

# 投资者高频情绪对股票日内收益率的预测作用

尹海员, 吴兴颖

**[摘要]** 利用数据挖掘手段从网络平台信息提取投资者情绪不仅增加了高频情绪数据的可得性,也有助于深入分析情绪与股票市场运行的互动关系。本文抓取上证指数股吧的实时发帖,通过文本语义分析构建了投资者日内高频情绪指标,并研究了其对股市盘中收益的预测效应。研究发现,中国股票市场的日内投资者情绪能正向预测股票市场运行,这种预测作用下午交易时段表现得更显著;尽管投资者情绪的预测作用独立于收益率自身的盘中动量效应,但在显著性程度上较前期收益率和波动水平要弱;牛市中投资者情绪对日内收益率的预测作用强于滞后收益率等变量,熊市则相反,但在暴涨或暴跌的极端市场环境中,情绪对日内收益率的影响程度相较于滞后收益率等变量更为显著;隔夜投资者情绪的释放会显著影响次日上午的市场收益率,但存在时滞性;午间休市期间的投资者情绪会与上午收益率一起正向影响下午的市场表现;进一步看,噪音交易是投资者情绪影响股票收益率的重要驱动力量。在考虑了月份效应、星期效应以及宏观经济变量的影响后结果仍然稳健。这些结论有助于从更高频率视角深入理解股市中情绪效应的特征及机理。

**[关键词]** 高频情绪; 股票收益率; 日内效应; 文本挖掘

**[中图分类号]**JF830 **[文献标识码]**A **[文章编号]**1006-480X(2019)08-0080-19

DOI:10.19581/j.cnki.ciejournal.2019.08.005

## 一、引言

党的十九大报告提出要建设网络强国、数字中国和智慧社会。在信息化时代发展的新阶段,互联网是基础设施,大数据是生产资料,智能化是转化成果。推动大数据、互联网与资本市场的深度融合,不仅对加快发展数字经济、提升经济社会智能化水平有重要意义,也可以更有效地对资本市场进行监管,坚守住不发生系统性金融风险的底线。从信息化时代的特点看,网络平台具备了信息创造的即时性、信息内容的多元性、信息传播的交互性等优势。网络媒体可以不断地通过反馈机制和注意力级联(Cascade)来塑造公众舆论,其对资本市场运行不仅发挥着“信息功能”,也发挥着“情绪功能”,那些有影响力的网络评述会影响投资者的情绪波动,最终对资产定价产生冲击。

自从 Keynes(1936)在解释经济运行的异常波动时提到“动物精神”以来,投资者情绪在资本市

**[收稿日期]** 2018-10-12

**[基金项目]** 教育部人文社会科学研究一般项目“股票流动性对投资者情绪波动的响应机制研究”(批准号16YJA790061);中央高校基本科研业务费专项项目“投资者情绪、股票流动性与资产配置效应”(批准号GK201803091)。

**[作者简介]** 尹海员,陕西师范大学国际商学院副教授,经济学博士;吴兴颖,陕西师范大学国际商学院硕士研究生。通讯作者:尹海员,电子邮箱:yinhaiyuan@snnu.edu.cn。感谢匿名评审专家和编辑部的宝贵意见,当然文责自负。

场运行中的作用就受到了广泛关注。近年来关于投资者情绪与股票市场互动关系的研究层出不穷,研究重点不再仅局限于判断情绪是否能预测收益水平,而是如何衡量投资者情绪并量化其更广泛的影响效应。现有研究主要有三种方法来衡量投资者情绪:基于调查问卷得到的直接指数、基于市场交易信息得到的间接指数、Baker and Wurgler(2006)通过对主客观基础指标进行主成分分析并剔除宏观经济变量影响后得到的复合情绪指数(以下简称 BW 复合指数)。其中,直接指数的成本较高,且易受主观因素以及时间频率的限制;间接指数虽更具客观性,但由于指标的单一性,难免有以偏概全之嫌;复合指数综合考虑了多种市场变量的影响,但依赖市场数据构建的指标很可能是情绪之外的许多经济力量的均衡结果,而且在选取基础指标的过程中,极有可能遗漏关键变量进而影响度量的全面性。

区别于传统的投资者情绪指标,依托海量在线信息来构建的情绪指标具有显著优势:首先,相对于大规模人工调查数据,通过数据挖掘手段获取网络文本信息更快速准确、更具成本和效率优势;其次,开放性的社交平台真实还原了投资者的思想和意见,使信息传递摆脱了时间、地域等约束,并能随时更新意见看法,这为高频情绪数据的产生搭建了一个较为理想的“实验环境”(黄燕芬和张超,2018)。正如 Da et al.(2015)所指出:“迄今为止,投资者情绪的高频数据只能在实验室环境中找到。”相比之下,许多基于调查的数据都是月度或季度频率,被现有研究广泛使用的 BW 复合指数也仅为月度频率。

在股票市场收益的预测方面,学者们已经基于月度、季度等低频数据展开了大量的研究。然而,多元化的信息会连续影响股价波动,数据频率越低,则过滤掉的市场信息越多,最终导致所构建的度量指标的有效性减弱。相对而言,高频数据会对丰富的日内波动性特征的信息迅速做出反应,实时揭示资产价格变化的动态特征。国内外已有学者基于分钟、小时等高频交易数据对股票市场收益率和波动性等变量的日内特征做出了探索性研究(Rahman et al.,2002; Renault,2017; Gao et al.,2018)。但中国股票市场的交易时间、竞价方式等特征与国外资本市场并不相同,这些差异化的特征可能导致其具有完全不同的盘中情绪效应模式,有必要利用新的思路探索中国股市的盘中情绪效应。

本文利用数据挖掘手段获取股票社交平台上的文本数据,通过情感分析算法构建日内高频投资者情绪指标,并研究高频投资者情绪对股票盘中收益率的预测作用,并且经过多维度验证后的实证结果仍然稳健。本文可能的创新之处主要体现在以下三个方面:

(1)直接对获取到的文本信息进行情感分析,构建了能更准确反映投资者情绪的日内高频指标。至少有两个理由促使本文构建日内情绪指标:从理论角度看,若投资者情绪是持续变化的,那么日内高频指标可以更精确地估算投资者情绪的真实波动情况;从实践角度看,投资者可能会将这种日内情绪效应纳入其预测模型并改善其交易策略。现有研究中广泛使用的网络搜索量、百度搜索指数等指标在很大程度上反映了投资主体的活跃度,但难以从语义角度判断情绪倾向。通过文本分析量化情绪,不仅可以丰富构建情绪指标的数据来源,也提高了情绪指标的解释能力。

(2)从多个维度验证了高频投资者情绪对中国股票日内收益率的可预测性,并针对中国股市的特殊交易机制,分析了隔夜信息以及午休期间的投资者情绪是否会对后期股价产生冲击。不同于美国等代表性的国外市场的日内连续交易,中国股市中的隔夜信息和午间休市为投资者获取并消化新的市场信息创造了条件。鉴于此,本文综合考虑了全样本和不同市场状态下情绪效应的差异,分析了非交易时间段内的投资者情绪的预测作用,进而分别从“月度效应”和“星期效应”两个层面验证了结果的稳健性。

(3)分别从交易量和市场长期收益反转的角度探讨了噪音交易者行为对情绪预测作用的影响。

相较于机构投资者,散户对股市信息的甄别和预判能力都处于绝对劣势,在信息的获取过程中往往奉行拿来主义,以噪音交易者的角色在股市涨跌过程中推波助澜,而社交媒体可以进一步促进不实的信息在市场中的快速传播,加剧市场波动。鉴于噪音交易往往会带来交易量上升和长期收益的反转,本文从实证角度验证了噪音交易行为在情绪预测市场收益过程中发挥的作用。

## 二、文献评述与研究思路

### 1. 投资者情绪度量

现有研究主要从直接指标、间接指标和复合指标三种思路试图准确衡量投资者情绪。直接指标方面,王美今和孙建军(2004)利用央视看盘指数构建了投资者情绪指数,发现它是一个影响股票均衡价格的系统性因子。间接指标方面,Derrien(2005)利用IPO发行量及首日收益衡量投资者情绪,发现该指标可反映投资者热情程度,且与情绪正相关。复合指标方面,Baker and Wurgler(2006)将多个单一指标(包括封闭式基金折价,IPO首日收益率,首次公开发行量和交易量)进行主成分分析,在控制宏观经济变量等理性因素后得到度量投资者情绪的复合指数。

近年来,基于对网络媒体内容进行文本分析的情感指标已逐渐被用到行为金融研究中。Wysocki(1998)最早开始研究网络论坛信息与股票市场的相关性,选取美国纽交所上市的50家样本公司,发现网络论坛上相关公司的讨论发帖量能预测次日该股票的收益状况。Antweiler and Frank(2004)从雅虎财经挖掘到150万条网络发帖,构建了情绪指标,发现从社交媒体内容中提取的情绪指标可显著正向预测股价收益。Bollen and Mao(2011)比较分析了传统投资者情绪度量指标和社交媒体预测证券市场的能力,发现从社交媒体中提取出来的情绪指标能更好地预测市场波动。Chen et al.(2014)发现社交网站中的公司信息可以显著预测股票回报,即使在控制了金融分析师和报纸文章等传统信息来源的影响后,这种预测性依然存在。Da et al.(2015)根据谷歌搜索结果,选择搜索关键词构建情绪指数,发现该指数可以预测短期回报逆转和波动性增加现象。You et al.(2017)利用Twitter内容构建了投资者情绪变量并对全球代表性股票市场进行了实证检验,结果发现情绪指标对10个国际股票市场收益有预测能力。

