

基于股评的投资者情绪对股票市场的影响^①

部 慧¹, 解 峥¹, 李佳鸿¹, 吴俊杰^{1,2*}

(1. 北京航空航天大学经济管理学院, 北京 100191;

2. 北京航空航天大学大数据科学与脑机智能高精尖创新中心, 北京 100191)

摘要: 探讨投资者情绪对我国股票市场的影响. 为刻画投资者情绪, 基于东方财富网股吧帖文与朴素贝叶斯方法, 提出融合股评看涨看跌预期和投资者关注程度的投资者情绪度量指标. 进一步, 利用 Granger 因果检验、瞬时 Granger 因果检验、跨期回归分析等方法, 探讨了投资者情绪对我国股票收益率、交易量和波动性是否具有预测能力及影响. 实证结果揭示: 虽然投资者情绪对股票市场收益率、交易量和波动性均无预测能力, 但投资者情绪对股票收益率和交易量有当期影响; 开盘前非交易时段的股评情绪对开盘价具有预测力, 开盘后交易时段的股评情绪对收盘价和日交易量具有更显著的影响. 此外, 股票收益率是投资者情绪的 Granger 原因, 即投资者情绪的形成依赖于前期市场收益率. 这些实证结果为深入理解参与股吧评论的交易者的行为以及行为对市场产生的影响提供了证据.

关键词: 投资者情绪; 噪声交易者; 文本挖掘; Granger 因果检验

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2018)04-0086-16

0 引言

证券市场资产定价是金融核心问题, 不同理论对此有不同理解. 古典定价理论认为基本面信息是驱动资产定价和价格波动的最主要因素. 有效市场假说(EMH)认为如果金融市场是信息有效的, 那么交易资产的价格应该反映了市场基本面的所有已知信息; 即使市场上存在一些非理性投资者, 他们的需求也会被套利者所抵消从而对资产价格没有显著影响. 然而, 自20世纪80年代中期开始, 市场发生的一些异常现象令研究者开始认真思考投资者理性这一重要假设, 一系列研究开始探讨有效市场假设未必成立的各种可能性. 例如, Black^[1]、DeLong等^[2]提出的噪声交易者理论认为, 如果投资者根据一些与基本面无关的噪声信号进行交易, 那么资产价格将偏离其内在价值. 他们的研究中假定了市场存在着两类交易

者——基本面交易者和噪声交易者, 基本面交易者是有理性的, 并且他们对资产价值具有无偏预期; 而噪声交易者依赖于不同的信号进行交易, 对资产价值的预期有偏. 他们的工作说明了噪声交易者会引起价格偏离, 这种观点被越来越多的学者所接受.

在我国股票市场上, 投资者非理性的现象比较明显, 尤其是普通的中小投资者囿于信息渠道以及知识水平等的限制, 对资产的估值很难做到完全理性. 此外, 由于我国市场交易机制的不完善, 例如对卖空的限制, 以及机构投资者的不成熟等原因, 使得套利未必能快速实现. 因此, 我国证券市场的资产定价问题更值得深入探讨, 尤其是投资者情绪是否会影响资产价格等问题. 由于投资者情绪无法直接观测, 因而该研究议题存在实证上的困难. 如何刻画投资者情绪也成为此论题中的一个重要问题. Bu和Pi^[3]对学术文献中出现

① 收稿日期: 2015-10-13; 修订日期: 2017-10-13.

基金项目: 国家自然科学基金资助项目(71671012; 71373001; 71531001; 71725002; 71471009); 北京航空航天大学青年科学家团队项目.

通讯作者: 吴俊杰(1979—), 男, 广东平远人, 教授, 博士生导师. Email: wujj@buaa.edu.cn

的刻画投资者情绪的方法进行了归类,发现 Baker 和 Wurgler^[4]提出的利用几个市场变量合成投资者情绪指数的方法是学术研究中使用最多的,而利用网站、社交媒体、网络留言板等互联网平台数据挖掘投资者情绪的方法是近几年的研究热点。在当前“大数据”的背景下,如何更好地利用广泛的互联网平台信息挖掘投资者情绪乃至其交易行为,值得继续深入研究。

本文基于互联网平台大数据利用文本挖掘方法,在融合股评情绪和投资者关注的基础上提出了新的投资者情绪指标。为此,收集了东方财富网旗下股票社区“股吧”中的沪深300成分股的所有评论帖子,采用朴素贝叶斯模型对股评情绪进行分类计算,以获取投资者对市场的未来预期或交易倾向,最终提出新的投资者情绪指标。本文围绕“投资者情绪是否对股票市场收益率、交易量和波动性具有预测能力”这一核心问题,选择了沪深300指数作为股票市场整体的代理变量,分别采用 Granger 因果检验、Granger 瞬时因果检验、跨期回归分析(intertemporal regression analysis)等方法,研究了投资者情绪对股票市场的影响并进行了实证分析。其实证研究结果便于深入分析那些参与股吧讨论的投资者行为以及这些行为产生的结果。

1 文献综述

文献中对投资者情绪的定义并不完全相同,Bu 和 Pi^[3]指出各种定义中都包含了一些基本内涵:投资者情绪是对未来收益的预期,这种预期由于不完全是基于基本面分析得到,因而存在偏差甚至是错误。根据已有文献,构建投资者情绪指数的方法大致可以分为三类:

第一类,通过市场调查得到投资者对于市场未来走势的看法,以度量投资者情绪的水平,即直接法。例如美国市场上有两个被普遍接受和使用的投资者情绪的调查指数,美国个人投资者协会构建的调查 AAII 以及 II(investor intelligence)调查。II 调查通常被认为是机构投资者情绪指标(Lee 等^[5])。利用上述两个市场调查的情绪指数进行研究的工作包括 Brown 和 Cliff^[6],Schmeling^[7]等等。我国市场上也有类似的情绪调查,例

如央视看盘。市场调查是对投资者情绪的直观反映,但有学者指出,问卷回答具有一定的主观性,并且投资者在实际的投资行为中要比在问卷回答时更加理性。

第二类,采用可观测的经济变量来度量,即间接法。常用的指标包括封闭式基金折价率、IPO 发行数量或首日评论收益率、市场换手率等。在 Lee 等^[8]、Neal 和 Wheatley^[9]、Ritter^[10]等学者的工作之上,Baker 和 Wurgler^[11,12]陆续研究了一些经济指标,并在 Baker 和 Wurgler^[4]的工作中利用封闭式基金折价率、股票市场换手率、IPO 发行数量、IPO 首日平均收益率、股票占直接融资(股票和债券)的发行份额、分红溢价6个指标合成了投资者情绪指数。这个指数在研究中的应用十分广泛,尤其是对美国股票市场的研究。此外,Brown 和 Cliff^[6],Berger 和 Turtle^[13]等人的工作除了上述指标还考虑了与衍生品市场投资者交易的相关指标。

第三类,基于新闻网站、社交媒体、网络留言板等网络平台,利用文本挖掘的方法,提取和构建投资者情绪指标。Antweiler 和 Frank^[14]选择了雅虎财经和 RagingBull.com 的网络留言板或称为股吧,研究了45家大公司在2000年当中近150万条文本信息,挖掘并构建了投资者情绪指标。Das 和 Chen^[15]、Kim 和 Kim^[16]同样利用雅虎财经股吧评论信息来提取并构建了投资者情绪指标。Chen 等^[17]利用社交网络的帖子、Sprenger 等^[18]和 Bartov 等^[19]利用推特上发布的帖子构建了投资者情绪指标。此外,还有一些研究基于搜索引擎上股票代码的在线搜索密度,以此来表达投资者关注,如 Joseph 等^[20]、Da 等^[21]。

大量研究利用前两类投资者情绪的度量方法分析了投资者情绪是否会影响资产价格、波动以及交易量。Lee 等^[5]利用投资者情报调查 II 研究了噪声交易者对市场收益和波动率的影响。该研究认为投资者情绪是系统性风险,因此应该包含在定价之中。另外一些研究则认为市场收益与投资者情绪是相互作用的过程:市场收益较高时,投资者情绪高涨更加乐观,而当市场收益较低甚至是下跌时,投资者情绪低落更加悲观,而投资者情绪又进一步作用于市场收益。例如,Brown 和 Cliff^[6]采用 VAR 方法对美国市场上投资者情绪

与市场收益之间关系的研究表明,过去的市场收益是投资者情绪的一个重要影响因子.此外,很多学者还研究了投资者情绪对横截面上不同资产价格的影响差异,如 Lee 等^[8]、Brown 和 Cliff^[22]、Baker 和 Wurgler^[4]等.

我国对投资者情绪的研究也很多,不过大多利用单一变量作为代理变量或者利用 Baker 和 Wurgler^[4]的方法利用多个市场变量合成投资者情绪指数进行分析.伍燕然和韩立岩^[23]以封闭式基金折价率和换手率作为投资者情绪的代理变量探讨了投资者情绪对股市收益的影响.刘维奇和刘新新^[24]通过将个人和机构月新增开户数作为代理变量,研究了我国个人和机构投资者情绪的相互影响.张强和杨淑娥^[25]、文风华等^[26]基于多个较好反映国内股票市场投资者情绪的指标,利用主成分分析法构建投资者情绪综合指数,并研究了投资者情绪对市场收益和价格行为的影响.黄德龙等^[27]、宋泽芳和李元^[28]利用我国股票投资者情绪指数探讨了投资者情绪对我国股票市场横截面收益的影响.目前,我国市场上利用文本挖掘方法构建投资者情绪指标的研究仍很少,因此,本文的工作能对已有文献提供有力补充.

