

中国股票网络论坛的信息含量分析

段江娇 刘红忠 曾剑平

(上海理工大学管理学院, 上海 200093;
复旦大学经济学院/计算机科学技术学院, 上海 200433)

摘 要: 本文选取东方财富网股吧论坛的个股帖子, 使用计算机文本处理技术提取帖子情绪, 结合证券分析师对个股的“中性”评级数据, 实证研究了我国股票网络论坛的信息含量问题。研究发现: 股票当日收益率受当日论坛情绪影响, 为显著正相关; 股票未来两日收益率与帖子数显著负相关; 股票当日的帖子数显著正向影响当日股价波动, 而且能影响未来两日的股价波动; 当日情绪分歧度越大, 未来两日的交易量越大。本研究不仅有助于理解我国股票网络论坛对股票市场的影响机制, 而且也为监管层对市场调控和监管提供了一定的决策依据。

关键词: 投资者情绪; 情绪分歧度; 股票收益率; 股价波动性; 股票交易量

JEL 分类号: G10, G14, G19 **文献标识码:** A **文章编号:** 1002-7246(2017)10-0178-15

一、引 言

随着网上股票交易及股票信息发布和共享的快速增长, 投资者越来越倾向于在股票网络论坛上表达自己对股票或股市的看法和预期。据中国互联网络信息中心(CNNIC)第 40 次调查报告, 截至 2017 年 6 月底, 我国网民规模达 7.51 亿人。随着我国网民规模与日俱增, 股票网络论坛的用户数亦快速增长。相应地, 股票网络论坛越来越受到研究者的关注。

国外在这方面的研究涉及论坛帖子数、帖子内容(Antweiler and Frank, 2004; Das and Chen, 2007; Zhang and Swanson, 2010)。论坛帖子数可用于投资者关注的代理变量, 帖子内容可用来提取投资者情绪。这些研究表明帖子数及帖子情绪对股票收益率、股价波动

收稿日期: 2014-10-26

作者简介: 段江娇, 理学博士, 副教授, 上海理工大学管理学院, Email: jjduan@usst.edu.cn.

刘红忠(通讯作者), 经济学博士, 教授, 复旦大学经济学院, Email: hzliu@fudan.edu.cn.

曾剑平, 理学博士, 副教授, 复旦大学计算机科学技术学院, Email: zjp@fudan.edu.cn.

* 本文感谢教育部人文社会科学研究规划基金项目(13YJAZH019)、国家自然科学基金项目(61073170)的资助, 感谢匿名审稿专家提出的宝贵修改意见。文责自负。

性和股票交易量存在显著影响。

我国一些学者也已开始股票网络论坛的相关研究(董大勇和肖作平 2011; 施荣盛和陈工孟 2012; Huang et al. 2016; 杨晓兰等 2016),但从相关文献看,这些研究对象多数仅限于帖子数,较少涉及从帖子内容提取情绪进行考察。因此,目前对于我国股票网络论坛的信息含量仍缺乏全面的研究。

本文提取了较完整的股票网络论坛特征维度,这些特征维度包括帖子数、帖子情绪及情绪一致性指标。帖子数是一个较容易统计的维度,在国内外对股票网络论坛的研究中常被用作投资者关注的代理变量。帖子情绪则是一个相对难以获得的特征,它包含在帖子内容中,反映了投资者对市场或个股涨跌的判断或预期。因此,本文选取我国股票网络论坛数据,从股票论坛帖子内容提取投资者情绪,研究股票网络论坛和股票价格的关系,分析我国股票网络论坛的信息含量。

本文的创新在于:(1)研究视角上,基于我国股票网络论坛,结合证券分析师对个股的“中性”评级数据,除了选用“帖子数”,还考虑了“帖子情绪”,拓展了国内相关研究的范围;(2)研究方法上,使用事件分析法、同期回归(Contemporaneous Regressions)和预测回归(Predictive Regressions)等方法分析了股票论坛对市场的影响。综合解决了赵龙凯等(2013)只研究同期、俞庆进和张兵(2012)仅研究滞后期的问题;(3)在研究内容上,针对国内相关研究只关注对股票收益的影响,本文增加了股票论坛对股价波动和交易量影响的研究。

以下内容的结构安排是:第二部分是文献回顾和研究假设;第三部分是研究方法和数据样本;第四部分是实证结果;第五部分是稳健性检验;最后是结论和启示。

二、文献回顾与研究假设

(一)论坛帖子信息对股票收益的影响

行为金融学认为,投资者情绪是影响股票收益不可忽视的因素(De Long et al., 1990)。按照投资者情绪理论模型(Barberis et al., 1998)的解释,人们进行投资决策时存在两种错误范式:其一是由于投资者心理存在保守性偏差或“锚定”心理,导致股价对信息反应不足,投资者情绪对股票收益的影响为正向;另一种是由于投资者心理存在代表性启发偏差,造成股价对信息反应过度,投资者情绪对股票收益的影响为负向。