国内学者近几年在这方面也做了不少探索性研究,林振兴(2011)收集东方财富网股吧的投资者发帖信息,从乐观情绪、意见分歧和关注度这三个方面构建指标,发现投资者关注度与IPO溢价率有较强的相关性,而乐观情绪和意见分歧与IPO溢价率无显著相关性。孟雪井等(2016)利用百度搜索指数数据衡量投资者情绪,发现中国股票市场的投资者情绪与市场指数之间存在联动机制。孟庆斌等(2017)对上市公司年报中的管理层讨论与分析信息进行文本向量化,研究发现管理层讨论与分析中的信息含量与股价崩盘风险负相关。段江娇等(2017)以上证A股公司的股吧发帖信息为研究对象,利用文本分析提取帖子情绪,发现论坛情绪会显著正向影响股票当日收益率,论坛发帖数会显著正向影响当日股价波动,显著负向影响股票未来两日收益率。石善冲等(2018)从股市类微信公众号的文章中提取投资者情绪,利用格兰杰因果检验研究不同类别的投资者情绪与上证综指收盘价和成交量之间的互动关系,发现乐观和中性情绪会影响成交量,而消极情绪主要影响收盘价和收益率。伊志宏等(2019)通过机器学习对上市公司分析师研究报告中的公司特质信息进行了文本分析,发现其中的信息含量越高,该公司的股价同步性越低。

### 2. 股票收益率的日内效应

一直以来,国内外学者围绕股市的预测做了许多尝试,但尚未发现准确预测股价变动的方法。Jegadeesh and Titman(1993)的开创性研究表明,股票收益率有延续前期运动方向的趋势。此后国内



外学者围绕股票收益率的“动量效应”展开了大量研究,发现资产的前期回报可以正向预测未来回报。比如,Cooper et al.(2006)研究了美国 1940—2003 年间 1 月份市场收益率回报的预测能力,发现 1 月的回报对未来 11 个月的市场回报具有预测能力,并且这种预测能力在不同类型的股票中普遍存在。Asness et al.(2013)的研究表明,时间序列的动量效应在股票、债券等资产类别中普遍存在。但上述研究依据的收益率时间序列大多都是月度或者季度频率数据。

在利用日内数据进行股票收益率规律研究方面,Narayan and Sharma(2016)利用日内数据进行实证研究,发现标准普尔 500 指数期货收益率可以在某个特定交易频率成功预测中国股票收益率。Gao et al.(2018)研究了标准普尔 500 指数的日内动量效应现象,发现市场首个半小时的回报可以预测市场上最后一个半小时的回报。借鉴这一思路,Zhang et al.(2018)利用半小时交易数据研究了中国股市的日内动量效应,发现第一个和第七个半小时的收益可以显著预测最后半小时的收益,并且早上的市场收益也能显著预测下午的市场收益。

投资者情绪也在股票收益的预测中扮演着重要角色。从理论角度,DeLong et al.(1990)的研究表明,在限制套利的情况下,噪音交易者情绪的变化会导致市场过度波动,使得股价偏离其基本价值。从实证角度,Hung(2012)、陆静和周媛(2015)分别基于不同的实证模型研究了投资者情绪的总效应,并从不同维度验证了投资者情绪对股票收益率的预测作用。王春(2014)研究了情绪的横截面效应,发现投资者情绪对股票市场的价值溢价具有预测能力,但这种溢出效应在不同的公司规模、行业种类中存在不同程度的差异。

也有学者从高频数据的视角研究了投资者与市场运行的互动关系。Antoniou and Doukas(2013)的研究表明,只有当投资者情绪乐观时,市场才会产生动量收益。Sun et al.(2016)发现标准普尔 500 指数的日内收益可利用投资者情绪进行预测,从投资角度来看,投资者情绪在市场交易策略中具有显著作用。刘善存和许敏(2007)以上证 50 指数为研究对象,以 30 分钟作为分割点,分别用收益的波动以及波动的差分值衡量投资者所面临的机会风险和逆向选择风险,以成交量变化衡量投资者的风险态度,结果发现投资者对股票收益波动以及波动的变化量均有一定程度的敏感性。谢军等(2012)以沪深 300 股指期货为研究对象,利用多空不均衡指标作为投资者情绪的代理变量,研究高频环境下情绪对股指期货的冲击效应,发现在总体上投资者情绪与股指期货收益显著正相关,在日内效应上情绪对股指期货的影响效应呈“ $\hookrightarrow$ ”型。

### 3. 研究思路与框架

综合分析,已有研究开始基于网络数据度量投资者情绪状态,并对未来经济状况、股票市场运行进行预测。然而在对网络数据有效信息的提取上,多数学者关注的是能对证券市场活跃度产生影响的数量指标,比如关键词搜索量、百度搜索指数、股吧发帖数等。数量指标固然在反应投资者对股票收益的关注程度上有其可行性,但是难以衡量投资者情绪倾向的深层次信息,而目前从文本语义的角度研究投资者情绪效应仍处于探索阶段。

进一步地,尽管投资者情绪在股票市场中的重要性已得到广泛认可,但鉴于高频投资者情绪的合理测度受到数据收集和量化的限制,在股票日内收益的预测上,多数学者仅考虑收益率本身的动量效应,或者利用可得性较高的交易数据从侧面度量投资者情绪。随着大数据挖掘技术的发展,如何利用网络信息传播路径来量化分析股票市场的日内情绪效应,值得进一步研究。本文利用数据挖掘手段获取主流网站股票论坛中的发帖文本信息,通过 Python 情感分析算法构建了上证指数的日内高频投资者情绪指标,从多个维度研究了中国股市中投资者情绪对股票日内收益的预测效应,并进一步探讨噪音交易在其中的驱动作用。具体研究框架如图 1 所示。

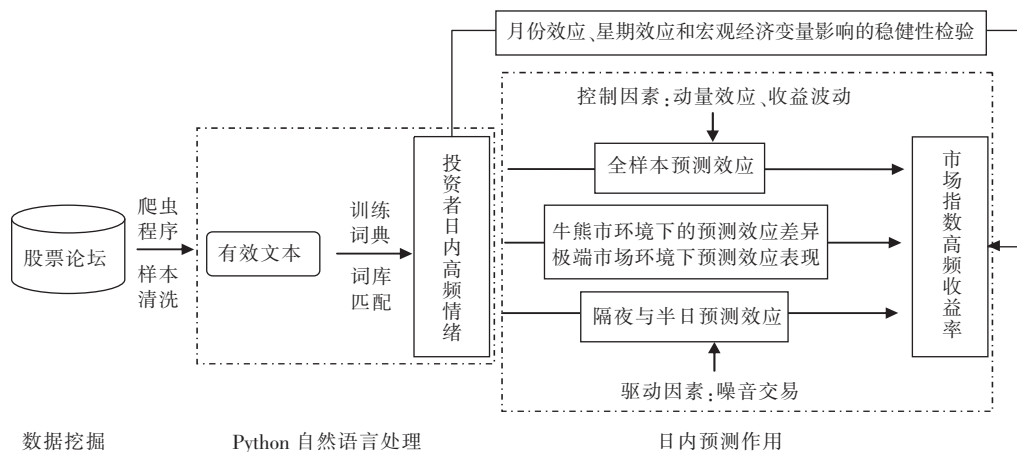


图1 本文研究思路框架

### 三、数据处理与研究设计

#### 1. 数据来源与处理

在利用网络媒体信息度量投资者情绪的过程中,选取有代表性的网络数据样本尤为必要。中国股票市场中个体交易者规模庞大而且交易更为活跃,规模方面,截至2017年底中国A股市场个体投资者占比超过99%;交易量方面,沪市个体投资者在2017年全年贡献的交易额是机构投资者的5倍<sup>①</sup>。但是无论是获取信息的渠道、预判能力,还是信息甄别能力,个体投资者相对机构投资者都处于绝对劣势,这促使前者在信息获取过程中往往奉行拿来主义。在众多网络平台上,股吧是众多中小投资者交流经验、传播信息、获得其他投资者看法的重要平台。

经过对多个股票论坛的比较,本文选取国内主流财经门户网站——“金融界”的股票论坛帖为研究对象。金融界网站主要优点在于股票论坛中的资讯丰富,发帖频率高,同时股吧论坛的专题分类页齐全,保留的发帖相较其他论坛有更长的历史追溯性,为日内数据的挖掘提供了便利。本文通过爬虫软件抓取了金融界论坛2013年7月1日至2018年8月14日的历史发帖信息,通过情感分析构建了上证指数的日内投资者情绪指数。具体的数据处理流程如下:

