2 I(\varepsilon_{i-1} < 0)$ 项不再显著,此时的市场非对称性已通过调整 af 来解释.这说明通过调整 af 确实能够更好地刻画市场的非对称效应.

5 结束语

本文利用卷积神经网络分类算法提取并测度了在线金融社区中用户评论所体现的支持倾向,能够达到74.8%的精度和74.7%的召回率,优于现有文献的分类算法.并且考察了支持倾向对沪深300指数成分股的收益和波动的影响.

在对市场收益的影响方面,用户当期的支持倾向对当期的市场收益率具有正向影响,但前1、2日的支持倾向对当日收益率具有负向影响,这一结果在控制用户情绪的条件下依旧显著.进一步,通过考察日间股市表现对用户支持倾向的影响,发现在羊群行为作用下,过去10日内的市场表现是用户形成支持倾向的重要原因,当市场历

史表现好的情况下,用户会形成看涨的支持倾向,反之形成看跌的支持倾向.在日内收益率与支持倾向的影响检验中,本文发现不仅日内的累计收益率会对用户支持倾向产生正向影响,而且用户已发表的评论中的累计支持倾向也会对后续用户的支持倾向产生正向影响,这进一步加强了在线金融社区用户追涨杀跌的“羊群效应”.

在对市场波动的影响方面,用户支持倾向的一致性对市场波动具有正向影响,即用户的分歧越小,市场的波动性越大.通过描述性统计和回归检验表明了股市中存在正向杠杆效应.通过引入的市场非对称效应的调整因子 af (asymmetric factor),本文很好地捕捉并刻画了市场的正向杠杆效应.

未来的研究一方面可以基于本文提出的在线金融社区用户的支持倾向测度方法,将本文对市场收益和波动的影响研究扩展至个股或股票组合层面.另一方面,针对在线金融社区投资者群体与机构、券商等其他投资群体支持倾向及一致性的差异研究,也值得进一步思考.

参考文献:

- [1] Fama E F. The behavior of stock-market prices [J]. Journal of Business, 1965, 38(1): 34-105.
- [2] Li F. Annual report readability, current earnings, and earnings persistence [J]. Journal of Accounting & Economics, 2008, 45(2): 221-247.
- [3] 薛健, 窦超. 并购重组过程中的信息泄露与财富转移 [J]. 金融研究, 2015, (6): 189-206.
Xue Jian, Dou Chao. Information leakage and wealth transfer effect in merge & acquisition [J]. Journal of Financial Research, 2015, (6): 189-206. (in Chinese)
- [4] 林翔. 对中国证券咨询机构预测的分析 [J]. 经济研究, 2000, (2): 56-65.
Lin Xiang. Financial derivatives and the development of capital market in China [J]. Economic Research Journal, 2000, (2): 56-65. (in Chinese)
- [5] Jagadish H V. Big data and science: Myths and reality [J]. Big Data Research, 2015, 2(2): 49-52.
- [6] Aarkstore Enterprise. Social media in financial services: The customer as the advisor [EB/OL]. <http://articles.pubarticles.com/aarkstore-enterprisesocial-media-in-financial-services-the-customer-as-the-advisor-1312199273,272759.html>, 2011.
- [7] Cogent Research. "Social Media's Impact on Personal Finance & Investing." [EB/OL]. <http://www.cogentresearch.com/>: Cogent Research, 2008.
- [8] John R N. Social mood and financial economics [J]. Journal of Behavioral Finance, 2005, 6(3): 144-160.
- [9] Antweiler W, Frank M Z. Is all that talk just noise? The information content of internet stock message boards [J]. Journal of Finance, 2004, 59(3): 1259-1294.
- [10] Chen H, De P, Hu Y, et al. Wisdom of crowds: The value of stock opinions transmitted through social media [J]. Review of Financial Studies, 2013, 27(5): 1367-1403.
- [11] Boehme R D, Danielsen B R, Sorescu S M. Short-sale constraints, differences of opinion, and overvaluation [J]. Journal of Financial & Quantitative Analysis, 2006, 41(2): 455-487.
- [12] Frieder L, Zittrain J L. Spam works: Evidence from stock touts and corresponding market activity [J]. Hastings Communi-