另外,投资者有限关注也会影响股票收益。投资者有限关注是一种稀缺资源(Kahneman, 1973)。投资者有限关注理论认为,由于时间和精力限制,个人投资者只会对自己已关注到的信息进行分析和判断,调整其投资者行为,从而导致了暂时性定价偏差的发生。Barber and Odean(2008)的价格压力假说认为大多数个人投资者没有办法在研究所有的股票之后再做购买股票的决定,而只是从高度吸引他的注意力的股票中进行选择,关注的增加会在短期内对股票产生向上的价格压力,之后会有反转。这种反转与基于有限关注的过度关注弱势假设一致(Seasholes and Wu, 2007; 饶育蕾和盛虎, 2010; 俞

庆进和张兵 2012)。

近年来,研究者开始从股票网络论坛选取个人投资者情绪和个人投资者关注的代理变量。从国外对股票论坛和股票收益的相关研究看,同时涉及论坛帖子数和帖子内容。论坛帖子数用于个人投资者关注的代理变量,帖子内容用来提取个人投资者情绪。Antweiler and Frank (2004) 使用计算机文本分类技术从 Yahoo 和 Raging Bull 股票留言板帖子中提取情绪。研究发现股票当期帖子情绪与收益率显著正相关,当日帖子数与未来一天的收益率显著负相关。Das and Chen (2007) 从 Yahoo 在线投资论坛股票留言板提取中小投资者情绪指数,利用网络论坛发帖者的看涨、看跌或看平观点,建立了一个基于 Morgan Stanley 高科技股票指数的情绪指数,研究表明:市场活跃程度与该情绪指数有很强的相关性。Tetlock et al. (2008) 研究发现从媒体报道的语言和内容中提取的情绪会影响股票未来收益。Sabherwal et al. (2011) 研究发现 TheLion 在线论坛中的热门股票存在“Pump - Bump”(拉高 - 出货)的模式¹,平均超常收益率在两天内上涨到最高值,而在成为热门股之后开始急剧下跌,出现负超常收益率,并且平均累积超常收益率在热门股事件发生后 30 天仍是负值。他们还发现股票当日帖子情绪与当日收益率显著正相关,与未来一天收益率显著负相关,当日情绪分歧度与未来一天收益率显著负相关。

从国内相关研究看,起步较晚,多数文献只使用论坛发帖量或网络搜索量作为投资者关注的代理变量,较少涉及帖子内容。董大勇和肖作平(2011)选用东方财富网股吧发帖量和沪深 A 股中注册地在四川和辽宁两省的 109 家上市公司数据,对信息交流中“家乡偏误”(Home bias)的存在性、影响因素及其对资产定价的影响进行了研究。施荣盛和陈工孟(2012)使用东方财富股票论坛中的超额发帖量作为个人投资者关注的代理变量,结合分析师中性评级事件,对投资者关注和股票收益率进行了实证研究,研究发现高关注度的股票投资组合的日超常收益率显著为负。赵龙凯等(2013)对同期的百度搜索量与股票收益率进行了实证研究,发现高关注度股票组合的平均收益率显著地大于低关注度股票组合,但是关注度的变化率对股票收益率的影响不显著,即不会系统地影响股票收益率。杨晓兰等(2016)选取创业板上市公司在东方财富网股吧 2012 年 9 月 18 日至 2013 年 10 月 15 日的样本数据,研究了本地偏好对股票收益率的影响,研究发现投资者本地关注现象对股票收益率的影响依赖于网络论坛帖子的投资者情绪。

上述国内相关研究较少涉及股票论坛帖子的文本内容,但是在股票网络论坛中,帖子的文本内容隐含了个人投资者情绪。投资者关注和投资者情绪是投资行为在不同阶段的表现。在认知心理学看来,投资行为是一个系统的信息处理过程,包括对投资感觉的输入、变换、简约、加工、存储直至形成具体投资行为的全过程。每个阶段都有可能存在认知偏差而导致股价异常。投资者关注度高的股票并不意味着投资者对股票的涨跌看法一

¹ 这种模式是一种股票买卖的骗局,串谋者事先买进特定股票,然后在论坛中大力吹捧这些股票,通常是发布虚假的乐观讯息,藉此推高股价,然后在高位卖出其持有的股票以赚取暴利。串谋者出货后,股价会急跌,即哄抬股价而导致股价暴跌。

致。根据 Barberis et al. (1998) 的投资者情绪理论模型,由于互联网对信息传播的快速与便捷性,投资者的“锚定”心理可能导致股价对信息的反应不足,表现为股票网络论坛的帖子情绪对股票收益短期有正向影响。另一方面,由于股票网络论坛信息的易得性,这种低成本和易访问的信息可能造成股价对信息的反应过度。从我国投资者行为的相关文献看,投资者情绪短期内影响股票收益。李心丹等(2002)的研究表明我国个体投资者存在认知偏差,而且认知偏差会影响收益。王美今和孙建军(2004)根据“央视看盘”节目构造 BSI 指标,实证发现同期投资者情绪变化显著正向影响收益,机构投资者是可能的噪声交易者风险源。刘红忠和何文忠(2012)对我国股票高频交易数据的研究表明:一方面,信息融入股价存在一个较为缓慢的过程,另一方面,也存在着股价对信息的过度反应。