2 基于股评的投资者情绪指标

2.1 基于股评的投资者情绪指标

借鉴 Antweiler 和 Frank^[14]提出的基于股吧帖子分类构建看涨指标的方法,即

$$B_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{pos} + M_t^{neg}} \quad (1)$$

式中 $M_t^c = \sum_{i \in D(t)} w_i x_i^c$ 表示类型为 $c \in \{pos, neu, neg\}$ 的消息在一段时间 $D(t)$ 内的加权数量之和,其中 pos 代表积极情绪, neg 代表消极情绪, neu 代表中性情绪; x_i^c 是指示变量,如果消息 i 属于某种类型 c ,其值为 1,否则为 0;特别地,当权重全部等于 1 时, M_t^c 就等于 $D(t)$ 时段内类型为 c 的消息总量.股评看涨指标 B_t 介于 -1 和 1 之间,表达了投资者们相对看涨的程度,该指标与帖子总数量无关.此外, Antweiler 和 Frank^[14]还定义了另外一个指标

$$B_t^* = \ln \left[\frac{1 + M_t^{pos}}{1 + M_t^{neg}} \right] \quad (2)$$

并给出 $B_t^* \approx B_t \ln(1 + (M_t^{pos} + M_t^{neg}))$.他们认为 B_t^* 不仅考虑了相对看涨的程度,还考虑了表达看涨看跌观点的帖子数量.他们的研究显示 B_t^* 指标更优,因此他们在论文中主要报告了以 B_t^* 指标度量的结果.

Antweiler 和 Frank^[14]构建的股评看涨指标 B_t^* 虽然考虑了帖子数量,但是仅考虑了表达看涨看跌情绪的帖子数量,并未考虑中性帖子数量.本文认为投资者关注程度同样是投资者情绪的一种表现,因此即使投资者表达的是中性预期,这个信息也是有价值的.鉴于此,本文提出了融合基于股评的看涨指标和用帖子总数表达的投资者关注的股评情绪指数 (B_t^{Att})

$$B_t^{Att} = B_t \ln(1 + M_t) \quad (3)$$

其中 $M_t = M_t^{pos} + M_t^{neg} + M_t^{neu}$ 是以帖子总量表达的投资者关注程度.由于帖子数量可能很大,因此这使得股评情绪指数的变化范围也会很大.为了解决这个问题,将股评情绪剔除掉过去一年的滑动平均值并且用滑动平均值标准化,构建异常股评情绪

$$Abn(S_t) = \frac{S_t - \text{average}(S_{(t-k:t-1)})}{\text{average}(S_{(t-k:t-1)})} \quad (4)$$

其中 S_t 表示投资者情绪,可以是式(1)至式(3)定义的任一基于股评的投资者情绪指标; $\text{average}(S_{(t-k:t-1)})$ 表示过去一年的滑动平均值,若是月度数据,那么 $k = 12$;若是周度数据,那么 $k = 52$;若是日度数据,那么 $k = 252$.

本文在月度、周度和日度频率上对比分析了上述几种投资者情绪的度量方法的异同,并且比较了不同情绪指标在研究投资者情绪对股票价格预测能力上的细微差别.由于投资者情绪指标的计算需要知道 M_t^c ,因此下文将聚焦于构建文本情绪分类模型.值得说明的是,本文的“投资者情绪”指投资者对未来走势的预期或市场交易倾向.例如,积极预期指看涨或买入倾向,消极预期指看跌或卖出倾向,中性预期指市场持平或无明显交易倾向.这与一般意义上的文本情绪分类^[29]不同,后者关注的是发帖人的心理情感状态;如股评“明天股市将大跌,我已经在高点全部卖出,哈哈!”反映了发帖人的高兴情绪,但从投资者预期来看却是消极看跌的.

2.2 基于朴素贝叶斯的股评预期分类模型

文本的情绪分类通常有基于情感词典和基于

分类器的两种方法^[30]. 中文股评通常长短不一、文法自由, 适用的情感词典极为匮乏^[31], 因此本文采用基于分类器的建模方法. 在具体分类器的选择上, 采用基于经验分布的朴素贝叶斯模型 (Tan^[32]). 该模型简单、鲁棒, 在海量股评数据上仍有极高的训练及使用效率. 部分研究如 Zhao 等^[29]表明, 在高维稀疏短文本如微博的情绪计算上, 简单模型如朴素贝叶斯的精度通常会高于诸如支持向量机的复杂模型.

给定一条股评 d , 朴素贝叶斯模型通过先验概率和类条件概率估计 d 属于任意类别 c_j 的后验概率. 具体而言, 首先对股评进行分词并做词性筛选, 只保留信息量丰富的动词、名词、形容词作为特征, 表示为 $d = (w_1, \dots, w_i, \dots, w_n)$, 其中 w_i 表示第 i 个非空特征值, 其在 c_j 类的经验分布如下

$$P(w_i | c_j) = \frac{TF(w_i, c_j) + 1}{\sum_q (TF(w_i, c_q) + 1)} \quad (5)$$

式中 $TF(w_i, c_j)$ 表示特征 w_i 在 c_j 类中出现的次数, 分母采用了拉普拉斯平滑以避免分母为零的情况. 因此 d 属于第 c_j 类的概率可估计为

$$P(c_j | d) = P(c_j) P(d | c_j) / P(d) \propto P(c_j) \prod_{i=1}^n P(w_i | c_j) \quad (6)$$

因此股评 d 的最终类别为

$$c_d = \arg \max_{c_j \in C} \left\{ P(c_j) \prod_{i=1}^n P(w_i | c_j) \right\} \quad (7)$$

本文的实验数据采自东方财富网旗下股票社区一股吧 (<http://guba.com.cn>). 利用 Scrapy 爬虫爬取了沪深 300 指数的历年成分股从 2009 年 1 月 1 日至 2014 年 10 月 31 日期间共约 1 800 万条主帖数据. 每条帖子共包含 12 个字段: 股票代码、用户名、股吧用户主页链接地址、是否是股吧认证用户、股票名称、帖子内容、帖子类型 (新闻、数据、

研报、公告、股东、None)、回帖数量、帖子标题、发帖用户 id、点击量、发帖时间. 为准备训练数据集, 从数据库中随机抽取经垃圾过滤的 5 000 条股评数据. 邀请 10 位金融从业人员进行情绪标注, 每人标注 1 000 条, 这样确保每条股评有两位人员标注; 对于标注结果相异的股评, 请第三位金融从业人员标注, 取三次标注结果中占多数标签作为该股评最终标签. 最终得到各类样本数量如下: 积极 - 1 586, 消极 - 1 765, 中性 - 1 649. 建立三类分类模型, 采用五折交叉检验, 如表 1 所示. 结果显示, 三类情绪的预测精度均达到了较高的 $f = 0.75$ 左右; 其中积极类的分类准确率较高, 而消极类的召回率较高, 说明有部分积极类样本被分到了消极类, 这与后者的训练样本数量相对占优应有一定关系, 但总体影响不大. 国际上已有部分文献采用朴素贝叶斯方法计算股评情绪. 例如, Das 和 Chen^[15]使用朴素贝叶斯方法对雅虎股评数据进行情绪分类, 测试准确率为 50%. Kim 和 Kim^[16]则在假设中性情感股评已经全部剔除情况下, 基于朴素贝叶斯模型将股评数据分为积极情绪和消极情绪两类, 分类召回率为 62.7%. Antweiler 和 Frank^[14]将股评数据分为积极情绪、消极情绪、中立三类, 样本测试结果显示积极情绪和消极情绪平均召回率为 72.3%. 与上述研究相比, 本文使用的分类模型在精度上是可比的甚至更高的. 值得指出的是, 上述文献的分析对象都是英文股评, 考虑到中文分词精度较低, 本文的分类模型精度总体而言是令人满意的.

另外, 还需注意到上述精度是针对单条股评而言的, 而在本文后续实证研究中考察的最细粒度是单只股票. 这意味着要汇总这只股票在某一时段内的所有股评情绪, 其实对单条股评的分类精度要求反而下降了. 为了更好地理解这一点, 假设一只股票每日股评有 n 条, 通过投票机制汇总

表 1 贝叶斯分类模型测试结果

Table 1 The results of Naïve Bayes classification algorithm

交叉检验 编号	消极类			中性类			积极类		
	$P(\%)$	$r(\%)$	$f(\%)$	$p(\%)$	$r(\%)$	$f(\%)$	$p(\%)$	$r(\%)$	$f(\%)$
1	67.37	89.33	76.81	79.88	75.84	77.81	85.52	63.92	73.16
2	70.37	85.93	77.38	81.52	76.92	79.16	82.32	68.53	74.79
3	63.11	84.15	72.13	80.66	72.64	76.44	80.92	63.73	71.30
4	69.51	83.78	75.98	76.54	75.27	75.90	87.14	69.71	77.46
5	71.19	90.32	79.62	83.03	78.29	80.59	86.36	66.28	75.00
平均值	68.31	86.70	76.38	80.33	75.79	77.98	84.45	66.43	74.34

注: p - 准确率, r - 召回率, $f = 2pr / (p + r)$, 三者均为正指标.

每条股评的情绪来确定该只股票当天的情绪极性,这就意味着只需分对 $a = \lceil (n+1)/2 \rceil$ 条股评的情绪极性即可. 假设单条股评的分类准确率为 f , 那么该只股票情绪的分对概率为 $P = \sum_{i=a}^n \binom{n}{i} f^i (1-f)^{n-i} \approx I_f(a, n+1-a)$, 其中 I 为不完全 β 函数 (incomplete Beta function) ^[33, 34]. 以 $n=30$ 为例, 当 $f=0.6$ 时, $P=82.5\%$; 而当 $f=0.7$ 时 $P=98.3\%$. 由此可见, 本文构建的朴素贝叶斯分类模型能够很好地支持股票的情绪计算.

3 数据和研究方法

3.1 数据和统计描述

本文选择了沪深 300 指数作为对股票市场整体的刻画, 选取 2009 年 1 月 1 日至 2014 年 10 月 31 日的月度、周度和日度频率的数据作为研究样本. 为了与沪深 300 指数收盘价相对应, 计算股评

情绪时 $D(t)$ 时段以每日收盘时刻、每周最后一天收盘时刻、每月最后一天收盘时刻为隔断点, 汇总计算不同频率时间段里面的股评帖子. 观察帖子数量, 发现中性帖子数量在 2013 年 4 月 1 日和 2013 年 5 月 2 日出现异常点, 为了避免数据异常带来的问题, 剔除了异常点, 并且用前 5 日的平均值替换异常数据.