- cations and Entertainment Law Journal, 2008, 30(3): 480–520.
- [13] Wurgler J A, Baker M P. Investor sentiment and the cross-section of stock returns [J]. Journal of Finance, 2006, 61(4): 1645–1680.
- [14] Berger D, Turtle H J. Cross-sectional performance and investor sentiment in a multiple risk factor model [J]. Journal of Banking & Finance, 2012, 36(4): 1107–1121.
- [15] Tetlock P C, Saar-Tsechansky M, Macskassy S. More than words: Quantifying language to measure firms' fundamentals [J]. Journal of Finance, 2008, 63(3): 1437–1467.
- [16] Cecchini M, Aytug H, Koehler G J, et al. Making words work: Using financial text as a predictor of financial events [J]. Decision Support Systems, 2011, 50(1): 164–175.
- [17] Tumarkin R, Whitelaw Robert F. News or noise? Internet postings and stock prices [J]. Financial Analysts Journal, 2001, 57(3): 41–51.
- [18] Das S R, Chen M Y. Yahoo! for Amazon: Sentiment extraction from small talk on the Web [J]. Management Science, 2007, 53(9): 1375–1388.
- [19] Renault T. Intraday online investor sentiment and return patterns in the US stock market [J]. Journal of Banking & Finance, 2017, 84: 25–40.
- [20] Eli B, Lucile F, Partha S. Can Twitter help predict firm-level earnings and stock returns? [J]. The Accounting Review, 2018, 93(3): 25–57.
- [21] 刘海飞, 许金涛, 柏巍, 等. 社交网络、投资者关注与股价同步性 [J]. 管理科学学报, 2017, 20(2): 53–62.
Liu Haifei, Xu Jintao, Bai Wei, et al. Social networks, investor attention and stock price synchronicity [J]. Journal of Management Sciences in China, 2017, 20(2): 53–62. (in Chinese)
- [22] 丁慧, 吕长江, 陈运佳. 投资者信息能力: 意见分歧与股价崩盘风险——来自社交媒体“上证 e 互动”的证据 [J]. 管理世界, 2018, 34(9): 167–177.
Ding Hui, Lü Changjiang, Chen Yunjia. Investor's information ability: Divergence of opinion and risk of stock market crash: Evidence from social media Shanghai “e” interaction [J]. Management World, 2018, 34(9): 167–177. (in Chinese)
- [23] 俞庆进, 张兵. 投资者有限关注与股票收益——以百度指数作为关注度的一项实证研究 [J]. 金融研究, 2012, (8): 152–165.
Yu Qingjin, Zhang Bing. Limited investor concern and stock return: An empirical study with Baidu index as concern degree [J]. Journal of Financial Research, 2012, (8): 152–165. (in Chinese)
- [24] 何贤杰, 王孝钰, 孙淑伟, 等. 网络新媒体信息披露的经济后果研究——基于股价同步性的视角 [J]. 管理科学学报, 2018, 21(6): 48–64.
He Xianjie, Wang Xiaoyu, Sun Shuwei, et al. Economic consequences of new media information disclosure: From the perspective of stock price synchronicity [J]. Journal of Management Sciences in China, 2018, 21(6): 48–64. (in Chinese)
- [25] 孙书娜, 孙谦. 投资者关注和股市表现——基于雪球关注度的研究 [J]. 管理科学学报, 2018, 21(6): 65–76.
Sun Shuna, Sun Qian. Investor attention and market performance: Evidence based on “Xueqiu attention” [J]. Journal of Management Sciences in China, 2018, 21(6): 65–76. (in Chinese)
- [26] 部慧, 解峥, 李佳鸿, 等. 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): 91–106.
Bu Hui, Xie Zheng, Li Jiahong, et al. Investor sentiment extracted from internet stock message boards and its effect on Chinese stock market [J]. Journal of Management Sciences in China, 2018, 21(4): 91–106. (in Chinese)
- [27] 艾瑞咨询, 2017 年 6 月 PC 网站榜单 top1000 [EB/OL]. <http://www.iresearch.cn/>: 艾瑞网, 2017.
- [28] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space [C]. Proceeding of the International Conference on Learning Representations, Arizona, USA, 2013: 1301–3781.
- [29] Brown G W, Cliff M T. Investor sentiment and the near-term stock market [J]. Journal of Empirical Finance, 2004, 11(1): 1–27.
- [30] Goodfellow I, Bengio Y, Courville A, et al. Deep Learning [M]. Cambridge: MIT Press, 2016.
- [31] Kim Y. Convolutional Neural Networks for Sentence Classification [C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha: Qatar, 2014: 1746–1751.
- [32] Tan P N, Steinbach M, Kumar V. 数据挖掘导论 [M]. 北京: 人民邮电出版社, 2011.