基于投资者情绪和有限关注的理论,结合股票网络论坛相关的研究文献,我们认为帖子情绪和帖子数对股票收益的影响机理为:当一只股票的论坛帖子数增加时,其受到的关注程度上升,帖子看涨情绪增多,被个人投资者购买的概率增加。在市场存在异质信念和卖空限制的情况下,短期内股价上升,股票收益率为正。另一方面,帖子数在网络论坛中是一种直观的信息,投资者对很容易处理的信息会产生过度反应。又由于我国投资者追涨杀跌、短线操作、频繁换手等投机性特征较浓,论坛中被高度关注的股票很容易转变为被过度关注。理性投资者可能会意识到股价中包含关注度溢价,帖子看跌情绪增多。理性投资者会卖出被高度关注的股票。随着帖子情绪看跌,个人投资者的关注发生转移,不再有买家跟进,之前帖子数越高的股票,未来价格回落幅度越大。所以,短期内股票收益随帖子情绪高涨而上升,随帖子情绪低落而下降。

综上,本文在提取我国股票网络论坛帖子情绪的基础上,提出如下的假设 1:

假设 1: 股票的帖子情绪短期内正向影响股票收益,而帖子数负向影响股票未来收益。

(二) 论坛帖子数对股票价格波动的影响

在这方面的研究,目前国内大多数文献将论坛帖子数或网络搜索量作为个人投资者关注代理变量,仅研究投资者关注对股票收益的影响,较少涉及投资者关注对股价波动性的影响。例如董大勇和肖作平(2011)、施荣盛和陈工孟(2012)、俞庆进和张兵(2012)、赵龙凯等(2013)的研究。姜富伟等(2011)对中国股票市场可预测性进行了实证研究,同样只涉及股票收益的预测。但是国外的研究文献表明,帖子数能影响股价波动。例如, Antweiler and Frank(2004)的研究表明,论坛帖子数与同期股价波动显著正相关,而且还能预测未来一天的股价波动。Das and Chen(2007)的研究也表明,论坛帖子数与股价波动显著正相关。而 Sabherwal et al. (2011)的研究选用每日帖子数进入前十名的股票作为样本,可能因为样本数偏少,没有发现帖子数显著影响股价波动的现象。

基于 Barber and Odean(2008)的价格压力假说,结合上述文献的研究成果,我们认为,当一只股票的论坛帖子数增加时,其受到的关注程度上升,对股票价格产生向上的压力。由于我国个人投资者存在过度自信等心理偏差(李心丹等,2002),导致频繁追涨等非理性的交易行为,所以个人投资者会对关注度高的股票进行积极交易,从而影响股价波动。同时“高明的投资者”(常被称为知情交易者)并没有与非理性投资者进行反向交易,

纠正股票定价错误,平抑股价波动(胡昌生和池阳春,2013)。而当这只股票因为过度关注股价开始下跌时,由于投资者过度恐惧等心理偏差,会出现频繁杀跌等投资行为,同样也会影响股价波动。所以,帖子数高的股票,短期由于价格上涨而使波动性增大,未来由于价格下跌,同样使波动性增大。因此,在我国股票网络论坛帖子数的基础上提出假设 2:

假设 2: 股票帖子数正向影响股价波动,有一定持续性。

(三) 论坛帖子情绪一致性对股票交易量的影响

关于投资者认知的一致性能否影响交易量,金融理论有两种不同的观点。经典的观点以 Hirshleifer (1977) 等为代表,认为分歧度越大或一致性越低,股票交易量越大。另一种观点是以 Milgrom and Stokey (1982) 为代表的“无交易定理”(no-trade theorem)。按照“无交易定理”,在理性预期均衡状态下,投资者担心交易对手拥有自己所不知道的资讯,因而不轻易下单,从而不会有交易发生。这样投资者认知的分歧度不会引起交易,导致市场价格和信念的重新调整。Antweiler and Frank (2004) 使用 15 分钟间隔的高频数据,研究发现帖子情绪一致性显著地负向影响同期的股票交易量,这个结论支持经典的关于情绪分歧度影响交易量的假设,即情绪分歧度越大,股票交易量越大。Antweiler and Frank (2004) 的研究还发现当日帖子情绪的一致性显著地正向影响未来一天的股票交易量,而在未来第二天又转变为负向影响,这个发现可以用 Cao et al. (2002) 的研究来解释。Cao et al. (2002) 揭示了抱观望态度的投资者会延迟交易,直到股票的价格走势验证了他预期的私有信息后才进行交易。这样,投资者情绪越一致,也有引起交易量增加的可能。

目前国内研究较少涉及帖子情绪一致性对股票交易量影响的研究。关于帖子情绪一致性对股票交易量影响的机理,我们认为:由于我国个人投资者的经验相对缺乏,非理性投资行为较多,投资者情绪越一致,短期内股票交易量就越大,可能导致未来股价偏离基础价值越大。随着互联网股票市场信息的传播,论坛帖子情绪分歧度增大,股价回归基础价值,股票交易量增大的趋势在未来可能反转。所以提出假设 3:

假设 3: 股票的帖子情绪一致性对股票交易量有影响,且具有短期效应。

三、研究方法和数据样本

(一) 股票论坛情绪指数的构造

1. 个股论坛帖子情绪的量化

本文选取我国股票市场上用户访问量最大、最具影响力的股票网络论坛——东方财富网股吧²为研究对象,从个股的发帖或回帖内容提取帖子情绪。用户可以进入自己所关注的股票频道进行发帖或回帖,一个帖子包含了帖子时间、内容以及用户标识等信息。情绪体现在帖子的文本内容中,情绪值用 5、4、3、2、1 表示,分别对应强烈买入、买入、中

² 在中国股票网络论坛最流行的前三名(东方财富网股吧、和讯、金融界)中,东方财富网股吧的用户数和浏览量最大(Huang et al., 2016)。

性、卖出和强烈卖出五档。图 1 是论坛帖子情绪类别识别的程序流程,其中情绪分类之前需要对帖子内容进行中文词汇的切分³。在词汇切分的基础上,本文通过检测词汇前后的否定词个数,进一步对带有情绪特征的词汇进行必要的转换。其它的一些特征词还包括疑问词、转折词等。这样,就把一个中文文本串转换为一个特征词汇序列,这个序列和英文文本的独立词汇形式相同。因此在对帖子情绪识别时,本文参照 Antweiler and Frank (2004) 并使用贝叶斯文本分类,例如,“时不我待! 坚决买入!”的情绪值为 5,而“垃圾股,快跑”的情绪值为 1。

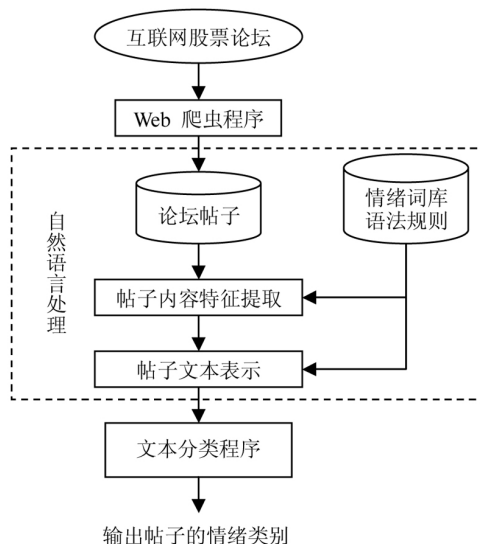


图 1 帖子情绪识别的程序流程

2. 论坛个股情绪指数和情绪一致性指数的构造

在个股帖子情绪量化的基础上,可计算股票 i 第 t 日的看涨情绪 $BSI_{i,t}$,如公式(1):

$$BSI_{i,t} = \frac{\text{看涨帖子数} - \text{看跌帖子数}}{\text{看涨帖子数} + \text{看跌帖子数}} \quad (1)$$

从 $BSI_{i,t}$ 的计算公式可知, $BSI_{i,t}$ 的最大值为 1,说明股票 i 第 t 日的所有帖子都是看涨。 $BSI_{i,t}$ 的最小值为 -1,说明股票 i 第 t 日的所有帖子都是看跌。如果 $BSI_{i,t}$ 越接近于 0,说明看涨帖子数与看跌帖子数越接近。

综合帖子数和看涨情绪可计算个股情绪指数 $MsgBSI_{i,t}$,如公式(2):

$$MsgBSI_{i,t} = BSI_{i,t} \times \ln(1 + Msg_{i,t}) \quad (2)$$

其中 $\ln(\cdot)$ 表示取自然对数,以下类同。 $Msg_{i,t}$ 表示股票 i 第 t 日的论坛帖子数。

股票 i 第 t 日的论坛情绪一致性指数的测度如公式(3),其最大值为 1,反映了帖子一

³ 本文采用中科院汉语词法分析系统 ICTCLAS(<http://www.ictclas.org/>) 对帖子内容进行词汇切分。ICTCLAS 是当前中文分词性能最好的系统,其切分的准确率可达 98.45%。

致看涨或一致看跌;其最小值为 0,反映了帖子情绪分歧度最大。

$$AgrIdx_{i,t} = 1 - \sqrt{1 - BSI_{i,t}^2} \quad (3)$$

(二) 个股情绪对股票收益率、波动性、交易量影响的检验模型

在检验模型中用到的变量名及含义或计算公式如表 1 所示。

表 1 变量名及含义或计算公式说明

变量名	含义或计算公式
$Msg_{i,t}$	股票 i 第 t 日的论坛帖子数
$BSI_{i,t}$	股票 i 第 t 日的论坛个股看涨情绪,计算见公式(1)
$MsgBSI_{i,t}$	股票 i 第 t 日的论坛个股情绪指数,计算见公式(2)
$AgrIdx_{i,t}$	股票 i 第 t 日的论坛情绪一致性指数,计算见公式(3)
$Price_{i,t}$	股票 i 第 t 日的收盘价
$Return_{i,t}$	股票 i 第 t 日的收益率,计算公式为: $\ln(Price_{i,t}) - \ln(Price_{i,t-1})$
$MktRet_t$	第 t 日市场指数的收益率
$AR_{i,t}$	股票 i 第 t 日按市场调整的超常收益率 $Return_{i,t} - MktRet_t$
$Volatility_{i,t}$	股票 i 第 t 日日内价格波动,取值为第 t 日日内价格的标准差
$\Delta Volume_{i,t}$	股票 i 第 t 日的调整交易量,计算见公式(4)、(5)
Mkt_t	第 t 日市场指数
$Size_{i,t}$	股票 i 第 t 日的市值,取值为上一财务年度的公司市值,年度更新
$B/M_{i,t}$	股票 i 第 t 日的账面市值比,为每股净资产与每股股价的比值,每股净资产取上季度的值,每股股价取日度收盘价 B/M 频率为日度数据