图 1 描绘了沪深 300 指数和沪深 300 所有成分股的股评总量随时间变化的情况; 图 2 描绘了本文新提出的股评情绪 B_t^{Att} 和剔除了均值水平的异常情绪 $Abn(B_t^{Att})$ ^②. 无论在何频率上, 股吧中消极 (看跌) 的帖子永远都比积极 (看涨) 的帖子多, 这使基于股评看涨指标 B_t 、 B_t^* 和股评情绪指数 B_t^{Att} 计算得到的投资者情绪均为负值. 这说明东方财富网股吧中的股评存在负向的系统偏差, 这从一个侧面揭示了在股吧中发表帖子的人更喜欢发表看跌的负面评论, 也说明散户投资者更容易产生消极情绪并且更愿意把消极情绪表达在网络平台中.

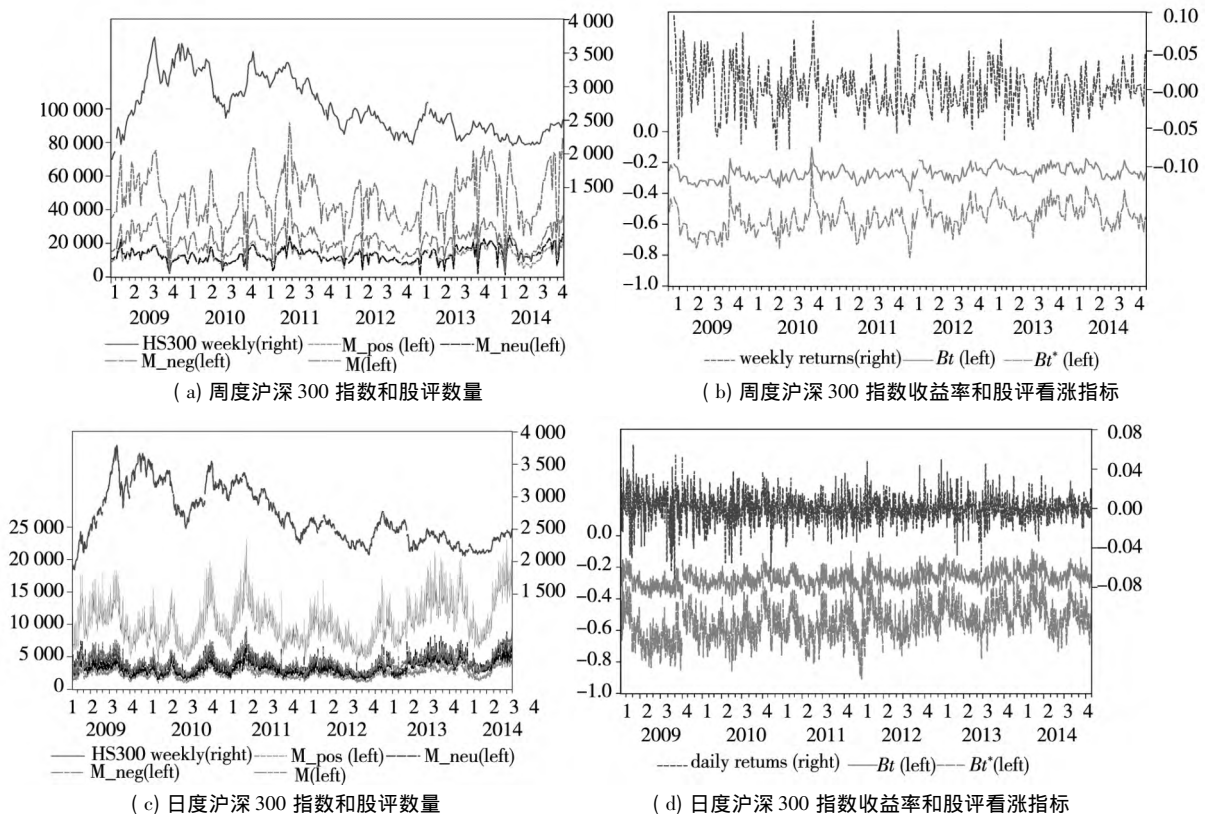


图 1 沪深 300 指数及收益率与股评数量及股评看涨指标

Fig. 1 CSI300 index and its returns, message number, and message bullishness indicators

② 为了节省空间, 图 1 和图 2 省略了月度结果.

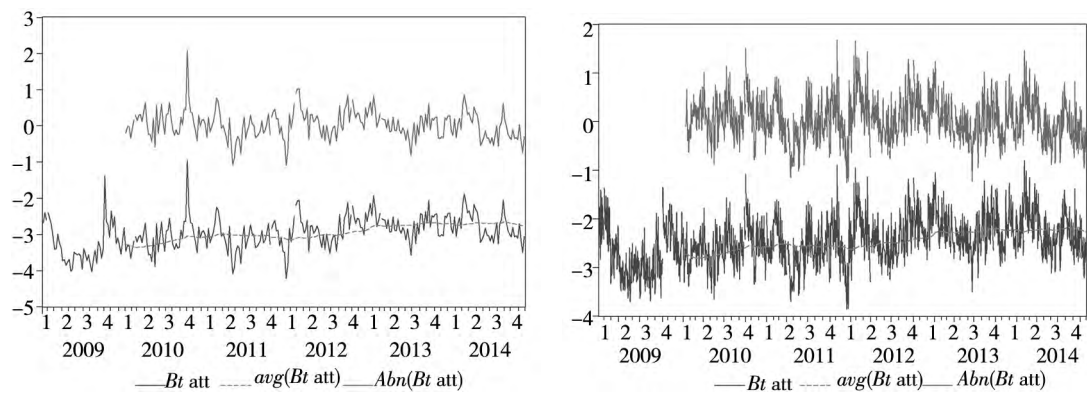


图 2 股评情绪 B_t^{Att} 和异常股评情绪 $Abn(B_t^{Att})$

Fig. 2 Investor sentiment extracted from messages B_t^{Att} and $Abn(B_t^{Att})$

表 2 沪深 300 指数与投资者情绪指标的描述统计

Table 2 Descriptive statistics of CSI300 index and investor sentiment indicators

	收益率	B_t	B_t^{Att}	收益率	B_t	B_t^{Att}	收益率	B_t	B_t^{Att}	交易量
样本数	70	70	70	297	297		1 412	1 412	1 412	1 412
均值	0.004 6	-0.265 9	-3.243 4	0.000 9	-0.270 8	-2.964 2	0.000 2	-0.268 6	-2.462 4	6.791 9
中位数	0.003 1	-0.2650	-3.239 3	0.001 4	-0.271 3	-2.976 4	0.000 4	-0.271 4	-2.484 1	6.133 2
最大值	0.165 0	-0.179 0	-2.212 4	0.095 9	-0.101 1	-0.953 4	0.064 6	-0.085 1	-0.803 8	23.917
最小值	-0.277 3	-0.339 9	-4.152 3	-0.091 0	-0.386 9	-4.218 1	-0.073 7	-0.426 1	-3.857 1	2.190 1
标准偏差	0.077 0	0.036 5	0.457 8	0.031 5	0.041 5	0.463 3	0.014 9	0.053 8	0.491 2	2.933 5
偏度	-0.454 8	-0.000 6	-0.101 2	0.125 2	0.249 0	0.211 9	-0.290 5	0.244 0	0.152 0	1.174 7
峰度	4.600 0	2.390 2	2.420 9	3.067 9	3.334 5	3.922 1	5.190 1	2.982 8	3.007 9	4.987 6
JB 统计量	9.879 6 [0.007 2]	1.084 7 [0.581 4]	1.097 5 [0.577 7]	0.833 5 [0.659 2]	4.454 9 [0.107 8]	12.787 [0.001 7]	302.05 [0.000 0]	14.031 [0.000 9]	5.442 5 [0.065 8]	557.56 [0.000 0]
ADF 检验	-8.532 9 [0.000 0]	-4.134 6 [0.001 6]	-4.304 2 [0.000 9]	-16.908 1 [0.000 0]	-6.012 3 [0.000 0]	-5.927 7 [0.000 0]	-36.835 [0.000 0]	-5.238 7 [0.000 0]	-5.318 8 [0.000 0]	-5.782 1 [0.000 0]

注： B_t^* 与 B_t 的统计性质相似，为了节省空间，本表并未论述。

本文分析了沪深 300 指数与股评数量的相关关系以及市场收益率和基于股评的投资者情绪指标的相关关系^③，发现沪深 300 指数与消极看跌的帖子数量相关性最高，达到 0.159 9，与积极看涨的帖子数量相关性只有 0.062 0 左右，与中性持平的帖子数量不相关，而与全部帖子总数的相关性为 0.069 9。从收益率上来，几个基于股评的投资者情绪指标均与市场收益率有较高的相关性，在 0.22 ~ 0.24 之间，其中股评情绪指数 B_t^{Att} 与收益率的相关性最高，达到 0.237 3。三种股评情绪的度量方法相关系数均在 0.98 以上，而股评情绪与其去掉趋势的异常情绪的相关系数均在 0.91 以上。这说明三种不同的基于股评的投资者

情绪的度量方法主要受股评看涨指标 B_t 影响。

接下来，分析了沪深 300 指数收益率、交易量和股评看涨指标 B_t 、 B_t^* 和股评情绪指数 B_t^{Att} 的统计性质，如表 2 所示^④。对股票收益率而言，周度收益率呈正态分布，但月度和日度收益率不是正态分布；对投资者情绪指标而言，其月度和周度频率是正态分布的，但日度指标不是正态分布；另外日度的股票交易量不是正态分布。ADF 单位根检验结果显示，各个频率的股票收益率、投资者情绪指标以及日度股票交易量均是平稳序列。