- Tan P N, Steinbach M, Kumar V. Introduction to Data Mining [M]. Beijing: Posts & Telecom Press, 2011. (in Chinese)
- [33] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- Zhou Zhihua. Machine Learning [M]. Beijing: Tsinghua University Press, 2016. (in Chinese)
- [34] Andersen T G, Bollerslev T, Diebold F X, et al. The distribution of realized stock return volatility [J]. Journal of Financial Economics, 2001, 61(1): 43–76.
- [35] Engle R F, Ng V K. Measuring and testing the impact of news on volatility [J]. Journal of Finance, 1993, 48(5): 1749–1778.
- [36] Schwert G W. Why does stock market volatility change over time? [J]. Journal of Finance, 1989, 44(5): 1115–1153.
- [37] Glosten L R, Jagannathan R, Runkle D E. On the relation between the expected value and the volatility of the nominal excess return on stocks [J]. Journal of Finance, 2012, 48(5): 1779–1801.
- [38] Tetlock P C. Giving content to investor sentiment: The role of media in the stock market [J]. Journal of Finance, 2007, 62(3): 1139–1168.
- [39] Karpoff J M. A theory of trading volume [J]. Journal of Finance, 1986, 41(5): 1069–1087.
- [40] Harris M, Raviv A. Differences of opinion make a horse race [J]. Review of Financial Studies, 1993, 6(3): 473–506.
- [41] 刘庆富, 周程远. 中国股票市场的非对称效应研究 [J]. 系统工程学报, 2012, 27(5): 648–655.
- Liu Qingfu, Zhou Chengyuan. Study on asymmetric effect of Chinese stock markets [J]. Journal of Systems Engineering, 2012, 27(5): 648–655. (in Chinese)
- [42] Jones C M, Kaul G, Lipson M L. Transactions, volume, and volatility [J]. Review of Financial Studies, 1994, 7(4): 631–651.
- [43] 许红伟, 陈欣. 我国推出融资融券交易促进了标的股票的定价效率吗? ——基于双重差分模型的实证研究 [J]. 管理世界, 2012, (5): 52–61.
- Xu Hongwei, Chen Xin. Does Chinese introduction of margin trading promote the pricing efficiency of underlying stocks?: An empirical study based on the double difference model [J]. Management World, 2012, (5): 52–61. (in Chinese)
- [44] Shiller R J. Irrational Exuberance: Revised and Expanded Third Edition [M]. Princeton: Princeton University Press, 2015, pp. 149–153.
- [45] Zhang J, Liu P. Rational herding in microloan markets [J]. Management Science, 2012, 58(5): 892–912.

Impact of online community support tendencies on returns and volatility in Chinese stock market

QIAN Yu, LI Zi-rao, LI Qiang, YUAN Hua*

School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: The rapid development of Internet has enabled non-professional individual investors to share information and express their tendency through online financial community. Using 5 178 824 user comments on eastmoney.com, this paper uses the convolutional neural network (CNN) classification algorithm to extract and measure the support tendency of online users for the future market (i.e. bullish or bearish) and the impact of user support tendency on the market is tested from two aspects of market returns and volatility. The results show that the current user's support tendency has a significant negative impact on future stock market returns, and the consistency of support tendency will amplify the market's volatility. Furthermore, the user support tendency is significantly based on the historical performance of the stock market, and the support tendency has a certain "herd effect".

Key words: text big data; online community; user support tendency; market returns; market volatility