$$Volume_{i,t} = \frac{\text{股票 } i \text{ 第 } t \text{ 日的成交量}}{\text{股票 } i \text{ 第 } t \text{ 日的流通股数}} \quad (4)$$

$$\Delta Volume_{i,t} = Volume_{i,t} - \text{Mean}(Volume_{i,t-3}, \dots, Volume_{i,t-43}) \quad (5)$$

其中, $\text{Mean}()$ 表示股票 i 发帖日 t 前两个月⁴交易量 $Volume$ 的均值。

同期股票收益率对帖子数、情绪一致性指数和个股情绪指数的回归如公式(6),其中控制变量的选取参照 Antweiler and Frank(2004)、Sabherwal et al. (2011) 的研究,包含市场收益率、公司规模及股票滞后一日的收益率,此外我们还增加了账面市值比作为控制变量。

$$Return_{i,t} = \beta_0 + \beta_1 \ln(1 + Msg_{i,t}) + \beta_2 AgrIdx_{i,t} + \beta_3 MsgBSI_{i,t} + \beta_4 MktRet_t + \beta_5 \ln(Size_{i,t}) + \beta_6 B/M_{i,t} + \beta_7 Return_{i,t-1} \quad (6)$$

同期股票波动性和交易量对帖子数、情绪一致性指数和个股情绪指数的回归分别如公式(7)和公式(8),其中控制变量的选取参照 Antweiler and Frank(2004) 的研究,与公式

⁴ 考虑到发帖事件的影响,时间窗口从第 $t-3$ 日至前两月计算,改变窗口长度为一个月或三个月,不影响本文结论。

(6) 不同的是选取市场指数,而不是市场收益率作为控制变量。

$$Volatility_{i,t} = \beta_0 + \beta_1 \ln(1 + Msg_{i,t}) + \beta_2 AgrIdx_{i,t} + \beta_3 MsgBSI_{i,t} + \beta_4 \ln(Mkt_t) + \beta_5 \ln(Size_{i,t}) + \beta_6 B/M_{i,t} + \beta_7 Volatility_{i,t-1} \quad (7)$$

$$Volume_{i,t} = \beta_0 + \beta_1 \ln(1 + Msg_{i,t}) + \beta_2 AgrIdx_{i,t} + \beta_3 MsgBSI_{i,t} + \beta_4 \ln(Mkt_t) + \beta_5 \ln(Size_{i,t}) + \beta_6 B/M_{i,t} + \beta_7 Volume_{i,t-1} \quad (8)$$

未来一至两期股票收益率、股票波动性和交易量对帖子数、情绪一致性指数和个股情绪指数的回归与同期类似,只是相应的回归变量取的时期不同。

(三) 数据来源和样本选择

本文选取东方财富网股吧论坛中关于上证 A 股的帖子,时间自 2011 年 1 月 5 日至 2012 年 12 月 31 日,覆盖 2011 年 1 月 1 日以前所有的上市公司。剔除没有标注情绪值、交易停牌、金融类股票的帖子后,有 4663111 个帖子。本文参照施荣盛和陈工孟(2012)的方法选取在分析师对个股评级为“中性”时的帖子⁵来分析股票论坛的信息含量,基于以下两点考虑:(1)与 Barber and Odean(2008)的理论模型中将交易者分成知情交易者与噪音交易者的假设一致。证券分析师在发布评级报告前,经常通过路演等方式提前通知机构投资者(Juergens and Lindsey 2009)。这样,个人投资者相对机构投资者在获取分析师评级信息时,存在时间上、质量上的劣势。(2)具有更纯净的样本,且便于样本间进行比较。相对而言,“中性”评级报告,即分析师不推荐“买入”,也不推荐“卖出”,在理论上信息本身不对股价造成显著影响;选取了同样的评级等级,很好地控制了由于股票间信息差异而对股价产生的不同影响。如果分析师的投资评级有调整,则很可能有别的基本面信息产生,而不是股票论坛中的信息,因此我们在分析时应尽量控制这些信息对股票的影响。考虑到“中性”评级事件影响的超前与滞后性,时间窗选为 $[-1, 1]$,帖子数有 73148 个,个股按日统计的样本数有 4355 个。分析师对个股的评级数据、股票交易及财务数据来源于 CSMAR 数据库,股票日内价格来源于锐思 RESSET 高频数据库。