3.2 研究方法

3.2.1 对收益率影响的 Granger 因果检验

首先，选取 Granger 因果检验方法来探讨投

③ 为了节省空间，未说明相关分析结果。
④ 为了节省空间，表 3 中只说明了部分结果。

投资者情绪与股票收益率的相互关系,并重点探讨投资者情绪对股票收益率在时间序列上的预测能力

$$R_t = \alpha_1 + \sum_{i=1}^m \beta_{1,i} R_{t-i} + \sum_{j=1}^n \gamma_{1,j} S_{t-j} + \varepsilon_{1,t} \quad (8)$$

$$S_t = \alpha_2 + \sum_{i=1}^m \beta_{2,i} R_{t-i} + \sum_{j=1}^n \gamma_{2,j} S_{t-j} + \varepsilon_{2,t} \quad (9)$$

其中 S_t 表示投资者情绪在时刻 t 的值(可用 B_t 、 B_t^* 、 B_t^{Alt} 和异常情绪指标 $Abn(S_t)$ 表达), R_t 是沪深 300 指数在时刻 t 的收盘价的对数收益率,即 $R_t = \ln(P_t) - \ln(P_{t-1})$ 。为了使模型设定更合理,需要 S_t 和 R_t 平稳,这是因为 Granger^[35] 在重新回顾其因果性定义时,指出“非平稳变量带来的问题过于复杂”。目前在进行实证研究时,一般认为只有平稳变量才能应用 F 统计量进行推断,否则结论可能是不可靠的。因此,对于上述模型的设定和估计方法,采用部慧等^[36] 中的具体方法。模型滞后期设定的基本原则为最小 AIC 准则,同时考虑了消除残差序列自相关的必要条件,并兼顾 SC 准则以令模型设定尽可能简单。模型残差的序列相关性用 LM (Lagrange Multiplier) 检验。如果存在序列相关,则增加解释变量的滞后期直到序列相关被消除。此外,本文还使用 White 检验方法检验了残差的异方差问题。如果模型残差存在异方差,那么会利用 White 的异方差一致协方差估计量进行修正。利用上述方程估计结果进行两个假设检验。其一,对式(8)检验原假设为 $H_{0,1}: \gamma_{1,j} = 0, j = 1, 2, \dots, n$, 即“投资者情绪不是股票收益率的 Granger 原因”;其二,对式(9)检验原假设为 $H_{0,2}: \beta_{2,j} = 0, j = 1, 2, \dots, n$, 即“股票收益率不是

投资者情绪的 Granger 原因”。假设检验以 Wald 卡方统计量来判断。

为了进一步检验投资者情绪对股票收益率的影响,将式(8)的 Granger 因果检验拓展为瞬时 Granger 因果检验

$$R_t = \alpha_3 + \sum_{i=1}^m \beta_{3,i} R_{t-i} + \gamma_{3,0} S_t + \sum_{j=1}^n \gamma_{3,j} S_{t-j} + \varepsilon_{3,t} \quad (10)$$

式(10)的原假设 $H_{0,3}: \gamma_{3,j} = 0, j = 0, 1, 2, \dots, n$ 。

实证研究中除了对月度、周度、日度频率的数据进行研究外,对于日度频率,本文还将日度投资者情绪的计算时段拆分为两段,即交易时段(当日开盘至当日收盘时)和非交易时段(前一交易日收盘后至当日交易日开盘前);然后分别计算两个时段的投资者情绪指标,并分别与当日收益率匹配,以区分全日时段中到底是哪一段的投资者情绪信息更有价值。这样处理,是为了剔除由于人为区隔数据引起的偏差。

3.2.2 对交易量和波动率影响的模型

接下来,探讨参与股吧评论的投资者是否为噪声交易者以及他们对市场的影响。根据 Black^[1]、Delong 等^[2] 的噪声交易者理论,噪声交易者会对未来价格产生错误预期,并将根据这种自己对市场看涨看跌的错误感知进行交易,从而对股票价格和股价波动率均产生影响。根据该理论,本文还构建了股评情绪对股票交易量和股价波动率的影响模型,如式(11)和式(12)所示。

$$V_t = a_0 + \sum_{i=1}^m a_i V_{t-i} + b_1 R_{t-1} + c_1 DisAgree_t + c_2 DisAgree_{t-1} + \varepsilon_t \quad (11)$$

$$Volatility_t = b_0 + b_1 Volatility_{t-1} + b_2 R_{t-1} + b_3 DisAgree_t + b_4 DisAgree_{t-1} + \varepsilon_t \quad (12)$$

其中 V_t 表示沪深 300 成分股票的成交量(单位为 10 亿股),模型中使用了投资者情绪的另外一个指标($DisAgree_t$)以衡量投资者情绪的分散程度, $DisAgree_t = 1 - |B_t|$ 。该指标的定义与 Kim^[16] 和 Kim^[16] 以及 Das 和 Chen^[15] 的定义一致。当投资者都普遍看涨或看跌,那么 $DisAgree_t$ 更接近于 0;而当投资者看涨与看跌比较均衡,那么 $DisAgree_t$ 更接近于 1。利用式(11),分别检验只引入投资者情绪分散程度滞后一期以及同时引入投资者情绪分散程度当期和滞后一期的结果。该模

型还控制了交易量的滞后变量以及收益率的滞后一期变量。这个模型与 Kim 和 Kim^[16] 讨论交易量时的模型略有差异。其一,使用了收益率滞后变量而非收益率绝对值的滞后变量,做这种设定是因为我国股票市场不能或限制卖空,这使得很多投资者只能采用买入持有的策略,这意味着只有价格上涨才会使得交易活跃;其二,方程中还检验了同时引入投资者情绪分散程度的当期和滞后一期的影响,而 Kim 和 Kim^[16] 只关注投资者情绪对交易量的可预测性,所以只引入了投资者情绪分散

程度的滞后一期变量. 式(11)中控制变量交易量的滞后变量阶数 m 的选择将根据 AIC 准则做出.

为了讨论投资者情绪对市场波动性的影响, 建立如式(12)所示的模型. 其中, 市场波动率采用了最简单的度量方法, 即日度收益率的标准差^⑤. 分别检验了只引入投资者情绪分散程度滞后变量或当期变量, 以及同时引入投资者情绪的当期变量和滞后变量的结果. 方程中控制了波动率滞后一期变量以及收益率的滞后一期变量, 这是因为大量结果揭示了波动率具有持续性, 并且与前期收益率相关^[17]. 投资者情绪对股票交易量和波动率影响的讨论聚焦于对日度频率数据的分析上.

4 实证结果及分析

4.1 对股票收益率影响的结果分析

4.1.1 对收益率的因果检验结果

投资者情绪与股票市场收益率的 Granger 因果检验结果如表3所示, 不同投资者情绪指标的结果高度相似^⑥. 表3中A栏至C栏的结果揭示, 不同频率下投资者情绪均不是股票收益率的 Granger 原因; 表3中D栏至F栏的结果揭示, 股票收益率是投资者情绪的 Granger 原因, 虽然不同频率下结果有略微差异. 月度和周度的股票收益率对投资者情绪的影响为负, 而日度股票收益率对投资者情绪的影响为正. 同时发现基于股评的投资者情绪指标均有较强的自相关, 这意味着投资者预期会延续一段时间. 表4中论述了投资者情绪对收益率的瞬时 Granger 因果检验结果, A栏至C栏的结果揭示三种不同频率下投资者情绪均是股票收益率的瞬时 Granger 原因, 投资者情绪对股票收益率的累积影响为正.

对比三个基于股评的投资者情绪指标, 发现综合考虑了股评看涨指标和股评帖子总数的 B_t^{All} 在周度和日度上对收益率的解释效果均好于其他两个股评指标 B_t 和 B_t^* ; 而后两个指标在各个频率上的表现都非常相似. 这从一个侧面说明, 以发帖总数反映的投资者关注程度对股票收益率是有

影响的. 而本文基于此修正的投资者情绪指标更具价值. 此外, 基于股评的异常情绪指标得到的结果与表4中结果相似, 为节省空间本文并未论述.

具体观察式(10)的实证结果(表4), 发现收益率的自身滞后变量并不显著, 现将实证结果揭示的方程列示如下

$$\hat{R}_t = c + c_1 S_t + c_2 S_{t-1} \quad (13)$$

实证结果揭示了 $c_1 > 0$ $c_2 < 0$, 并且均在1%的显著性水平下显著; 并且投资者情绪指标 S_t 和 S_{t-1} 系数估计结果的绝对值非常接近, 为此推测对收益率有影响的是投资者情绪的变化, 即 $\Delta S_t = S_t - S_{t-1}$, 因此投资者情绪和股票收益率之间的关系应该是

$$\hat{R}_t = c + c_1 \Delta S_t \quad (14)$$

利用三个日度投资者情绪指标构建的投资者情绪变化 ΔS_t 对日度收益率重新估计了方程, 方程结果与表3中C栏日度频率的估计结果非常相似. 这里只论述以 B_t^{All} 的结果

$$\hat{R}_t = 0.0002 + 0.0083 \Delta B_t^{All} \quad (15)$$

(0.53) (9.37) Adj. $R^2 = 0.0580$

结果显示投资者情绪变化 ΔB_t^{All} 在1%的显著性水平下显著, 投资者情绪的变化对股票收益率有正向影响.

4.1.2 投资者情绪对日内价格的影响

为了更深入地研究基于股评的投资者情绪对股票价格的影响, 将日度投资者情绪的计算时段拆分成两段, 即开盘前非交易时段(前一交易日收盘 15:00PM 后至当日交易日开盘 9:00AM 前)和开盘后交易时段(当日开盘 9:00AM 至当日收盘 15:00PM 时), 两个时段计算的投资者情绪分别记为 $B_{t \text{ pre-open}}^{all}$ 和 $B_{t \text{ trading}}^{all}$. 为了对应, 每日收益率差分为开盘收益率 $R_t^{open} = \ln(P_t^{open}) - \ln(P_{t-1}^{close})$ 和交易日收益率 $R_t^{trading} = \ln(P_t^{close}) - \ln(P_t^{open})$. 利用上述指标估计如式(10)所示的瞬时 Granger 因果检验, 以讨论投资者情绪对股票开盘价和收盘价的影响. 结果如表5所示. 表5中A栏结果显示, 开盘前非交易时段的情绪 $B_{t \text{ pre-open}}^{all}$

⑤ 实证中, 采用了历史22天, 即接近一个月的日度收益率数据计算.