(1)原始数据抓取:①在Java环境的支持下调用R中的Rselenium和Rwebdriver等程序包实现动态网页爬虫。逐一打开股票论坛“全部贴”分类页中的历史发帖信息,在详情页中抓取每条发帖的标题、作者、阅读量、详细时间等字段。在数据抓取过程中忽略无效链接并剔除图片、广告贴,最终初步导出文本信息486415条。②对原始文本内容进行数据清洗,去除无效符号和冗余字段、删除重复发帖、剔除空缺行等。初步过滤后,仍存在少量与股市无关的发帖,在Python语言中导入jieba,对所有文本内容进行分词,再导入Collections模块的Counter类,跟踪统计各单词或者双词出现的次数,在剔除停用词如“的”“了”之后得到排名前1000的关键词。③人工筛选出与股市无关的关键词,以此为依据查找并剔除无效发帖80281条,得到有效帖406134条,论坛的日均发帖数量大约为160条。以半小时为间隔,按照不同时间区间对所有帖子分类汇总,将发帖日期与股市行情数据的日期进行匹配,得到交易时间段内的发帖120665条,非交易时间段的发帖285469条。

<sup>①</sup> 投资者规模数据来自于中国证券登记结算有限公司发布的《中国证券登记结算统计年鉴》(2017);投资者交易量数据来自于上海证券交易所发布的《上海证券交易所统计年鉴》(2018卷)。

(2)投资者情绪的量化。利用 Python 中的类库 SnowNLP 来量化论坛帖中蕴含的情绪倾向,具体步骤如下:

首先,构建量化分析所需的词典语料。将 SnowNLP 中自带的 pos(neg)语料库、中国知网(HowNet)情感词典和台湾大学简体中文情感极性词典作为基础情感词典语料库。考虑到不同行业或领域中的情感表达在用词上风格各异,为提高情感识别的准确度,再选取 BosonNLP 情感词典<sup>①</sup>作为领域情感词典语料库。将基础语料库、领域语料库中的正(负)面情感词语词典分别消重后进行整合,获得训练数据。对这些数据集中加以训练,并将训练结果单独存储。

其次,调用 SnowNLP 中的 sentiment 模块并编写循环语句,基于词库匹配法对所有样本数据逐条进行量化处理,返回情感为“正面”的概率,取值区间为[0,1]。考虑到部分文本内容中没有情感词汇,这会导致少量分类结果中含有情绪值为 0.5 的中性类别的量化结果。针对中性类别的内容,利用 Python 编写程序,安装 jieba 并在 Python 中导入 NLTK 资料库,对这些内容进行分词后,跟踪统计排名前 1000 的单词和双词,把单词和双词一起作为特征,重新构建训练需要的数据格式,为积极文本赋予“pos1”,为消极文本赋予“neg1”,形成一个新字典并再次利用 SnowNLP 将这些原始信息重新进行划分。本文最终获得 288163 乐观文本、117971 个悲观文本。

最后,将所有交易日的帖子按照半小时划分后,分别在每个半小时中抽取 50 条,人工对这些帖子进行情感类别划分并标注。将人工标注的类别与利用 SnowNLP 划分的结果进行比对发现分类的准确率为 83.2%,情绪分类总体较为理想。

一个容易引起争议的问题是,股吧里的发帖有没有可能是某些投资者为了故意诱导其他投资者而表达的假观点,以及大量情绪宣泄类的跟帖?<sup>②</sup>本文认为,这些不足以影响研究的科学性,主要原因有三个:①对多数注册用户来说,通过发表对股票的看法,如果能预测成功将极大提升其论坛影响力并吸引追随者,所以发表符合其真实意思的看法符合发帖者自身利益;②论坛中有数量众多的发帖者和追随者,并且没有形成微博、博客中的大 V 认证的“权威人士”,单个发帖者的看法很难影响众人的看法和判断;③即使大量情绪宣泄贴存在,或者一些发帖者并未持有相关股票而发帖,这仍然反应了发帖者的看法,阅读浏览者会根据其对相关股票的看法来选择相应的操作策略。但为了保证科学性和真实性,本文在 Python 文本分析程序设计时,降低了对部分极值词汇的权重赋值,以便过滤和减弱极端情绪宣泄对情绪度量的影响。

(3)市场运行数据。后续实证中涉及的市场收益率、交易量、宏观经济变量数据均来源于万德(Wind)金融数据库。同时为确保数据的准确性,对这些市场交易数据又通过锐思(RESSET)金融数据库、国家统计局网站进行了分批逐一核实。

## 2. 变量指标设计

(1)投资者高频情绪。本文以半小时为间隔将中国股市每日的交易时间分为 8 个 30 分钟时间段,分别汇总各半小时内的乐观/悲观帖子数,以此构建高频投资者情绪指数。具体地,将量化后的情绪值介于[0,0.5)的视为悲观帖子,情绪值介于(0.5,1]的视为乐观帖子,则第  $t$  个半小时投资者情绪为:

$$sentiment_{i,t} = \ln[(1 + pos_{i,t}) / (1 + neg_{i,t})] \quad (1)$$

公式(1)中的  $pos_{i,t}$  代表第  $i$  日第  $t$  半小时的乐观帖子数,  $neg_{i,t}$  代表悲观帖子数。可以发现,当乐观帖与悲观帖子数相等时,投资者情绪保持中立;乐观帖子数大于悲观帖子数时,投资者情绪值大

① BosonNLP 情感词典是基于微博、新闻和网络论坛等数据来源构建得到的情感极性词典,囊括了大量网络词汇以及非正式简称,对非规范语言表述也有较高的覆盖率,比较适用于网络社交媒体的文本分析。

② 感谢匿名审稿专家对这一问题提供的启发。

于0,反之小于0。

本文投资者情绪指数是通过数据挖掘和文本分析构建的高频指标,为验证其作为情绪衡量指标的有效性,分别选取封闭式基金折价率、BW复合指数作为传统的情绪代理变量  $proxy_t$ ,利用公式(2)分别进行回归。为了与传统指标的数据频率保持一致,预先对  $sentiment$  数据集进行月度汇总。回归结果表明<sup>①</sup>,两种传统的情绪代理变量都与  $sentiment$  正相关,其中与 BW 复合指数的一致性程度更好,在 10% 的水平下显著,且拟合度略高于封闭式基金折价率,这说明本文构建的投资者情绪变量  $sentiment$  是可行的。

$$sentiment_{i,t} = \beta_0 + \beta_1 proxy_t + \varepsilon_t \quad (2)$$

(2)股票收益率。为与情绪指标的时间相匹配,将半小时收盘价取对数差分后乘以 100,得到上证指数收益率:

$$R_{i,t} = 100 \times [\ln(P_{i,t}) - \ln(P_{i,t-1})] \quad (3)$$

公式(3)中,  $P_{i,t}$  表示第  $i$  日第  $t$  半小时的收盘价。在计算各交易日第一个半小时的收益率时,使用前一个交易日的收盘价作为起始价格  $P_{i,t-1}$ 。基于此,前半个小时的收益就会捕捉到在前一天的市场收盘后发布的信息。第五个半小时的回报是基于 11:30 和 13:30 的收盘价计算的。由于中国股市在 11:30—13:00 之间有 90 分钟的午休时间,每个交易日共获得 8 个交易日时段观测值。

## 四、实证分析

### 1. 描述性统计

图 2 和图 3 分别展示了市场收益率、论坛发帖数量在不同时间段的基本特征,柱状图中各变量取值都是横坐标对应时间段在 2013 年 7 月至 2018 年 8 月的加权平均值,其中,图 2 中各个半小时内的收益率波动通过 5 分钟频率的收益率标准差衡量。从图 2 可以看出,市场收益率通常在交易日开盘后第 1 个半小时波动的最为剧烈,随后逐渐降低,下午开盘后波动强度又逐渐增大,在时间分布上呈现较为明显的“U”型特征。从图 3 可以看出,上午发帖量总体上高于下午发帖量,且开盘后的一小时是投资者发帖的高峰期,随后逐渐减少,但会在下午开盘后恢复上涨趋势,这一特征和市场收益率的波动状况十分接近。此外,各时间段内的乐观帖普遍多于悲观帖。

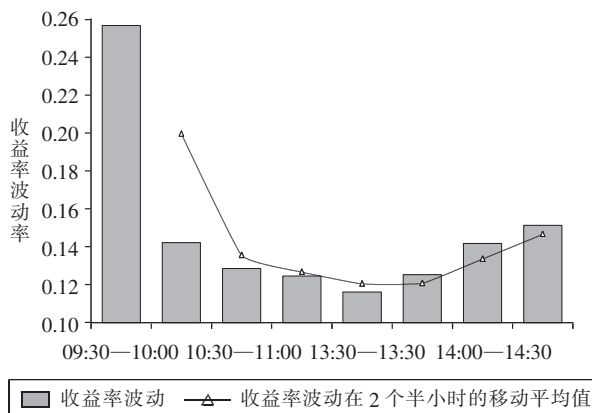


图2 市场(上证指数)收益率日内波动率

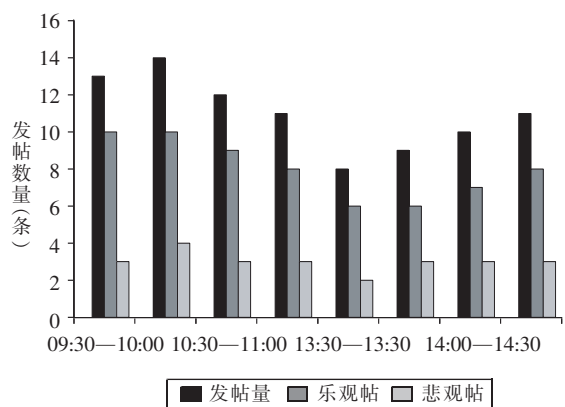


图3 论坛日各时间段发帖数量均值

① 完整的回归结果请参见《中国工业经济》网站(<http://www.ciejournal.org>)公开附件。



进一步对主要变量进行描述性统计分析,其中发帖信息同时包含了交易和非交易时段。可以发现,论坛在每半小时大约可提供8条有效帖,且每条帖子的阅读量接近3000次,说明论坛发帖信息在很大范围上得到了传播。投资者情绪和市场收益率的均值都大于0,但极值相差很大,成交量的极值差异和标准差都较高,情绪和市场波动程度都较为明显。偏度、峰度和JB统计量表明各变量都显著的异于正态分布,ADF检验结果表明各变量都平稳<sup>①</sup>。