四、实证结果与分析

(一) 股票论坛在分析师不同投资评级时的描述性统计

为了比较股票论坛数据在分析师推荐“中性”评级与其他评级的情况,表 2 给出了在分析师推荐“买入”、“中性”和“卖出”评级时股票论坛帖子的描述性统计。从表 2 可以看出,在分析师推荐“中性”评级时的样本数,介于分析师推荐“买入”和“卖出”评级时的样本数之间。在分析师推荐“中性”时的个股日看涨情绪均值,介于分析师推荐“买入”和“卖出”评级时情绪的均值。同样,在分析师推荐“中性”时的个股日收益率均值,介于分

⁵ 分析师的评级值从 1 至 5 依次对应“卖出”至“买入”,考虑到同一只股票同一日有多个分析师的评级值,而且可能存在乐观倾向(Jackson 2005; Agrawal and Chen 2008; 张雪兰和何德旭 2008),我们取某只股票同一日分析师评级值的均值在 $[3, 4]$ 之间的为“中性”。

析师推荐“买入”时和“卖出”评级时收益率的均值⁶。论坛个股日收益率的均值在分析师推荐“买入”评级时大于0,在分析师推荐“卖出”时小于0,说明分析师“买入”和“卖出”评级分别给市场传递了利好和利空信息。从表2描述性统计可看出,选取分析师“中性”评级对应的股票论坛数据可基本排除其他基本面信息对个股的影响。因此,本文以下对股票论坛数据的分析是在分析师“中性”评级时的样本。

表2 股票论坛在分析师不同投资评级时的描述性统计

论坛个股日帖子数 <i>Msg</i>					
	样本数	平均值	最大值	最小值	标准差
所有	392273	11	3361	1	25.99
在分析师推荐“买入”评级时	13033	15	637	1	26.79
在分析师推荐“中性”评级时	4355	16	630	1	28.60
在分析师推荐“卖出”评级时	47	14	69	1	15.66
论坛个股日看涨情绪 <i>BSI</i>					
	样本数	平均值	最大值	最小值	标准差
所有	392273	0.116	1	-1	0.644
在分析师推荐“买入”评级时	13033	0.156	1	-1	0.586
在分析师推荐“中性”评级时	4355	0.084	1	-1	0.581
在分析师推荐“卖出”评级时	47	0.004	1	-1	0.569
论坛个股日收益率 <i>Return</i>					
	样本数	平均值	最大值	最小值	标准差
所有	392273	-0.06%	33.57%	-22.64%	2.59%
在分析师推荐“买入”评级时	13033	0.24%	10%	-10%	2.66%
在分析师推荐“中性”评级时	4355	-0.33%	10%	-10%	2.77%
在分析师推荐“卖出”评级时	47	-0.35%	3.41%	-5.03%	2.23%

(二) 事件分析和相关分析

表3给出了股票在论坛发帖前后从 $t-5$ 至 $t+5$ 交易日的平均帖子数和平均超常收益率 AR ,其中 t 表示股票的发帖日期。从表3可看出,在发帖日 t 的平均帖子数均比前后5日的要多。在发帖日 t 的前几日,股票的平均超常收益率显著为正,而在发帖日 t 之后,转变为显著为负。相应的变化如图2所示。另外,本文还进行了相关性分析,从相关性分析结果可以看出,论坛情绪指数与股票收益率显著正相关,股票帖子数与股价波动性显著正相关,情绪一致性指数与交易量显著负相关,相关性的检验结果等符合本文假设。限于篇幅没有列出。

⁶ 表2中论坛所有样本的个股日收益率最大值为33.57%、最小值为-22.64%,这是由于恢复上市交易时出现的。恢复上市交易首日不受涨跌幅限制。

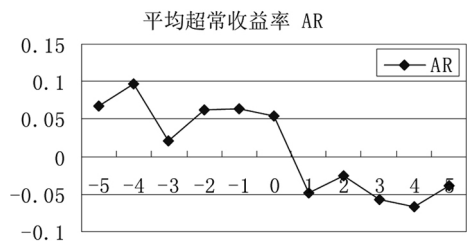


图 2 股票在发帖日前后的平均超常收益率

表 3 股票在发帖前后的平均帖子数、平均超常收益

时窗	<i>M</i> _{sg}	Sig. (2 – tailed)	<i>AR</i>	Sig. (2 – tailed)
<i>t</i> – 5	16. 8 ***	0	0. 0671 ***	0
<i>t</i> – 4	16. 613 ***	0	0. 0975 ***	0
<i>t</i> – 3	15. 322 ***	0	0. 020438	0. 177
<i>t</i> – 2	13. 853 ***	0	0. 061899 ***	0
<i>t</i> – 1	17. 117 ***	0	0. 06355 ***	0
<i>t</i>	20. 714 ***	0	0. 053246 ***	0. 001
<i>t</i> + 1	19. 25 ***	0	– 0. 04935 ***	0. 001
<i>t</i> + 2	16. 91 ***	0	– 0. 02565 *	0. 084
<i>t</i> + 3	18. 941 ***	0	– 0. 05632 ***	0
<i>t</i> + 4	17. 255 ***	0	– 0. 06641 ***	0
<i>t</i> + 5	11. 164 ***	0	– 0. 0393	0. 19