⑥ 若使用剔除趋势的异常情绪指标 $Abn(S_t)$, 结果与不剔除趋势的股评情绪指标相似. 为了节省空间, 表3只列了三个指标的检验结果.

对开盘收益率 R_t^{open} 有正向显著影响,甚至在控制了更早时间段的投资者情绪和前期的收益率,开盘前非交易时段的情绪对开盘价仍有显著影响.本文认为开盘前形成的市场情绪尤其是当日开盘前的非交易时段的情绪 $B_{t,pre-open}^{att}$ 对当日开盘价具有正向的预测能力.表 5 中 B 栏结果显示,开盘前各时段的投资者情绪均对当日收盘价无预测能力,开盘后交易时段的投资者情绪对当日的股票市场收盘价有显著的正向影响.对比不同投资者情绪的指标,结果非常相似,其中本文提出的新指标 B_t^{att} 的结果更优.

4.2 对股票交易量和波动性的影响分析

分析投资者情绪对股票市场交易量影响的方程估计结果(表 6 中 A 栏所示),本文发现若在模型中只引入 $DisAgree_t$ 或 $DisAgree_{t-1}$,投资者情绪分散程度指标均不显著;但当同时引入 $DisAgree_t$ 和 $DisAgree_{t-1}$,投资者情绪分散程度指标对交易量的影响增强,其中全日数据的结果在 10% 的水平下显著,日交易时段数据的结果在 5% 的显著性水平下显著,只有日非交易时段数据的结果不显著.若方程中分别引入 $DisAgree_{pre-open,t}$ 和 $DisAgree_{trading,t-1}$,则投资者情绪指标系数显著.对比最后两列的结果,本文认为是因为方程中引入了 $DisAgree_{trading,t-1}$,才改善了投资者情绪指标的显著性.因此认为日交易时段的投资者情绪指标对交易量的解释能力最为显著,实证结果如式(16)所示,该结果中所有系数除了截距项均显著.另外,观察到方程中 $DisAgree_t$ 或 $DisAgree_{t-1}$ 的系数绝对值非常近似,本文推测对股票交易量有解释能力的变量可能是投资者情绪的变化,即

$$\Delta DisAgree_t = DisAgree_t - DisAgree_{t-1}$$

因此,利用投资者情绪的变化重新估计了式

(11),估计结果如式(17)所示.

接下来,检验了实证结果的稳健性.首先,采用不同的投资者情绪得到的结果非常类似^⑦.其次,检验了不同控制变量对实证结果的影响.例如,若方程中控制前期收益率的绝对值 $|R_{t-1}|$,如 Kim 和 Kim^[16] 的模型,结果显示投资者情绪指标仍没有预测能力,而投资者情绪的变化仍显著为正, $|R_{t-1}|$ 在模型中的估计结果全部不显著,只有交易量的滞后变量全部显著^⑧.基于这个结果,本文认为之前设定的式(11)更为合理.从上述结果看,投资者情绪对股票市场交易量无预测能力,投资者情绪对同期的股票交易量具有显著正向影响;同时,投资者情绪的变化对同期的股票交易量具有显著的正向影响,并且日交易时段形成的投资者情绪指标最为显著.另外,实证结果揭示了式(11)中的控制变量全部在 1% 的显著性水平下显著.具体而言,交易量三期的滞后变量均显著为正,这说明股票市场的交易量有一定的惯性,市场交易活跃具有一定持续性.滞后一期的收益率在模型中显著为正,这说明市场上存在很多交易者根据市场的价格趋势进行交易.

表 6 中 B 栏汇报了投资者情绪对股票收益率波动性影响的实证结果.无论选择哪个时间段数据构建的投资者情绪指标,投资者情绪对股票收益率的波动性均无预测能力,也没有当期影响.式(13)的估计结果显示,只有两个控制变量波动率的滞后一期和收益率的滞后一期在 1% 的显著性水平下显著.波动率滞后一期变量的系数为正,并且系数数值近似于 1,说明收益率波动性具有很强的持续性.收益率滞后一期变量的系数为负,说明前期收益率若为正,那么接下来的市场波动性较低,若前期收益率为负,则接下来的市场波动性较高.

$$\hat{V}_t = 0.677\ 3 + 0.579\ 0V_{t-1} + 0.195\ 7V_{t-2} + 0.114\ 8V_{t-3} + 33.536\ 0R_{t-1} + 2.484\ 1DisAgree_{trading,t} - 2.364\ 5DisAgree_{trading,t-1} \quad (16)$$

(0.92) (16.42) (5.58) (3.59) (11.13) (2.31) (-2.40) Adj. $R^2 = 0.750\ 5$

$$\hat{V}_t = 0.760\ 5 + 0.579\ 2V_{t-1} + 0.195\ 7V_{t-2} + 0.114\ 6V_{t-3} + 33.627R_{t-1} + 2.427\ 4\Delta DisAgree_{trading,t} \quad (17)$$

(6.44) (16.45) (5.58) (3.59) (11.27) (2.73) Adj. $R^2 = 0.750\ 7$

⑦ 为了节省空间,论文中并未论述该结果.

⑧ 为了节省空间,论文中并未论述该结果.

表3 投资者情绪与股票收益率的 Granger 因果检验结果

Table 3 Granger causality test results of investor sentiment and stock returns

式(8): $R_t = \alpha_1 + \sum_{i=1}^m \beta_{1,i} R_{t-i} + \sum_{j=1}^n \gamma_{1,j} S_{t-j} + \varepsilon_{1,t}$											
A: 月度频率				B: 周度频率				C: 日度频率			
	B_t	B_t^*	B_t^{4u}		B_t	B_t^*	B_t^{4u}		B_t	B_t^*	B_t^{4u}
α_1	-0.037 5 (-0.52)	-0.035 9 (-0.52)	-0.014 1 (-0.20)	α_1	0.007 7 (0.54)	0.007 0 (0.52)	0.015 0 (1.10)	α_1	-0.002 1 (-1.06)	-0.002 1 (-1.05)	-0.001 4 (-0.68)
R_{t-1}	-0.006 0 (-0.05)	-0.005 8 (-0.05)	-0.018 0 (-0.14)	R_{t-5}	-0.122 1 (-2.19) **	-0.122 0 (-2.19) **	-0.119 5 (-2.15) **	R_{t-1}	0.026 0 (0.90)	0.026 0 (0.90)	0.024 3 (0.84)
S_{t-1}	-0.152 5 (-0.57)	0.071 4 (-0.58)	-0.005 3 (-0.25)	R_{t-6}	0.140 8 (2.32) **	0.140 7 (2.32) **	0.146 6 (2.40) ***	S_{t-1}	-0.008 6 (-1.12)	-0.004 0 (-1.12)	-0.000 6 (-0.76)
				S_{t-1}	0.026 4 (0.51)	0.011 7 (0.49)	0.004 9 (1.08)				
调整 R^2	-0.024 4	-0.024 3	-0.028 5	调整 R^2	0.026 9	0.026 8	0.031 1	调整 R^2	-0.000 074	-0.000 065	-0.000 57
DW 统计 量	2.032 8	2.033 3	2.030 4	DW	1.938 2	1.937 2	1.969 1	DW	1.996 2	1.996 1	1.996 5
卡方 统计 量	0.325 2	0.332 0	0.061 3	卡方	0.263 3	0.244 2	1.156 3	卡方	1.262 5	1.261 8	0.582 7
[p 值]	[0.568 5]	[0.564 5]	[0.804 4]	[p 值]	[0.607 8]	[0.621 2]	[0.282 2]	[p 值]	[0.261 2]	[0.261 3]	[0.445 2]
式(9): $S_t = \alpha_2 + \sum_{i=1}^m \beta_{2,i} R_{t-i} + \sum_{j=1}^n \gamma_{2,j} S_{t-j} + \varepsilon_{2,t}$											
D: 月度频率				E: 周度频率				F: 日度频率			
	B_t	B_t^*	B_t^{4u}		B_t	B_t^*	B_t^{4u}		B_t	B_t^*	B_t^{4u}
α_2	-0.090 9 (-3.67) **	-0.185 8 (-3.67) ***	-1.185 9 (-3.95) ***	α_2	-0.072 3 (-5.20) ***	-0.147 6 (-5.17) ***	-0.686 1 (-4.42) ***	α_2	-0.032 8 (-4.64) ***	-0.067 7 (-4.66) ***	-0.318 9 (-5.01) ***
R_{t-1}	-0.100 3 (-2.28) **	-0.218 0 (-2.31) **	-1.588 2 (-2.91) ***	R_{t-1}	0.121 2 (1.97) **	0.259 7 (1.96) *	-0.679 0 (-0.96)	R_{t-1}	1.057 4 (15.18) ***	2.290 3 (15.22) ***	9.314 9 (14.89) ***
R_{t-2}	-0.149 7 (3.59) ***	-0.323 9 (-3.62) ***	-1.956 2 (-3.77) ***	R_{t-2}	-0.159 5 (-2.66) ***	-0.347 4 (-2.69) ***	-2.157 5 (-3.12) ***	S_{t-1}	0.252 8 (11.00) ***	0.254 6 (11.07) ***	0.300 8 (12.74) ***
S_{t-1}	0.657 5 (7.15) ***	0.658 6 (7.18) ***	0.633 5 (6.92) ***	S_{t-1}	0.494 0 (8.04) ***	0.494 2 (8.03) ***	0.500 3 (8.08) ***	S_{t-2}	0.129 3 (5.34) ***	0.130 3 (5.35) ***	0.133 4 (5.36) ***
				S_{t-2}	0.239 9 (4.04) ***	0.241 3 (4.06) ***	0.268 6 (4.38) ***	S_{t-5}	0.232 0 (9.43) ***	0.229 0 (9.34) ***	0.204 0 (8.33) ***
								S_{t-10}	0.166 5 (7.18) ***	0.167 0 (7.24) ***	0.146 5 (6.49) ***
								S_{t-15}	0.098 2 (4.42) ***	0.097 7 (4.44) ***	0.087 1 (3.98) ***
调整 R^2	0.474 0	0.477 2	0.483 0	调整 R^2	0.453 8	0.455 4	0.441 1	调整 R^2	0.539 5	0.540 9	0.552 9
DW 统计 量	1.798 3	1.793 1	1.829 5	DW	2.061 5	2.029 2	2.073 6	DW	1.887 1	1.888 4	1.904 7
卡方 统计 量	17.589 9	17.891 3	22.34	卡方	11.213	11.306	10.871	卡方	230.30	231.68	221.80
[p 值]	[0.000 2]	[0.000 1]	[0.000 0]	[p 值]	[0.003 7]	[0.003 5]	[0.004 4]	[p 值]	[0.000 0]	[0.000 0]	[0.000 0]
影响	(-) ***	(-) ***	(-) ***	影响	(-) ***	(-) ***	(-) ***	影响	(+) ***	(+) ***	(+) ***