## 2. 日内投资者情绪对市场收益率的预测作用

建立基础模型(4)分析滞后的投资者情绪是否可以预测上证指数未来半小时的收益率,其中, $R_{i,t}$ 为上证指数在第*i*交易日第*t*个半小时的收益率, $\text{sentiment}_{i,t-1}$ 为滞后半小时的投资者情绪。由于本文主要关注情绪的日内效应,所以预测回归从各交易日的第二个半小时开始。此外,对每个回归进行了怀特异方差检验及LM检验,对于部分存在异方差或自相关的情况,对标准差进行了Newey-West调整。

$$R_{i,t} = \beta_0 + \beta_1 \text{sentiment}_{i,t-1} + \varepsilon_{i,t}, \quad t=2, \dots, 8 \quad (4)$$

表1报告的结果表明,投资者情绪在总体上都能对后续半小时的收益率产生正向影响。其中,各交易日开盘后第一个半小时的情绪能显著预测第二个半小时的市场收益率,但上午其他半小时内的情绪对收益的预测作用并不显著,系数 $\beta_1$ 逐渐减小。与之相反的是,情绪在下午交易的系数逐渐增大,至少都在5%的显著性水平下通过了检验。

表1 投资者情绪对日内收益的预测作用

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投资者情绪	0.0266* (1.6726)	0.0228 (1.2790)	0.0042 (0.2097)	0.0127** (2.2363)	0.0250*** (2.5956)	0.0272*** (2.6208)	0.0497** (2.2840)
常数项	0.0366** (2.2346)	-0.0235 (-1.3017)	0.0046 (0.2182)	0.0361** (1.9720)	0.0250 (1.5258)	-0.0284 (-1.5306)	0.0076 (0.3334)
N	1171	1173	1175	1169	1121	1131	1144
Adj-R <sup>2</sup>	0.0615	0.1050	0.0617	0.1004	0.0516	0.0706	0.0937
F	2.7977	1.6359	0.0440	3.5134	2.7743	3.7318	5.2167

注:\*,\*\*,\*\*\* 分别表示在10%、5%和1%水平下显著,括号内的数值为t值,以下各表同。

图2已经初步表明中国股市的日内收益率有延续前期趋势的特征,因此考虑到收益率自身的盘中动量效应,本文参照Sun et al.(2016)的做法,进一步在预测模型中控制收益率滞后项。同时,考虑到市场波动对收益率的影响,加入市场收益在滞后半小时中的波动水平,分析在控制了收益率滞后项和市场波动水平的情况下,情绪的预测作用是否有所减弱。建立模型(5)如下所示:

$$R_{i,t} = \beta_0 + \beta_1 \text{sentiment}_{i,t-1} + \beta_2 R_{i,1} + \beta_3 R_{i,t-1} + \beta_4 \sigma_{i,t-1} + \varepsilon_{i,t}, \quad t=2, \dots, 8 \quad (5)$$

其中, $R_{i,1}$ 为第*i*交易日开盘后第一个半小时的收益率, $R_{i,t-1}$ 为滞后半小时的收益率, $\sigma_{i,t-1}$ 为市场收益在滞后半小时中的波动水平,通过对该时间段内的5分钟频率的收益率求标准差获得。

表2报告了回归结果,在控制了收益率滞后项、市场波动水平后,与表1相比,除第五个半小时的收益率外,投资者情绪的预测作用大部分仍然显著,但个别回归系数的大小和显著性低于表1。

① 描述性统计结果请参见《中国工业经济》网站(<http://www.ciejjournal.org>)公开附件。



除此之外,收益率滞后项方面,开盘后第一个半小时的收益对下午其他时间段的收益率都能产生显著的正向影响,第二、四、六、七个半小时的收益率都能促进后续半小时的收益率的增长,而且显著性水平总体上高于情绪变量。

表 2	投资者情绪、滞后收益和收益波动对日内收益的预测作用						
	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投资者情绪	0.0249* (1.6845)	0.0244 (1.5037)	0.0050 (0.2091)	0.0097 (0.5734)	0.0200** (2.2654)	0.0219** (2.0844)	0.0397* (1.6737)
第一个半小时收益率	0.0346** (2.3454)	0.0241 (0.6079)	-0.0228 (-1.2601)	0.0399** (2.4647)	0.0421*** (2.7539)	0.0539*** (3.3567)	0.0651*** (3.3735)
滞后半小时收益率		0.0969*** (3.2423)	-0.0301 (-0.8579)	0.1292*** (5.1058)	0.0270 (0.9468)	0.1201*** (3.7212)	0.2269*** (6.7125)
滞后半小时波动率	0.1087** (2.3479)	0.0207 (0.2253)	0.0135 (0.1019)	0.1724 (0.6087)	0.3492*** (3.9907)	-0.0782 (-0.3490)	-0.2546** (-2.2061)
常数项	0.0098 (0.1275)	-0.0243 (-1.2321)	0.0004 (0.0147)	0.0137 (0.6113)	-0.0167 (-0.8478)	-0.0136 (-0.5348)	0.0588** (2.1371)
N	1171	1173	1175	1169	1121	1131	1144
Adj-R <sup>2</sup>	0.0893	0.0720	0.0622	0.0874	0.0724	0.0646	0.0622
F	3.6526	3.5467	0.6448	8.1892	6.3928	7.0981	18.9003

总体上,表 1 和表 2 实证结果表明,中国股票市场的日内投资者情绪确实能影响市场收益,这种影响独立于收益率自身的日内动量效应,但与 Sun et al.(2016)的发现不同的是,情绪效应比收益率的动量效应要弱。从时间上看,尽管这种影响效应在上午的显著性较低,但在下午交易时段始终保持显著。产生这一现象的原因可能在于投资者交易行为的延迟,隔夜信息流往往会带来不确定性,加之各类市场新闻大多会在股市开盘前后发布,由于“避险情绪”的存在,投资者并不会选择在开盘后马上进行交易,而是在消化市场信息后做出判断和决策。相对于上午,投资者可以更好地把握股市在下午第五、六、七个半小时这几个时间段的信息,这提高了投资者的交易动机,使情绪的预测效应得以体现。

3. 日内情绪的预测作用在不同市场环境下的表现

以上分析都是基于全样本得到的预测结果,为了区分投资者情绪的预测作用在不同市场状态下的表现,借鉴 Pagan and Sossounov(2003)的方法,将市场环境划分为牛熊市两种状态。首先,对市场指数设定一个单向运行周期,将样本区间细分为多个子区间;其次,将各个运行周期中第  $i$  日的市场指数与相邻交易日进行对比,运行周期中会交替出现极高点 and 极低点,选取这些连续极值点中的最高者(或连续波谷中的最低者)作为该周期内牛熊市状态转换的转折点。本文将上证指数的单向运行周期设定为 5 个月,可以将样本区间大致划分为牛市和熊市各五个阶段<sup>①</sup>。

表 3 和表 4 分别为牛市和熊市下的回归结果,可以发现,在牛市环境下投资者情绪主要在第二、三、六、七、八这五个半小时的收益率产生显著的正向影响,显著的时间区间甚至多于全样本;熊市环境下,情绪的预测作用远不及牛市,但投资者情绪至少在收盘前的一小时内始终保持显著;滞后收益率等控制变量在熊市中的显著性总体上高于牛市。这说明投资者情绪的预测作用在牛市中

① 牛熊市市场状态阶段的划分结果请参见《中国工业经济》网站(<http://www.ciejjournal.org>)公开附件。

表现得更显著。一个可能的原因是,牛市中良好的市场环境会提升投资者的乐观情绪倾向,使投资者更积极地参与实际交易中,显著影响市场收益率,而熊市环境下的投资者在日内并不会频繁交易,市场运行主要受到前期收益率和波动水平的影响。

表 3 牛市环境下投资者情绪、滞后收益和收益波动对日内收益的预测作用

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投资者情绪	0.0388** (2.1881)	0.0537*** (2.6962)	0.0048 (0.1391)	0.0005 (0.0305)	0.0176** (2.4491)	0.0267*** (2.6382)	0.0287** (2.3198)
第一个半小时收益率	0.0003 (0.0103)	0.0738** (2.4122)	-0.0488 (-1.0296)	0.0338* (1.6756)	0.0227 (0.8159)	0.0234 (1.0486)	0.0615** (2.2695)
滞后半小时收益率		-0.1625*** (-3.1001)	0.0191 (0.2944)	-0.0384 (-0.5093)	0.0177 (0.2239)	0.2058*** (5.3281)	0.1951*** (4.3446)
滞后半小时波动率	0.1448 (0.9262)	0.1910 (0.9053)	-0.2030 (-0.8135)	-0.1721 (-0.5086)	0.3180** (2.5334)	0.1195 (0.7524)	-0.1085 (-0.6810)
常数项	0.0302 (1.1181)	-0.0580* (-1.9265)	0.0293 (0.7569)	0.0593 (1.4870)	-0.0112 (-0.5326)	-0.0119 (-0.4124)	0.0566 (1.5805)
N	729	731	731	728	693	706	703
Adj-R <sup>2</sup>	0.0715	0.0604	0.0665	0.0889	0.120	0.419	0.0888
F	2.6743	9.6349	1.1543	2.7476	2.3253	7.6729	7.0403