注：*、**、***分别表示 10%、5% 和 1% 显著性水平显著，下同。

(三) 最优回归模型选取

本文计量方法是基于面板数据的方法，每个面板为一个公司，共有 418 个面板。固定效应模型使用似然比检验，随机效应检验使用 Hausman 检验。从检验结果可知，上文所述 9 个回归的最优模型都是固定效应回归模型。限于篇幅没有列出。

(四) 同期回归分析

同期回归分析中使用的是日度数据，论坛帖子数及情绪是按自然日来统计⁷。表 4 给出了股票收益率、股价波动、股票交易量的同期回归结果。从表 4 可知，当日论坛情

⁷ 由于我国股票市场存在显著的“隔夜效应”(刘红忠和何文忠 2012)，在每日收盘后，论坛帖子有可能是对当日收盘前交易的评论，体现的是收盘前的交易信息，所以我们在统计论坛帖子数及情绪时没有区分收盘时间。未来的工作是使用高频数据进行检验。

绪指数与股票收益率收益显著正相关,进一步验证了本文假设 1 中帖子情绪对股票收益的正向影响。这个结果和 Antweiler and Frank(2004) 使用 15 分钟间隔的高频数据同期回归结果一致,说明和美国股票市场相比,我国股票市场信息融入股价的过程较为缓慢。当日股票帖子数与股价波动性显著正相关,说明本文假设 2 帖子数正向影响股价波动成立。当日情绪一致性指数与交易量显著正相关,与股票收益率显著负相关,很可能是非理性情绪造成的,和本文假设 3 股票的帖子情绪一致性对股票交易量有影响相一致。

表 4 股票收益率、股价波动、股票交易量的同期回归结果

解释变量	因变量					
	$Return_t$		$Volatility_t$		$\Delta Volume_t$	
	系数	t 统计量	系数	t 统计量	系数	t 统计量
截距项	0.001	0.291	18.904	1.501	7.067 ***	3.73
$\ln(1 + Msg_t)$	-0.002 ***	-5.249	0.743 ***	3.554	0.31 ***	8.164
$AgrIdx_t$	-0.003 ***	-3.125	0.526	0.977	0.269 **	2.494
$MsgBSI_t$	0.004 ***	10.955	0.054	0.344	-0.089 *	-1.648
$MktRet_t$	0.013 ***	37.71				
$\ln(Mkt_t)$			-2.683 *	-1.738	-0.803 ***	-3.458
$\ln(Size_t)$	0.001 *	1.774	0.335 *	1.92	-0.091 ***	-3.438
B/M_t	0	0.401	-4.818 ***	-7.77	0.037	0.429
$Return_{t-1}$	0.08 ***	5.999				
$Volatility_{t-1}$			0.657 ***	49.201		
$\Delta Volume_{t-1}$					0.693 ***	47.569
调整 R^2	0.292		0.569		0.539	

(五) 预测回归分析

未来一日和两日的回归结果分别由表 5 和表 6 给出。综合表 5 和表 6 可知,当日情绪指数对股票未来两日收益率没有影响,而帖子数却有显著负向影响。而且帖子数对未来两日的股价波动的影响仍显著为正。情绪一致性指数对未来两日交易量的影响由当日显著为正转变为显著为负,说明这种正向影响具有短期效应。预测回归结果和本文假设 3 股票的帖子情绪一致性对股票交易量影响具有短期效应相一致。

表 5 未来一日股票收益率、股价波动、股票交易量的回归结果

解释变量	因变量					
	$Return_{t+1}$		$Volatility_{t+1}$		$\Delta Volume_{t+1}$	
	系数	t 统计量	系数	t 统计量	系数	t 统计量
截距项	-0.007 ***	-3.752	2.1	0.183	2.665	1.467
$\ln(1 + Msg_t)$	-0.002 ***	-4.531	0.296 *	1.709	-0.097 ***	-2.661
$AgrIdx_t$	-0.002 *	-1.731	-0.194	-0.402	-0.148	-1.497
$MsgBSI_t$	0	1.238	-0.026	-0.182	0.044	0.856
$MktRet_{t+1}$	0.013 ***	37.06				
$\ln(Mkt_{t+1})$			-0.276	-0.198	-0.233	-1.056
$\ln(Size_{t+1})$	0.002 ***	3.536	0.396 ****	2.492	-0.038	-1.49
B/M_{t+1}	0.002 ***	4.602	-5.597 ***	-9.943	0.024	0.289
$Return_t$	0.055 ***	4.393				
$Volatility_t$			0.574 ***	48.395		
$\Delta Volume_t$					0.726 ***	51.176
调整 R^2	0.261		0.542		0.506	