注: a. 本文分别论述了以 B_t 、 B_t^* 、 B_t^{All} 作为投资者情绪指标的估计结果, 每一列为一个方程的估计结果. 经检验, 所有频率下, 式(8)和式(9)的残差均不存在自相关. 但是, 周度频率的式(8)和日度频率的式(8)和式(9)残差序列存在异方差特征, 为此, 利用 White 的异方差一致协方差估计量对标准误和 t 统计量进行了修订. 表中列出的是修订后的结果.

b. 括号中是方程参数估计的 t 统计量. ***、**、* 分别表示系数在 1%、5%、10% 的水平下显著异于零.

c. 式(8)的原假设为 $H_{0.1}: \gamma_{1j} = 0 \quad j = 1, 2, \dots, n$; 式(2)的原假设为 $H_{0.2}: \beta_{2j} = 0 \quad j = 1, 2, \dots, n$. 对上述两个原假设以 Wald 系数检验的卡方统计量判别. 方括号中表示的是卡方统计量的 p 值.

d. “影响”表示的是收益率的所有滞后期对投资者情绪的累积影响, (+) 或 (-) 表示 $\sum \beta_{2j}$ 的符号. 其后的显著性表示的是对应的卡方统计量的显著性水平.

表 4 投资者情绪对股票收益率的瞬时 Granger 因果检验结果

Table 4 Granger instantaneous causality test results of investor sentiment to stock returns

A: 月度频率				B: 周度频率				C: 日度频率			
解释变量	B_t	B_t^*	B_t^{All}	解释变量	B_t	B_t^*	B_t^{All}	解释变量	B_t	B_t^*	B_t^{All}
α_3	0.067 0 (0.69)	0.063 4 (0.69)	0.081 0 (1.14)	α_3	0.042 8 (2.94) ***	0.040 7 (2.95) ***	0.053 2 (3.61) ***	α_3	0.009 3 (3.85) ***	0.008 9 (3.83)	0.010 1 (4.16) ***
R_{t-1}	0.098 1 (0.56)	0.099 8 (0.57)	0.111 2 (0.90)	R_{t-5}	-0.069 5 (-1.36)	-0.068 5 (-1.34)	-0.064 4 (-1.27)	R_{t-1}	-0.052 4 (-1.75) *	-0.052 9 (-1.77) *	-0.059 5 (-2.00) **
R_{t-2}				R_{t-6}	0.155 5 (2.70) ***	0.154 7 (2.69) ***	0.166 4 (2.92) ***				
S_t	1.144 6 (3.93) ***	0.533 3 (3.94) ***	0.084 7 (3.44) ***	S_t	0.362 8 (7.52) ***	0.169 4 (7.61) ***	0.034 0 (7.61) ***	S_t	0.086 9 (9.03) ***	0.040 5 (9.02) ***	0.010 6 (9.48) ***
S_{t-1}	-0.904 8 (-2.65) **	-0.422 9 (-2.66) ***	-0.060 8 (-2.37) **	S_{t-1}	-0.206 8 (-3.22) ***	-0.097 4 (-3.35) ***	-0.016 3 (-2.79) ***	S_{t-1}	-0.052 9 (-6.10) ***	-0.024 7 (-6.12) ***	-0.006 5 (-6.62) ***
调整 R^2	0.147 9	0.148 9	0.116 2	调整 R^2	0.164 2	0.166 0	0.188 0	调整 R^2	0.061 7	0.061 8	0.069 4
DW 统计量	2.000 1	2.001 9	2.015 2	DW 统计量	1.991 2	1.989 4	1.896 8	DW 统计量	2.005 8	2.005 6	2.006 3
卡方统计量	17.84	18.043 8	11.877	卡方统计量	58.394	59.626	61.719	卡方统计量	84.855	84.829	92.933
[p 值]	[0.000 1]	[0.000 1]	[0.002 6]	[p 值]	[0.000 0]	[0.000 0]	[0.000 0]	[p 值]	[0.000 0]	[0.000 0]	[0.000 0]
影响	(+) ***	(+) ***	(+) ***	影响	(+) ***	(+) ***	(+) ***	影响	(+) ***	(+) ***	(+) ***

注: a. 式(10) $R_t = \alpha_3 + \sum_{i=1}^m \beta_{3i} R_{t-i} + \gamma_{30} S_t + \sum_{j=1}^n \gamma_{3j} S_{t-j} + \varepsilon_{3t}$ 的原假设为 $H_{0.3}: \gamma_{3j} = 0 \quad j = 0, 1, 2, \dots, n$.

b. 所有结果均利用了 White 的异方差一致协方差估计量对标准误和 t 统计量进行了修订. 表中报告的是修订后的结果.

c. 表中其他说明同表 4 注释所示.

表 5 投资者情绪对股票开盘价和日内涨幅的 Granger 因果检验和瞬时 Granger 因果检验结果

Table 5 The results of Granger causality of investor sentiment to open price changes and trading day price changes

A: 日度开盘价涨幅 (R_t^{open})					B: 日内涨幅 ($R_t^{trading}$)				
c	0.002 4 (6.53) ***	0.000 9 (1.05)	-0.001 1 (-1.28)	-0.001 0 (-1.09)	c	-0.000 2 (-0.16)	-8.23E-05 (-0.03)	0.029 0 (11.10) ***	0.020 9 (7.19) ***
R_{t-1}^{open}	-0.059 9 (-1.88) *	-0.032 9 (-1.00)	-0.067 1 (-2.11) **	-0.053 3 (-1.63)	$R_{t-1}^{trading}$	-0.025 8 (-0.77)	-0.036 3 (-1.19)	0.010 5 (0.32)	0.020 2 (0.65)
$B_{pre-open,t}^{att}$	0.002 0 (8.059 1) ***		0.002 8 (9.24) ***	0.002 9 (9.29) ***	$B_{trading,t}^{att}$			0.014 4 (12.78) ***	0.016 6 (14.05) ***
$B_{trading,t-1}^{att}$		0.000 6 (1.85) *	-0.001 8 (-4.55) ***	-0.001 4 (-3.42) ***	$B_{pre-open,t}^{att}$	-0.000 7 (-1.01)		-0.006 4 (-8.78) ***	-0.004 7 (-5.94) ***
$B_{pre-open,t-1}^{att}$				-0.000 7 (-2.36) **	$B_{trading,t-1}^{att}$		-0.000 4 (-0.42)		-0.006 2 (-5.91) ***
调整 R^2	0.050 8	0.002 1	0.063 1	0.066 4	调整 R^2	0.001 0	0.000 3	14.36	0.163 3
DW 统计量	1.960 9	1.996 3	1.981 1	2.012 0	DW 统计量	2.000 0	1.998 6	1.932 3	2.025 2
卡方统计量	64.949	4.412 6	86.228	88.133	卡方统计量	1.013 8	0.175 3	171.96	207.98
[p 值]	[0.000 0]	[0.064 7]	[0.000 0]	[0.000 0]	[p 值]	[0.314 0]	[0.675 5]	[0.000 0]	[0.000 0]
影响	(+) ***	(+) *	(+) ***	(+) ***	影响			(+) ***	(+) ***

注：a. 式(10) $R_t = \alpha_3 + \sum_{i=1}^m \beta_{3,i} R_{t-i} + \gamma_{3,0} S_t + \sum_{j=1}^n \gamma_{3,j} S_{t-j} + \varepsilon_{3,t}$ 的原假设为 $H_{0,3} : \gamma_{3,j} = 0 \quad j = 0, 1, 2, \dots, n$ 。
b. 所有结果均利用了 White 的异方差一致协方差估计量对标准误和 t 统计量进行了修订，表中列出的是修订后的结果。
c. 表中其他说明同表 4 注释所示。

4.3 对实证结果的分析与讨论

本文的实证分析揭示了几点重要结果：基于股评的投资者情绪对未来股票收益率、股票交易量不具有预测能力，只有每日开盘前的投资者情绪对股票开盘价具有预测能力；但是，基于股评的投资者情绪的变化对同期收益率和同期交易量有显著的正向影响，其中日交易时段的投资者情绪对交易量的影响更显著。基于股评的投资者情绪的形成依赖于前期的收益率。投资者情绪对股票收益率的波动性既没有预测能力，也没有当期影响。本文的结果与一些已有文献结果一致。例如，Kim 和 Kim^[16] 同样发现基于雅虎财经股吧评论构建的投资者情绪对美国未来股票收益率无预测能力。另一方面，本文的结果与那些揭示了投资者情绪对股票收益率具有预测能力的结果并不矛盾。例如，Antweiler 和 Frank^[14] 揭示了由雅虎财经股吧评论形成的投资者情绪对美国股票高频收益率有显著影响。本文将日度数据拆分成开盘前的

非交易时段和开盘后的交易时段，发现开盘前的非交易时段形成的投资者情绪对股票开盘价具有预测能力，而交易时段形成的投资者情绪对收盘价具有正向影响。从一个侧面上，其结果说明基于股评的投资者情绪对股票价格的影响不会超过一个交易日。若进行高频的探讨，也许可以更进一步地揭示投资者情绪与股票收益率之间的 Granger 因果关系。