表 4 熊市环境下投资者情绪、滞后收益和收益波动对日内收益的预测作用

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投资者情绪	0.0265 (0.9829)	0.0231 (0.7947)	0.0078 (0.2261)	0.0130 (0.4095)	0.0263 (1.0611)	0.0336** (2.4325)	0.0570* (1.7679)
第一个半小时收益率	0.0645*** (2.8374)	0.0299 (0.4649)	0.0045 (0.1702)	0.0424* (1.8048)	0.0562** (2.5190)	0.0732* (1.8468)	0.0586** (2.0265)
滞后半小时收益率		0.0065*** (2.7780)	0.0694 (1.3617)	0.3563*** (8.1526)	0.0306*** (2.7199)	0.0107 (0.1106)	0.2606*** (4.9668)
滞后半小时波动率	0.1230** (2.0355)	-0.2256 (-1.2861)	0.3374 (1.5811)	0.4849*** (2.9429)	0.3801*** (2.9401)	-0.1889 (-0.6150)	0.3975** (2.2948)
常数项	-0.0220 (-0.6873)	0.0163 (0.3821)	-0.0442 (-0.9580)	-0.0418 (-1.0251)	-0.0223 (-0.6863)	-0.0370 (-0.9032)	0.0518 (1.1711)
N	444	442	444	441	428	425	441
Adj-R <sup>2</sup>	0.0418	0.0564	0.0801	0.1449	0.0772	0.0884	0.0906
F	3.2713	2.6989	1.1177	18.4734	6.0899	3.8759	10.8637

#### 4. 日内情绪的预测效应在极端市场环境中的表现

上述对市场环境作了牛市和熊市的划分,但是极端市场条件下,高频情绪对收益率的预测作用有何规律?考虑到 2015 年股灾期间是一个较好的时间窗口,根据上证综指走势的极高/极低点,选择 2015 年 6 月 12 日至 2015 年 7 月 8 日这一暴跌时间段;再以 2015 年 6 月 12 日(当日上证综指

最高至 5178.19 点)为极高点,向前追溯到前一个极低点(2015 年 5 月 7 日),作为暴涨时间段,重点分析极端市场环境中的日内情绪效应。

本文的回归结果表明,以 2015 年 6 月 12 日上证综指的最高点为界,无论前期暴涨阶段,还是之后的暴跌阶段,投资者情绪对日内收益率的影响方向始终为正。但暴涨阶段时的投资者情绪比暴跌阶段对后续半小时收益率的影响更为显著,这与牛市下的情况一致。从回归系数大小上看,极端环境期间的情绪系数普遍增加;从显著性上看,情绪效应通过显著的检验的半小时区间个数比牛/熊市环境下还要多。这一方面表明当市场处于极端环境时,投资者情绪对市场后续收益率的影响强度会显著放大;另一方面,与全样本下的“牛市中情绪主导收益率、熊市中市场变量主导收益率”不同的是,在暴涨/暴跌的极端市场环境中,投资者情绪对日内收益率的影响都不容小觑,比滞后收益率、滞后波动率的影响效应更显著<sup>①</sup>。

## 五、进一步研究

### 1. 隔夜投资者情绪的预测作用

考虑到资本市场中的隔夜信息流和开盘前信息也可能是诱发投资者情绪的重要原因。在这段时间的情绪是否也如日内情绪一样影响开盘后市场收益率,值得做进一步探索。鉴于此,建立计量模型(6), $\text{sentiment}_{i,bo}$ 表示第  $i-1$  个交易日收盘到第  $i$  个交易日开盘前的情绪,定义为隔夜投资者情绪:

$$R_{i,t} = \beta_0 + \beta_1 \text{sentiment}_{i,bo} + \beta_2 R_{i,1} + \beta_3 R_{i,t-1} + \beta_4 \sigma_{i,t-1} + \varepsilon_{i,t}, \quad t=1, \dots, 4 \quad (6)$$

表 5 报告了分别对应开盘后四个半小时时间段的预测结果,除了第一个半小时(09:30—10:00)外,其他半小时都包括两列回归结果,前列仅包含情绪项解释变量,后列控制了其他市场变量。出乎意料的是,隔夜情绪并没有显著预测第一个半小时的收益率,当单独考虑情绪这一个解释变量时,它能对第二至四个半小时的收益率产生正向显著影响。进一步控制市场相关变量后,情绪对第二个半小时的预测作用变得不再显著,但仍显著影响第三、四个半小时的收益率。这一结果说明隔夜情绪的释放具有时滞性,它并非在开盘后短时间内就能冲击市场,而是集中作用于第三个半小时(10:00—10:30)的收益率,并在后续时间里持续减弱。

### 2. 投资者情绪的“半日效应”

中国股市交易机制的一个特点是从 11:30—13:00 有 90 分钟的午休时间,这段时间也是网络媒介活跃和信息传递的高峰期,投资者可以在此期间获取、消化新的市场信息,并可能作用于下午开盘后的市场运行。鉴于此,考虑午休期间投资者情绪的影响,建立模型(7):

$$R_{i,qft} = \beta_0 + \beta_1 \text{sentiment}_{i,mor} + \beta_2 \text{sentiment}_{i,noon} + \beta_3 R_{i,mor} + \varepsilon_i \quad (7)$$

区别于前述模型,模型(7)将情绪的预测作用从半小时扩展到了半日。其中, $R_{i,qft}$ 为市场在第  $i$  日 13:00—15:00 的收益率, $\text{sentiment}_{i,noon}$ 为午休期间的投资者情绪, $\text{sentiment}_{i,mor}$ 和  $R_{i,mor}$ 分别为市场在 9:30—11:30 的投资者情绪和收益率。

表 6 的逐步回归结果表明,仅考虑上午的投资者情绪时,它能显著正向影响下午的市场收益率,但随着解释变量的增加,上午的投资者情绪变得不再显著,且系数值逐渐较小。整个过程中,上午收益率的系数始终正向显著,这与 Zhang et al.(2018)对股票市场日内收益率规律的研究结论一致。重要的是,午间休市时间段的投资者情绪的系数明显较上午有所增大且通过了 5%显著性水平

<sup>①</sup> 具体详细的回归结果请参见《中国工业经济》网站(<http://www.ciejjournal.org>)公开附件。



表 5 隔夜投资者情绪对股票收益率的预测作用

	第一个半小时 *		第二个半小时		第三个半小时		第四个半小时	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
隔夜投资者情绪	0.0543 (0.9469)		0.0271* (1.7347)	0.0257 (0.8976)	0.0268*** (2.5921)	0.0311** (2.1409)	0.0656* (1.8186)	0.0662* (1.8337)
第一个半小时收益率				0.0377** (2.3532)		0.0200 (1.2689)		-0.0263 (-1.3893)
滞后半小时收益率						0.1193*** (-3.8500)		-0.0262 (-0.7365)
滞后半小时波动率				0.1072** (2.3712)		0.0042 (0.0405)		0.0320 (0.2402)
常数项	-0.1084** (-2.3545)		0.0303 (1.3189)	0.0046 (0.1796)	0.0083 (0.3428)	0.0135 (0.4798)	-0.0336 (-1.1600)	-0.0403 (-1.2045)
N	1140		1140	1140	1140	1140	1140	1140
Adj-R <sup>2</sup>	0.0908		0.0618	0.0778	0.0707	0.0847	0.0929	0.1053
F	1.8967		2.8953	2.9794	2.7959	4.2223	3.3075	3.0227

注：\* 当开盘后第一个半小时收益率为被解释变量时，解释变量只包含隔夜投资者情绪，无法控制第一个半小时收益率和滞后半小时波动率。因此，第一个半小时对应的第(2)列回归结果为空白。

表 6 投资者情绪和收益率的“半日效应”

	模型(7-1)	模型(7-2)	模型(7-3)
上午投资者情绪	0.0323* (1.8314)	0.0268 (0.4435)	0.0154 (0.2979)
上午股票收益率		0.1071*** (3.0769)	0.1049*** (3.0427)
午休投资者情绪			0.0388** (2.3245)
常数项	0.0355 (0.5748)	0.0457 (0.0843)	0.0217 (0.3566)
N	1187	1187	1187
Adj-R <sup>2</sup>	0.1002	0.0945	0.1076
F	3.4824	7.3464	6.3765

检验，说明在中国股市收益率的预测中，确实存在“半日效应”，这种效应不仅取决于上午的总体收益率，还来源于上午和午休期间的投资者情绪，相比之下，投资者在午休期间的情绪状态发挥着更明显的作用。