本文注意到东方财富股吧的股评中负面预期的帖子数量更多，这使得股评情绪指标均为负值，为此计算了去掉趋势的股评情绪。研究结果显示，无论是否去掉股吧中负向情绪偏多的系统偏差，投资者情绪对股票收益率的影响都非常相似。这意味着股评情绪的绝对值并不重要，相对值更重要。正如式(15)所揭示的，投资者情绪的变化对同期的股票收益率具有正向影响。这个结果与 DeLong 等^[2] 噪声交易者理论所揭示的关系一致。噪声交易者理论指出噪声交易者对未来价格的错误预期会影响当期的股票价格。

表 6 投资者情绪对日度沪深 300 股票交易量和波动性影响的实证结果

Table 6 Empirical results of the effect of investor sentiment on daily trading volume and volatility of CSI300

A: 对交易量的影响. 式(11): $V_t = a_0 + \sum_{i=1}^m a_i V_{t-i} + b_1 R_{t-1} + c_1 DisAgree_t + c_2 DisAgree_{t-1} + \varepsilon_t$										
解释变量	日度 $DisAgree$			日交易时段的 $DisAgree_{open}$			日非交易时段 $DisAgree_{close}$			
a_0	1.096 8 (1.90) *	-0.015 3 (-0.03)	0.432 3 (0.64)	1.469 3 (2.37) **	-0.107 2 (-0.16)	0.677 3 (0.92)	0.714 9 (1.72) *	0.301 8 (0.68)	0.392 7 (0.80)	1.178 3 (1.91) *
V_{t-1}	0.580 2 (16.36) ***	0.578 5 (16.34) ***	0.583 2 (16.42) ***	0.579 5 (16.43) ***	0.577 8 (16.32) ***	0.579 0 (16.42) ***	0.578 1 (16.17) ***	0.578 9 (16.39) ***	0.581 3 (16.21) ***	0.581 2 (16.55) ***
V_{t-2}	0.193 9 (5.51) ***	0.192 3 (5.47) ***	0.191 1 (5.44) ***	0.194 5 (5.55) ***	0.194 2 (5.51) ***	0.195 7 (5.58) ***	0.194 0 (5.51) ***	0.191 5 (5.47) ***	0.190 5 (5.42) ***	0.189 0 (5.42) ***
V_{t-3}	0.115 2 (3.61) ***	0.116 2 (3.64) ***	0.114 8 (3.60) ***	0.113 8 (3.57) ***	0.117 4 (3.68) ***	0.114 8 (3.59) ***	0.115 8 (3.63) ***	0.115 5 (3.62) ***	0.115 1 (3.61) ***	0.111 2 (3.49) ***
R_{t-1}	33.155 (11.15) ***	31.535 (10.23) ***	31.408 (10.20) ***	33.796 (11.21) ***	32.130 (10.76) ***	33.536 (11.13) ***	32.885 (11.04) ***	31.440 (9.68) ***	31.025 (9.12) ***	31.302 (9.68) ***
$DisAgree_t$		1.084 0 (1.28)	1.865 4 (1.86) *		1.255 3 (1.33)	2.484 1 (2.31) **		0.601 9 (1.08)	0.769 4 (1.17)	1.530 0 (2.11) **
$DisAgree_{t-1}$	-0.458 7 (-0.58)		-1.413 5 (-1.50)	-1.009 9 (-1.16)		-2.364 5 (-2.40) ***	0.067 0 (0.13)		-0.289 7 (-0.46)	-2.298 5 (-2.03) ***
调整 R^2	0.749 4	0.749 7	0.749 9	0.749 6	0.749 7	0.750 5	0.749 3	0.749 5	0.749 3	0.750 2
DW 统计量	2.013 9	2.005 8	2.017 7	2.014 8	2.005 2	2.019 8	2.009 8	2.008 1	2.012 0	2.017 5
B: 对收益率波动性的影响. 式(13): $Volatility_t = b_0 + b_1 Volatility_{t-1} + b_2 R_{t-1} + b_3 DisAgree_{t-1} + b_4 DisAgree_{t-1} + \varepsilon_t$										
解释变量	日度 $DisAgree$			日交易时段 $DisAgree_{open}$			日非交易时段 $DisAgree_{close}$			
b_0	1.35E-04 (0.49)	2.67E-04 (0.75)	1.98E+04 (0.50)	-0.000 14 (-0.37)	8.22E-05 (0.23)	-1.09E-04 (-0.26)	5.85E-04 (2.21) **	5.61E-04 (1.89) *	6.79E-4 (2.11) **	
$Volatility_{t-1}$	0.983 9 (195) ***	0.983 6 (196) ***	0.983 8 (195) ***	0.984 6 (195) ***	0.984 0 (196) ***	0.984 5 (195) ***	0.982 6 (195) ***	0.982 6 (194) ***	0.982 2 (194) ***	
R_{t-1}	-0.006 9 (-4.38) ***	-0.006 8 (-4.09) ***	-0.006 8 (-4.09) ***	-0.007 3 (-4.51) ***	-0.007 0 (-4.41) ***	-0.007 3E-04 (-4.50) ***	-0.006 8 (-4.36) ***	-0.005 9 (-3.33) ***	-0.006 3 (-3.46) ***	
$DisAgree_t$		-5.82E-05 (-0.13)	-1.71E-04 (-0.32)		1.98E-04 (0.40)	-8.33E-05 (-0.14)		-0.000 4 (-1.17)	-0.000 21 (-0.51)	
$DisAgree_{t-1}$	1.18E-04 (0.27)		2.04E-04 (-0.39)	5.04E-04 (1.00)		5.49E-04 (0.93)	-0.000 43 (-1.42)		-0.000 34 (-0.95)	
调整 R^2	0.965 9	0.965 9	0.965 8	0.965 9	0.965 9	0.965 9	0.965 9	0.965 9	0.965 9	
DW 统计量	2.033 3	2.033 4	2.032 6	2.033 9	2.035 0	2.033 6	2.037 7	2.033 4	2.035 9	

注: a. A 栏中对交易量建模的结果利用了 White 的异方差一致协方差估计量对标准误和 t 统计量进行了修订. 表中报告的是修订后的结果.

b. A 栏利用非交易时段投资者情绪分散程度指标建模的倒数第二列模型为 $V_t = a_0 + \sum_{i=1}^m a_i V_{t-i} + b_1 R_{t-1} + c_1 DisAgree_{close,t} + c_2 DisAgree_{close,t-1} + \varepsilon_t$, 而最后一列的模型为 $V_t = a_0 + \sum_{i=1}^m a_i V_{t-i} + b_1 R_{t-1} + c_1 DisAgree_{close,t} + c_2 DisAgree_{open,t-1} + \varepsilon_t$.

c. 表中其他说明同表 4 的注释所示.

股票收益率是投资者情绪的 Granger 原因, 这说明投资者情绪的形成是依赖于前期市场价格趋势的, 同时长短期有差异. 在半个月至两个

月的时间跨度上, 前期收益率对投资者情绪的影响为负, 而较短的时间跨度上, 前期收益率对投资者情绪的影响为正. 在式(9)的估计结果中

还发现投资者情绪具有惯性,这种惯性通常会持续两周左右.投资者情绪滞后期阶数与收益率滞后期阶数的结果建议了以两周作为区分长期和短期的分界点.这些反映了参与股吧讨论的这些投资者在短期内是趋势追随者,倾向于根据之前股票市场的涨跌形成看涨看跌的正向预期,即追涨杀跌;而在长期这些投资者又倾向于认为股票市场是均值回复的,会形成与市场趋势相反的情绪.

本文揭示了投资者情绪对股票交易量无预测能力,这个结果与 Antweiler 和 Frank^[14]、Kim 和 Kim^[16]等均相同,这说明股票交易量是更难预测的.同时,本文揭示了投资者情绪分散程度的变化对股票交易量有当期正向影响.投资者情绪分散程度与交易量的正向关系符合金融学的理论逻辑,因为只有投资者的预期不一致,有人看涨有人看跌时,市场上的交易才会呈现出有人买有人卖的局面,才会有更多的成交交易;否则如果大家的预期一致,那么便没有人愿意充当买方(卖方)的交易对手,从而会导致市场上的买单(卖单)无法成交.此外,本文还揭示了投资者情绪对股票收益率波动性无预测能力也无影响,这与绝大多数的实证结果一致,例如 Kim 和 Kim^[16].

综上所述,上述研究结果能有助于理解参与股吧讨论的这些交易者的交易行为.交易者们根据市场价格趋势形成自己对走势的预期或者信念,称之为投资者情绪,这种投资者情绪的形成短期内倾向于趋势追随而长期倾向于趋势反转.他们将自己的看涨看跌情绪(对未来市场走势的预期)作为交易信号来决定买入或卖出.他们各自都相信自己的预期判断,所以他们的投资决策非常主观.由于每人的情绪(预期)也都不尽相同,所以市场会有买有卖,当投资者情绪差异越大时,交易越活跃,成交量越高.从市场整体来看,市场的投资者情绪越高,说明对股票的需求越高,股价也就越高.从这些行为和行为产生的结果看,本文倾向于认为参与我国股吧讨论的这些投资者属于噪声投资者.这意味着这些投资者虽然是非理性投资者,是基于自己的情绪做出投资决策,但他们确实能够对股票价格造成影响,引起价格偏离其内在价值.

5 结束语

本文利用从股吧评论文本中提取的投资者情绪,研究了基于股评的投资者情绪对股票市场收益率、交易量和波动性的可预测能力和影响.为了挖掘投资者情绪,采集了东方财富网旗下股票社区“股吧”中的沪深300成分股的所有评论帖子,根据发帖的数量稳定程度,将研究样本确定为2009年1月1日至2014年10月31日.为了挖掘互联网平台上这些由各个投资者发布的消息、帖子等看似杂乱的文本信息内容的价值,并且考虑到中文股评通常长短不一、文法自由、适用的情感词典极为匮乏,因此本文采用了基于分类器的建模方法.在具体分类器的选择上,采用基于经验分布的朴素贝叶斯模型,该模型简单、鲁棒,在海量股评数据上有极高的训练及使用效率.本文致力于发掘文本中投资者对股票市场未来趋势的预期或买卖倾向,建立了看涨、看跌、中性三类投资者情绪分类模型.五折交叉检验结果显示,三类情绪的预测精度均达到了75%左右.基于这种方法构建的投资者情绪指标,对沪深300指数月度、周度、日度三个频率的收益率,以及对日度的交易量和波动性进行了实证分析.

本文分别采用 Granger 因果检验、Granger 瞬时因果检验、跨期回归分析的方法,揭示了在月度、周度和日度频率上,投资者情绪对股票市场收益率、交易量、波动性均无预测能力,但对当期的股票市场收益率和交易量有显著影响.在更高频率上,发现开盘前的股评情绪对开盘价具有预测能力;开盘后交易时段的股评情绪对当日收盘价具有当期正向影响.具体而言,投资者情绪对股票价格存在当期正向影响,说明从市场整体来看市场的投资者情绪越高,对股票的需求越高,股价也就越高,反之亦然.投资者情绪分散程度与交易量存在同期的正向关系,投资者情绪分散程度越高,交易越容易达成,交易量越大.同时,揭示了投资者情绪的形成依赖于市场前期的收益率水平,参与股吧讨论的这些投资者在短期属于趋势追随者,价格上涨才会导致投资者情绪的看涨,而在长期属于趋势反转者,长时间的价格上涨容易令投资者形成看跌情绪.

从实证结果反映出来的参与股吧评论的这些投资者的行为以及这些行为所产生的结果看 我们倾向于认为这些投资者属于噪声投资者,虽然他们是

基于自己对未来的预期做出投资决策,但他们的活动会对市场交易量和资产价格产生影响. 论文结果在一定程度上支持了噪声交易者理论.

参 考 文 献:

- [1] Black F. Noise [J]. *Journal of Finance*, 1986, 41: 529 – 543.
- [2] DeLong J B, Shleifer A, Summers L H, et al. Noise trader risk in financial markets [J]. *Journal of Political Economy*, 1990, 98(4): 703 – 738.
- [3] Bu H, Pi L. Does investor sentiment predict stock returns? The evidence from Chinese stock market [J]. *Journal of Systems Sciences and Complexity*, 2014, 27(1): 130 – 143.
- [4] Baker M, Wurgler J. Investor sentiment and the cross-section of stock returns [J]. *Journal of Finance*, 2006, 61(4): 1645 – 1680.
- [5] Lee W Y, Jiang C X, Indro D C. Stock market volatility, excess returns, and the role of investor sentiment [J]. *Journal of Banking and Finance*, 2002, 26(12): 2277 – 2299.
- [6] Brown G W, Cliff M T. Investor sentiment and the near-term stock market [J]. *Journal of Empirical Finance*, 2004, 11(1): 1 – 27.
- [7] Schmeling M. Investor sentiment and stock returns: Some international evidence [J]. *Journal of Empirical Finance*, 2009, 16(3): 394 – 408.
- [8] Lee C, Shleifer A, Thaler R H. Investor sentiment and the closed-end fund puzzle [J]. *Journal of Finance*, 1991, 46(1): 75 – 109.
- [9] Neal R, Wheatley S M. Do measures of investor sentiment predict returns? [J]. *Journal of Financial and Quantitative Analysis*, 1998, 33(4): 523 – 547.
- [10] Ritter J R. The long-run performance of initial public offerings [J]. *Journal of Finance*, 1991, 46(1): 3 – 27.
- [11] Baker M, Wurgler J. The equity share in new issues and aggregate stock returns [J]. *Journal of Finance*, 2000, 55(5): 2219 – 2257.
- [12] Baker M, Wurgler J. A catering theory of dividends [J]. *Journal of Finance*, 2004, 59(3): 1125 – 1165.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Kim S H, Kim D. Investor sentiment from internet message postings and the predictability of stock returns [J]. *Journal of Economic Behavior & Organization*, 2014, 107: 708 – 729.
- [17] Chen H, De P, Hu Y J, et al. Wisdom of crowds: The value of stock opinions transmitted through social media [J]. *Review of Financial Studies*, 2014, 27(5): 1367 – 1403.
- [18] Sprenger T O, Tumasjan A, Sandner P G, et al. Tweets and trades: The information content of stock microblogs [J]. *European Financial Management*, 2014, 20(5): 926 – 957.
- [19] Bartov E, Faurel L, Mohanram P S. Can Twitter Help Predict Firm-level Earnings and Stock Returns? [R]. Rotman School of Management Working Paper No. 2631421, Available at SSRN: <https://ssrn.com/abstract=2782236>, 2016.
- [20] Joseph K, Wintoki M B, Zhang Z L. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search [J]. *International Journal of Forecasting*, 2011, 27(4): 1116 – 1127.
- [21] Da Z, Engelberg J, Gao P. In search of attention [J]. *Journal of Finance*, 2011, 66(5): 1461 – 1499.
- [22] Brown G W, Cliff M T. Investor sentiment and asset valuation [J]. *Journal of Business*, 2005, 78(2): 405 – 439.
- [23] 伍燕然, 韩立岩. 不完全理性、投资者情绪与封闭式基金折价之谜 [J]. *经济研究*, 2007, (3): 117 – 129.
Wu Yanran, Han Liyan. Imperfect rationality, sentiment and closed-end-fund puzzle [J]. *Economic Research Journal*, 2007, (3): 117 – 129. (in Chinese)
- [24] 刘维奇, 刘新新. 个人和机构投资者情绪与股票收益——基于上证 A 股市场研究 [J]. *管理科学学报*, 2014, 17(3): 70 – 87.
Liu Weiqi, Liu Xinxin. Individual /institutional investor sentiment and stock returns: Study based on Shanghai A-share market [J]. *Journal of Management Sciences in China*, 2014, 17(3): 70 – 87. (in Chinese)

- [25]张 强,杨淑娥. 噪音交易、投资者情绪波动与股票收益[J]. 系统工程理论与实践, 2009, 29(3): 40–47.
Zhang Qiang, Yang Shue. Noise trading, investor sentiment volatility and stock returns[J]. Systems Engineering: Theory & Practice, 2009, 29(3): 40–47. (in Chinese)
- [26]文凤华,肖金利,黄创霞,等. 投资者情绪特征对股票价格行为的影响研究[J]. 管理科学学报, 2014, 17(3): 60–69.
Wen Fenghua, Xiao Jinli, Huang Chuangxia, et al. The effects of characteristics of investor sentiment on stock price behaviors[J]. Journal of Management Sciences in China, 2014, 17(3): 60–69. (in Chinese)
- [27]黄德龙,文凤华,杨晓光. 投资者情绪指数及中国股市的实证[J]. 系统科学与数学, 2009, 29(1): 1–13.
Huang Delong, Wen Fenghua, Yang Xiaoguang. Investor sentiment index and empirical evidence from China's stock market[J]. Journal of Systems Science and Mathematical Sciences, 2009, 29(1): 1–13. (in Chinese)
- [28]宋泽芳,李 元. 投资者情绪与股票特征关系[J]. 系统工程理论与实践, 2012, 32(1): 27–33.
Li Zefang, Li Yuan. Relationship between investor sentiment and stock characteristic[J]. Systems Engineering: Theory & Practice, 2012, 32(1): 27–33. (in Chinese)
- [29]Zhao J, Dong L, Wu J, et al. Moodlens: An emoticon-based sentiment analysis system for Chinese tweets[C]. New York: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM, 2012, 1528–1531.
- [30]Liu B, Zhang L. A Survey of Opinion Mining and Sentiment Analysis[M]. New York: Mining Text Data, Springer US, 2012: 415–463.
- [31]Zhao Y Y, Qin B, Liu T. Sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834–1848.
- [32]Tan P N. Introduction to Data Mining[M]. Chennai: Pearson Education India, 2006.
- [33]Wu J, Yuan H, Xiong H, et al. Validation of overlapping clustering: A random clustering perspective[J]. Information Sciences, 2010, 180(22): 4353–4369.
- [34]Dutka J. The incomplete beta function: A historical profile[J]. Archive for History of Exact Sciences, 1981, 24(1): 11–29.
- [35]Granger C W J. Testing for causality: A personal view[J]. Journal of Economic Dynamics and Control, 1980, 2: 329–352.
- [36]部 慧,李 艺,汪寿阳. 国际基金持仓与大豆商品期货价格关系的实证研究[J]. 管理评论, 2008, 20(5): 3–8.
Bu Hui, Li Yi, Wang Shouyang. Relationship between funds and commodity futures prices: An empirical study based on soybean futures market[J]. Financial Management, 2008, 20(5): 3–8. (in Chinese)

Investor sentiment extracted from internet stock message boards and its effect on Chinese stock market

BU Hui¹, XIE Zheng¹, LI Jia-hong¹, WU Jun-jie^{1 2*}

1. School of Economics and Management, Beihang University, Beijing 100191, China;

2. Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China

Abstract: This paper examines whether investor sentiment expressed in online user generated content (UGC) in the stock message boards of Easymoney.com has a predictive power for Chinese stock market. Naive Bayes classifier is used to obtain the expectation of every message, and then a new indicator of investor sentiment is put forward that combines the idea of bullishness and investor attention. Granger causality test, instantaneous Granger causality test and intertemporal regression analysis are applied to study the predictability of investor sentiment for stock returns, volatility, and trading volume. It is found that investor sentiment has no predictable power for returns, volatility, or trading volume; while investor sentiment does have a significant positive effect on the contemporaneous stock price, and disagreement of investor sentiment can affect contemporaneous trading volume. Investor sentiment using the messages before market opening has the predictive power of open price, and investor sentiment from the trading hours is more valuable for closing price and trading volume. Moreover, stock returns are the Granger cause of investor sentiment. Our results can help to understand the behavior of investors and their effects on stock market.

Key words: investor sentiment; noise trader; text mining; Granger causality test