### 3. 噪音交易对日内情绪预测作用的影响

前述实证证明了高频投资者情绪对上证指数的日内收益有显著的预测作用。一个可能的解释是，这种预测效应主要由噪音交易者的交易行为驱动。接下来，本文试图从交易量和股票长期收益的角度来验证这一猜想。Barber and Odean(2008)指出，噪音交易活动的主要表现是市场上的成交量显著提高，且成交量可作为噪音交易的代理变量。本文认为，若噪音交易假设成立，则投资者情绪的预测作用理应会在成交量高的时间区间内更显著。因此，进一步建立模型(8)以验证噪音交易假设在日内频率上是否成立：

$$R_{i,t} = \beta_0 + \beta_1 sentiment_{i,t-1} + \beta_2 HighVol_{i,t} + \beta_3 sentiment_{i,t-1} HighVol_{i,t} +$$

$$\beta_4 R_{i,1} + \beta_5 R_{i,t-1} + \beta_6 \sigma_{i,t-1} + \varepsilon_{i,t} \quad t=2, \cdots, 8 \tag{8}$$

其中,  $HighVol_{i,t}$  为交易量在  $i$  日第  $t$  个半小时的虚拟变量, 如果第  $t$  个半小时的交易量高于所有样本在该时段的均值, 则取值为 1, 否则取值为 0。  $sentiment_{i,t-1} HighVol_{i,t}$  为交易量和情绪的交互项, 若噪音交易假设成立, 该交互项系数  $\beta_3$  理应为正, 且在统计上显著。特别地, 考虑到交易量通常具有长期趋势, 本文参照 Sun et al.(2016) 首先对其进行对数转换, 再对各交易量数据减去相应 500 个半小时的移动平均值, 以减弱交易量的趋势性。

表 7 报告的回归结果显示, 在加入交易量与情绪的交互项后, 上午时段的第二个、下午时段的第六、七、八个半小时的收益率都至少在 10% 的显著性水平下受到滞后投资者情绪的影响。交互项在第六、七、八个半小时也通过了显著性检验且系数为正, 说明相比于交易量低的半小时时段, 情绪效应在更高交易量的时间段具有更显著的预测效果, 这也证明了噪音交易行为是情绪影响市场收益率的可能原因。

表 7 交易量和投资者情绪对日内收益的预测作用							
	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投资者情绪	0.0009* (1.8388)	0.0087 (0.5189)	0.0211 (1.4008)	0.0081 (0.5463)	0.0115** (2.5324)	0.0157** (2.2512)	0.0274* (1.8756)
第一个半小时收益率	0.0346 (1.1279)	0.0246 (0.5849)	-0.0196 (-0.5160)	0.0378 (1.3956)	0.0424 (1.3437)	0.0543** (2.0813)	0.0660** (2.0467)
滞后半小时收益率		0.0974 (1.4997)	-0.0339 (-0.9685)	0.1278 (1.6332)	0.0259 (0.4565)	0.1190** (2.4141)	0.2257*** (3.5740)
滞后半小时波动率	0.1109 (1.1071)	0.0209 (0.1075)	0.0678 (0.2737)	0.1518 (0.5271)	0.3437 (1.0573)	-0.0777 (-0.3852)	-0.2418 (-0.9187)
交易量	0.0211 (0.6325)	-0.0300 (-0.8672)	-0.1031* (-1.8103)	0.0265 (0.6585)	0.0220 (0.6219)	-0.0231 (-0.6326)	-0.0562 (-1.1296)
交易量×滞后半小时投资者情绪	0.0473* (1.6667)	0.0357 (1.1508)	0.0613 (1.1422)	0.0027 (0.0694)	0.0212** (2.2928)	0.0298*** (2.8954)	0.0269** (2.5123)
常数项	-0.0022 (-0.0829)	-0.0114 (-0.3487)	0.0387 (1.1280)	0.0036 (0.1111)	-0.0266 (-0.0840)	-0.0026 (-0.1074)	0.0842** (2.0056)
N	1171	1173	1175	1169	1121	1131	1144
Adj-R <sup>2</sup>	0.1153	0.0828	0.0773	0.0682	0.0929	0.0553	0.0637
F	3.7588	2.1213	1.4112	2.6845	8.7840	9.3257	13.4364

从另一个角度看, 如果投资者日内的情绪效应是由噪声交易驱动的话, 那么这种预测作用会在长期内发生反转, 因为噪音交易活动只会在短期内推动股票价格偏离基本价值, 而股价在更长的时间内会发生反转。因此, 进一步构建模型(9)验证这一假设:

$$R_{i,t+h} = \beta_0 + \beta_1 sentiment_{i,t-1} + \beta_2 R_{i,1} + \beta_3 R_{i,t-1} + \beta_4 \sigma_{i,t-1} + \varepsilon_{i,t+h} \tag{9}$$

其中,  $h$  表示不同的时间区间,  $R_{i,t+h}$  为上证指数从第  $i$  天的  $t$  半小时起, 后续  $h$  个半小时时段的收益率, 比如  $h=[0,5]$  表示股票在第  $[t+0,t+5]$  这一持有期内的收益率。本文以  $h=[0,5]$  衡量短期, 以  $h=[6,20]$  衡量长期, 研究情绪的预测作用在长期内是否会发生反转。

表 8 中的各列分别对应股票市场从第  $t$  半小时开始, 后续五个半小时内的收益率, 各行对应第

$t-1$  半小时的解释变量;表 9 为情绪对更长期收益率的影响。通过对比两表的预测结果可以看出,当把股票持有期扩展到五个半小时后,投资者情绪的系数值为正,而把持有期扩展到更长期时,尽管投资者情绪的显著性有所降低,但系数符号方向的改变揭示了股票收益率的确会在长期发生反转,这进一步验证了噪音交易行为在投资者情绪对收益率的预测作用中扮演的重要角色。

表 8 短期内  $h=[0,5]$  投资者情绪对日内收益的预测作用

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时	0.0027	0.0379	0.0822	0.0582*	0.0396***	0.0400**	0.0705**
投资者情绪	(0.0798)	(0.7381)	(1.2615)	(1.6534)	(2.8368)	(2.3951)	(2.4465)
第一半小时	0.0650	0.1925***	0.1982**	0.1797*	0.1756***	0.0518	0.0267
收益率	(0.9722)	(2.9253)	(2.0483)	(1.9117)	(3.6158)	(0.5165)	(0.6022)
滞后半小时		0.0940	0.0340	0.0025	0.4469***	0.2777*	0.1786**
收益率		(0.6018)	(0.2115)	(0.0160)	(4.9278)	(1.7879)	(2.3005)
滞后半小时	-0.3028	0.7320	-0.6239	-0.5781	-0.5141*	-1.2923**	-0.5216*
波动率	(-1.5936)	(1.3163)	(-0.9765)	(-0.9593)	(-1.8496)	(-2.0392)	(-1.9685)
常数项	0.1023*	-0.0459	0.1627*	0.1223	0.0520	0.1319	0.1088*
	(1.9384)	(-0.5219)	(1.8913)	(1.4836)	(0.8313)	(1.6481)	(1.7220)
N	1171	1173	1174	1169	1120	1131	1144
Adj-R <sup>2</sup>	0.0634	0.0462	0.0986	0.0758	0.0685	0.0626	0.1009
F	5.3004	6.6259	5.5318	4.6831	11.1674	6.5047	5.1280

表 9 长期内 ( $h=[6,20]$ ) 投资者情绪对日内收益的预测作用

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
滞后半小时投	-0.0040	-0.1480	-0.1163	-0.1494*	-0.0274	-0.0039**	-0.0428*
投资者情绪	(-0.0425)	(-1.6111)	(-1.3684)	(-1.8176)	(-0.3382)	(-2.4465)	(-1.6780)
第一半小时收	0.1524	0.0498	0.0621	0.0058	0.0105	0.0335	0.0269
收益率	(0.8637)	(0.2653)	(0.4232)	(0.0705)	(0.1263)	(0.1921)	(0.3303)
滞后半小时收	—	0.3760	0.0981	-0.2525*	-0.3449**	-0.6381***	-0.5625***
收益率		(1.3866)	(0.4993)	(-1.8894)	(-2.2259)	(-2.7362)	(-3.9439)
滞后半小时波	0.3954	0.4745	0.6695	-1.5641***	-1.3825***	-1.6621	-0.7843
动率	(0.7032)	(0.4161)	(0.5995)	(-2.8227)	(-2.9130)	(-1.2133)	(-1.6108)
常数项	-0.0622	-0.2048	-0.0818	0.1113	0.1977*	0.2565*	0.1585
	(-0.4056)	(-1.2501)	(-0.7153)	(0.9398)	(1.8508)	(1.6585)	(1.3651)
N	1169	1171	1173	1168	1119	1131	1144
Adj-R <sup>2</sup>	0.1033	0.0973	0.0535	0.0617	0.0626	0.0421	0.0558
F	1.3420	2.1340	1.5432	4.4659	3.5602	5.435	4.5767

#### 4. 内生性问题的探讨

对于可能存在的内生性问题使实证结果产生偏误,本文主要从以下几个方面梳理并排除:

(1) 由于本文是利用半小时区间的投资者情绪预测后续  $t+1$  期的市场收益率,所以可以排除因变量和自变量之间存在交互影响的可能性,避免这一问题可能带来的偏误。同时,本文所有变量均来自于实际测量数据,没有使用代理变量,测度误差带来内生性的可能性也较小。



(2)变量遗漏也可能导致核心解释变量(投资者高频情绪)与残差项相关,产生内生性问题。在这里主要考虑残差项中包含的“价格”因素,因为第  $t$  期投资者情绪尽管无法受到  $t+1$  期收益率的影响,但很可能受同期收益率的影响。鉴于此,前述实证中除投资者情绪外,又逐步加入了  $t$  期收益率、收益率波动等可能表征“价格”因素的变量,最大程度排除产生内生性的可能。

(3)尽管控制了滞后收益率等变量,仍然可能存在遗漏变量以及对情绪和股票收益率都有影响的变量。所以,接下来选取日内半小时区间的市场流动性作为投资者情绪的工具变量,运用 2SLS 分别对日内 7 个半小时区间中的情绪对收益率的预测作用进行更精确的衡量。两阶段最小二乘法估计的结果表明<sup>①</sup>,工具变量都通过了至少 10% 显著性水平的检验,Hansen J 统计量也通过了至少 5% 的显著性水平检验,证实所选工具变量较合适。通过将 2SLS 回归与表 2 中的 OLS 结果作对比可以发现,各变量的系数大小和方向基本一致,工具变量的显著时间段、系数符号方向与表 2 高度一致,进一步说明了实证结果的可靠性。

## 六、稳健性检验

本部分通过对样本按照月份和工作日进行划分,从多个维度来检验投资者情绪对日内收益预测作用的稳健性。此外,通过进一步控制股票市场之外的宏观经济变量,分析基于文本构建的投资者情绪指数对股票市场运行的影响是否会受到其他因素的干扰。

### 1. 月份效应和星期效应

季节性往往是资产定价异常的重要特征,Cooper et al.(2006)发现美国股市收益存在显著的“一月效应”,Da et al.(2015)也发现基于 Google 搜索量构建的投资者情绪指数存在显著的“星期效应”。鉴于此,本文按照不同月份进行模型(10)中的预测回归,按照不同工作日进行模型(11)中的预测回归,以检验前述实证结果是否受特定月份和工作日的影响<sup>②</sup>。

模型(10)中, $R_{mi,t}$ 代表上证指数在  $m$  月中第  $i$  天的第  $t$  半小时的收益率, $sentiment_{mi,t-1}$ 和  $R_{mi,t-1}$ 分别为滞后半小时的投资者情绪和收益率, $R_{mi,1}$ 为开盘后第一个半小时的收益率。

$$R_{mi,t} = \beta_0 + \beta_1 sentiment_{mi,t-1} + \beta_2 R_{mi,1} + \beta_3 R_{mi,t-1} + \beta_4 \sigma_{mi,t-1} + \varepsilon_{mi,t} \quad m=1, \dots, 12 \quad (10)$$

通过观察由模型(10)得到的完整回归结果中的  $\beta_1$  的值,表 10 中汇总报告了投资者情绪对日内收益率有显著预测作用的月份,可以发现投资者情绪在一年中大部分的月份都能预测市场收益率,这种预测效果在开盘后的第五、六、七、八个半小时表现的最为显著。

同理,将样本数据按照工作日划分,进行模型(11)中的预测回归,其中  $R_{wi,t}$ 代表第  $i$  个星期  $w$  ( $w$  取值一、二、三、四、五)中第  $t$  个半小时的收益率:

$$R_{wi,t} = \beta_0 + \beta_1 sentiment_{wi,t-1} + \beta_2 R_{wi,1} + \beta_3 R_{wi,t-1} + \beta_4 \sigma_{wi,t-1} + \varepsilon_{wi,t} \quad w=1, \dots, 5 \quad (11)$$

表 11 汇总报告了投资者情绪对收益率有显著预测作用的工作日,可以看出,情绪的预测作用同样在交易日的下午更显著。具体来说,第二和第三个半小时的预测作用只在周三和周五显著,第四个半小时的预测作用只在周二显著,但第六至八个半小时的收益率几乎在所有工作日都能被显著预测。

总而言之,本文发现情绪的影响效应在不同时间维度普遍存在,尽管这种预测作用在上午表现相对较弱,但在下午交易时段第六至八个半小时表现出非常明显的预测效应。

① 完整回归结果请参见《中国工业经济》网站(<http://www.ciejournal.org>)公开附件。

② 模型(10)和模型(11)的完整回归结果请参见《中国工业经济》网站(<http://www.ciejournal.org>)公开附件。

表 10 投资者情绪对日内收益率有显著预测作用的月份汇总

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
投资者情绪对 日内收益率有 显著预测作用 的月份	7 月、8 月 10 月、11 月	1 月、4 月 5 月、12 月	2 月、4 月 10 月、11 月	2 月、6 月、 7 月、8 月、 9 月、12 月	1 月、2 月 6 月、7 月 8 月、9 月、 10 月	1 月、3 月、 5 月、7 月 8 月、9 月 10 月、11 月	2 月、4 月 7 月、8 月 9 月、11 月、 12 月

表 11 投资者情绪对日内收益有显著预测作用的工作日汇总

	第二个 半小时	第三个 半小时	第四个 半小时	第五个 半小时	第六个 半小时	第七个 半小时	第八个 半小时
投资者情绪对 收益率有显著 预测作用的工 作日	星期三 星期五	星期三 星期五	星期二	星期二 星期五	星期一 星期三 星期四	星期一 星期三 星期四 星期五	星期一 星期二 星期三 星期五

## 2. 滞后宏观经济变量的影响

接下来进一步探讨在控制了宏观经济变量后,投资者情绪的预测作用是否受到影响。借鉴 Sun et al.(2016)的做法,本文选取违约利差(*DEF*)、期限利差(*TERM*)和短期无风险利率(*RATE*)作为宏观经济变量的代表。其中,违约利差(*DEF*)选取中信标普企业债指数收益率和国债指数收益率,利用两者之差来表示违约利差;期限利差(*TERM*)选取 10 年期和 1 年期国债到期收益率之差;短期无风险利率利用 SHIBOR 隔夜拆借利率衡量。由于日内高频的宏观经济变量不可获取,本文以滞后一天的宏观经济变量作为滞后半小时宏观经济状况的代理变量,建立模型(12)。回归结果表明,宏观经济变量并没有影响投资者情绪效应和收益率动量效应,除了违约利差和期限利差对第二个半小时的收益产生影响外,其他时间区间内的宏观经济变量都不显著<sup>①</sup>。

$$R_{i,t} = \beta_0 + \beta_1 sentiment_{i,t-1} + \beta_2 R_{i,t-1} + \beta_3 R_{i,t-1} + \beta_4 \sigma_{i,t-1} + \beta_5 rate_{i-1} + \beta_6 Def_{i-1} + \beta_7 Term_{i-1} + \varepsilon_{i,t} \quad (12)$$

### 3. 1 小时间隔的投资者情绪对日内收益率的预测作用

考虑到不同的时间段分割长度可能会对实证结果产生影响,进一步以 1 小时为分割点对各变量数据重新进行整理,分析扩大时间区间后投资者情绪对日内收益率的预测作用是否仍然稳健。回归结果表明,将交易时间区间按照 1 小时分割后,尽管投资者情绪的系数有减小的趋势,但仍能对后续 1 小时的收益率产生显著的正向影响,这说明本文研究的高频投资者情绪对日内收益率的预测作用是稳健的<sup>②</sup>。

## 七、研究结论与启示

### 1. 研究结论

投资者情绪能够很大程度上影响股票收益率,这一观点在近年来大量的文献中已经得到印证。但限于数据指标构建的困难,投资者高频情绪的度量目前还处于探索阶段。本文利用数据挖掘手段

① 滞后宏观经济变量的影响的回归结果请参见《中国工业经济》网站(<http://www.ciejjournal.org>)公开附件。

② 以 1 小时为分割点的回归结果请参见《中国工业经济》网站(<http://www.ciejjournal.org>)公开附件。

获取“金融界”股票论坛中的发帖信息,通过 Python 情感分析算法构建了上证指数的日内高频投资者情绪指标,从多个维度研究了中国股市中投资者情绪对股票日内收益的预测效应,并从交易量和市场长期收益反转的角度探讨了噪音交易在这一过程中所扮演的角色。实证发现:①中国股票市场中的日内投资者情绪确实能正向影响股票市场运行,这种预测作用在交易日的下午表现更显著,它独立于收益率自身的日内动量效应,但较收益率的动量效应要弱。②牛市中投资者情绪对日内收益率的预测作用强于滞后收益率等市场变量,熊市则相反。但在极端市场环境中,无论是暴涨还是暴跌,投资者情绪相较于市场滞后收益率等变量,对日内收益率的影响程度都更为显著。③隔夜情绪的释放会显著影响次日上午的市场收益率,但存在时滞性,午休期间的投资者情绪会联同上午收益率一起影响下午交易时间段的市場表现。④投资者的噪音交易行为是投资者情绪影响股票收益率的重要驱动力量。⑤稳健性检验中,投资者高频情绪对股票日内收益的预测效应在考虑了月份效应、星期效应以及宏观经济变量后,在不同时间维度仍然普遍存在。

本文利用互联网大数据挖掘和文本分析的方法,提取投资者在网络平台里的实时发帖背后隐藏的情绪状态,并从实证角度揭示了高频投资者情绪在中国股市日内收益预测中发挥的重要作用。必须指出的是,本文发现尽管噪音交易者在股价波动中也发挥着重要作用,但其情绪效应在总体上并没有超过市场收益率滞后变量以及市场波动的影响。同时,隔夜情绪会对市场次日开盘后产生冲击,午间休市期间的投资者情绪也在股票市场“半日效应”中发挥重要作用。这些发现不仅是对已有文献的有益补充,也有助于从更高频率的视角深入理解股市中情绪效应的特征和机理。

## 2. 启示与政策建议

(1)深刻理解股票市场信息结构的多样化和复杂性。目前对股票市场信息的监管多关注成熟的传统新闻媒介,如报纸、电视等等,但随着网络和移动通信技术的发展,信息产生和传播方式都发生了巨大变化。一方面,由于互联网的匿名性,投资者更加倾向于表达内心真实情感,发布对股市未来走势的真实看法;另一方面,信息通过不同渠道的分享功能得以快速传播并交互影响。投资者不再是信息的单纯的被动接受者,还能通过发帖、留言等方式发布自己的观点和想法。这些网络信息一般不会在市场交易数据中得到反应,却蕴含着对市场运行有显著解释力的有效信息,从根本上改变了股票市场的信息结构。

(2)从上市公司角度看,应跟随互联网、移动通信技术快速发展的步伐,加强对互联网、移动通信媒介的应用和管理,通过多种手段充分与投资者进行双向沟通,并接受投资者的合理建议。这样不仅可以提升信息的传递效率和准确度,更可以及时澄清网络上良莠不齐的错误信息,保障投资者利益。

(3)对投资者而言,互联网给投资者获取信息提供了便利,但也使信息的真伪甄别更加困难,投资者有可能落入信息陷阱。投资者应理性认识股票市场信息传播渠道,多方位、全面了解上市公司和各类市场信息,提升理性决策水平。

(4)对监管部门来说,股票市场运行的基础是信息传递的及时性和有效性,特别是在互联网背景下,论坛、股吧、微博和微信等新型传播媒介方式不断涌现,深刻影响着投资者决策心理及行为,进而影响股票市场运行。监管部门应动态调整信息披露的渠道和机制,将互联网和自媒体等传播渠道纳入信息监管的范畴,加大违规散布虚假信息行为的惩罚力度。在技术手段上,应该提升并充分利用大数据工具进行市场信息监管,加强对互联网信息的实时监控。同时,监管部门还应该加强对个体投资者的教育,帮助提高其理性程度、信息甄别能力,避免成为市场虚假信息和消极情绪的传播者。

## 〔参考文献〕

- [1]段江娇,刘红忠,曾剑平.中国股票网络论坛的信息含量分析[J].金融研究,2017,(10):178-192.



- [2]黄燕芬,张超. 大数据情绪指数与经济学研究:现状、问题与展望[J]. 教学与研究, 2018,(5):40-50.
- [3]林振兴. 网络讨论、投资者情绪与IPO抑价[J]. 山西财经大学学报, 2011,(2):23-29.
- [4]刘善存,许敏. 基于高频交易数据的上海证券市场投资者风险态度实证研究[J]. 系统工程, 2007,(7):7-12.
- [5]陆静,周媛. 投资者情绪对股价的影响——基于AH股交叉上市股票的实证分析[J]. 中国管理科学, 2015,(11):21-28.
- [6]孟庆斌,杨俊华,鲁冰. 管理层讨论与分析披露的信息含量与股价崩盘风险——基于文本向量化方法的研究[J]. 中国工业经济, 2017,(12):134-152.
- [7]孟雪井,孟祥兰,胡杨洋. 基于文本挖掘和百度指数的投资者情绪指数研究[J]. 宏观经济研究, 2016,(1):144-153.
- [8]石善冲,朱颖楠,赵志刚. 基于微信文本挖掘的投资者情绪与股票市场表现 [J]. 系统工程理论与实践, 2018,(6):1404-1412.
- [9]王春. 投资者情绪对股票市场收益和波动的影响——基于开放式股票型基金资金净流入的实证研究[J]. 中国管理科学, 2014,(9):49-56.
- [10]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.
- [11]谢军,杨春鹏,闫伟. 高频环境下股指期货市场情绪冲击效应[J]. 系统工程, 2012,(9):27-36.
- [12]伊志宏,杨圣之,陈钦源. 分析师能降低股价同步性吗——基于研究报告文本分析的实证研究[J]. 中国工业经济, 2019,(2):134-152.
- [13]Antonioni, C., and J. A. Doukas. Cognitive Dissonance, Sentiment, and Momentum [J]. Journal of Financial & Quantitative Analysis, 2013,48(1):245-275.
- [14]Antweiler, W., and M. Z. Frank. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards[J]. Journal of Finance, 2004,59(3):1259-1294.
- [15]Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. Value and Momentum Everywhere[J]. Journal of Finance, 2013,68(3):929-985.
- [16]Baker, M., and J. Wurgler. Investor Sentiment and the Cross-Section of Stock Returns[J]. Journal of Finance, 2006,61(4):1645-1680.
- [17]Barber, B. M., and T. Odean. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors[J]. Review of Financial Studies, 2008,21(2):785-818.
- [18]Bollen, J., and H. Mao. Twitter Mood as a Stock Market Predictor[J]. Computer, 2011, 44(10):91-94.
- [19]Chen, H., P. De, and Y. Hu. Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media[J]. Review of Financial Studies, 2014,27(5):1367-1403.
- [20]Cooper, M. J., J. J. McConnell, and A. V. Ovtchinnikov. The Other January Effect [J]. Social Science Electronic Publishing, 2006,82(2):315-341.
- [21]Da, Z., J. Engelberg, and P. Gao. The Sum of All Fears Investor Sentiment and Asset Prices [J]. Social Science Electronic Publishing, 2015,28(10):1-32.
- [22]DeLong, J. B., A. Shleifer, and L. H. Summers. Noise Trader Risk in Financial Markets [J]. Journal of Political Economy, 1990,98(4):703-738.
- [23]Derrien, F. IPO Pricing in “Hot” Market Conditions: Who Leaves Money on the Table[J]. Social Science Electronic Publishing, 2005,60(1):487-521.
- [24]Gao, L., Y. Han, and S. Z. Li. Market Intraday Momentum [J]. Journal of Financial Economics, 2018,129(2):394-414.
- [25]Hung C. H. When Does Investor Sentiment Predict Stock Returns[J]. Journal of Empirical Finance, 2012,19(2):217-240.
- [26]Jegadeesh, N., and S. Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency[J]. Journal of Finance, 1993,48(1):65-91.

- [27]Keynes, J. M. The General Theory of Employment, Interest, and Money[M]. London: MacMillan, 1936.
- [28]Narayan, P. K., and S. S. Sharma. Intraday Return Predictability, Portfolio Maximization, and Hedging[J]. Emerging Markets Review, 2016,28(3):105–116.
- [29]Pagan, A. R., and K. A. Sossounov. A Simple Framework for Analyzing Bull and Bear Markets [J]. Journal of Applied Econometrics, 2003,18(1):23–46.
- [30]Rahman, S., C. F. Lee, and K. P. Ang. Intraday Return Volatility Process: Evidence from NASDAQ Stocks[J]. Review of Quantitative Finance & Accounting, 2002,19(2):155–180.
- [31]Renault, T. Intraday Online Investor Sentiment and Return Patterns in the U.S. Stock Market [J]. Journal of Banking & Finance, 2017,84(7):25–40.
- [32]Sun, L., M. Najand, and J. Shen. Stock Return Predictability and Investor Sentiment: A High-Frequency Perspective[J]. Journal of Banking & Finance, 2016,73(7):147–164.
- [33]Wysocki, P. D. Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards[J]. SSRN Electronic Journal, 1998,20(2):56–70.
- [34]You, W., Y. Guo, and C. Peng. Twitter's Daily Happiness Sentiment and the Predictability of Stock Returns[J]. Finance Research Letters, 2017,23(11):54–70.
- [35]Zhang, Y., F. Ma, and B. Zhu. Intraday Momentum and Stock Return Predictability Evidence from China[J]. Economic Modeling, 2018,33(6):36–52.

## Predictive Effect of High-frequency Investor Sentiment on the Intraday Stocks Return

YIN Hai-yuan, WU Xing-ying

(International Business School of Shaanxi Normal University, Xi'an 710119, China)

**Abstract:** The method using data mining to extract investor sentiment from network platform information not only increases the availability of high-frequency sentiment index, but also helps to deeply analyze the interaction between investor sentiment and stock market. The paper obtains the real-time posts of Shanghai Securities Composite Index, and constructs the intraday high-frequency sentiment indicators through text analysis, studying its forecasting effect on the stock market intraday returns. The study finds that the intraday investor sentiment can positively predict the stock market return, and the forecasting effect is more pronounced in the afternoon trading session. Although the forecasting effect of investor sentiment is independent from the intraday momentum effect to the stock return itself, it's less significant than the previous return and volatility. In the bull market, the investor sentiment predicts the intraday return more strongly than the lag return, while the bear market is opposite. In the extreme market environment such as skyrocketing or plunging, the influence of investor sentiment on intraday return is more significant than that of lagging return. The release of overnight investor sentiment will significantly affect the stock return of next morning, but there is a time lag. The investor sentiment during midday break can positively affect market performance in the afternoon along with the stock return in the morning. Further, noise trading is an important driving force for investor sentiment to influence stock return. The results are still robust after considering the effects of the month, the week and macroeconomic variables. These conclusions will help to understand the characteristics and mechanisms of sentiment effects in the stock market from a higher frequency perspective.

**Key Words:** high-frequency sentiment; stock return; intraday effect; text mining

**JEL Classification:** G12 G14 C32

[责任编辑:许明]