表 6 未来两日股票收益率、股价波动、股票交易量的回归结果

解释变量	因变量					
	$Return_{t+2}$		$Volatility_{t+2}$		$\Delta Volume_{t+2}$	
	系数	t 统计量	系数	t 统计量	系数	t 统计量
截距项	0	-0.032	-5.468	-0.457	1.926	1.195
$\ln(1 + Msg_t)$	-0.001 ***	-3.253	0.439 **	2.209	-0.101 ***	-2.842
$AgrIdx_t$	-0.001	-1.205	0.276	0.55	-0.162 *	-1.694
$MsgBSI_t$	0	0.586	-0.149	-0.993	0.022	0.438
$MktRet_{t+2}$	0.013 ***	40.984				
$\ln(Mkt_{t+2})$			0.229	0.158	-0.215	-1.05
$\ln(Size_{t+2})$	0.001 *	1.906	0.552 ***	3.305	-0.021	-0.657
B/M_{t+2}	0	0.539	-5.822 ***	-9.686	0.029	0.883
$Return_{t+1}$	0.065 ***	5.187				
$Volatility_{t+1}$			0.554 ***	40.205		
$\Delta Volume_{t+1}$					0.706 ***	49.979
调整 R^2	0.296		0.528		0.485	

五、稳健性检验

(一) 排除公司公告、兼并收购、配股、增发、红利分配事件的影响

相对于分析师对个股的“买入”或“卖出”评级,虽然“中性”评级控制了相关基本面信息的影响,但股票论坛仍可能同时受到公司公告、兼并收购、配股、增发、红利分配等事件的影响。在排除了公司公告等基本面信息后,结果不变。类似的,未来两日的预测回归结果保持不变,限于篇幅没有列出。

(二) 控制周一效应、节假日后首日效应的影响

本文参照 Antweiler and Frank (2004) 设置一哑变量表示周一或节假日后的首个交易日来控制周一效应、节假日后首日效应所带来的影响,回归结果与本文第四部分的结论一致(限于篇幅没有列出)⁸。

六、结论与启示

本文从股吧论坛个股的帖子内容提取帖子情绪,结合证券分析师对个股的“中性”评级数据,检验了我国股票网络论坛的信息含量。主要研究结论如下:

(1) 股票当日收益率受当日论坛情绪影响,为显著正相关,说明帖子情绪有一定的信息含量。股票当日帖子数显著负向影响当日及未来两日的股票收益率,说明帖子数对股票收益率有一定的预测力。

(2) 股票当日的帖子数显著正向影响当日股票价格波动,而且还能影响未来两日的股价波动,从而拓展了国内以论坛帖子数或网络搜索量作为投资者关注代理变量的研究范围。

(3) 股票当日交易量受当日情绪一致性的影响,情绪越一致,交易量越大。但是,这种趋势在未来一日将反转,当日情绪分歧度越大,未来两日的交易量反而越大。

本文研究得到如下启示:在公司没有新的重大新闻、基本面信息情况下,股票网络论坛对股票市场的影响是不能忽视的。论坛作为一个大量股民言论的聚集处,为投资者情绪和投资者关注的获取和衡量提供了直接有效的途径。证券市场监管部门可利用这些信息来提升监管效率:(1) 重视论坛发帖量的监测。证券市场监管部门可以在股票网络论坛这种相对容易监测的环境中获得对未来股票价格波动的先知信号,可在股票价格出现剧烈波动之前采取适当措施,防范股价暴涨、暴跌的风险。(2) 重视论坛帖子情绪的监测。从帖子情绪涨跌一致性对股票交易量和收益率的影响来看,一致性高的情绪导致交易量放大,收益率降低,说明投资者非理性情绪较浓,通过情绪监测,可防范论坛中的股票被操控的风险。

⁸ 如需详细结果,可来函索取。

参 考 文 献

- [1]董大勇和肖作平 2011,《证券信息交流家乡偏误及其对股票价格的影响:来自股票论坛的证据》,《管理世界》第 1 期,第 52 ~ 61 + 188 页。
- [2]胡昌生和池阳春 2013,《投资者情绪、资产估值与股票市场波动》,《金融研究》第 10 期,第 181 ~ 193 页。
- [3]姜富伟、涂俊、David E. Rapach、Jack K. Strauss 和周国富 2011,《中国股票市场可预测性的实证研究》,《金融研究》第 9 期,第 107 ~ 121 页。
- [4]李心丹、王冀宁和傅浩 2002,《中国个体证券投资者交易行为的实证研究》,《经济研究》第 11 期,第 54 ~ 63 页。
- [5]刘红忠和何文忠 2012,《中国股票市场上的“隔夜效应”和“午间效应”研究》,《金融研究》第 2 期,第 155 ~ 167 页。
- [6]饶育蕾和盛虎,2010,《行为金融学》机械工业出版社。
- [7]施荣盛和陈工孟,2012,《网络时代的股市“自行车定理”——基于股票论坛数据及分析师评级的研究》,《上海金融》第 7 期,第 68 ~ 73 页。
- [8]王美今和孙建军 2004,《中国股市收益、收益波动与投资者情绪》,《经济研究》第 10 期,第 75 ~ 83 页。
- [9]杨晓兰、沈翰彬和祝宇 2016,《本地偏好、投资者情绪与股票收益率:来自网络论坛的经验证据》,《金融研究》第 12 期,第 143 ~ 158 页。
- [10]俞庆进和张兵 2012,《投资者有限关注与股票收益》,《金融研究》第 8 期,第 152 ~ 165 页。
- [11]张雪兰和何德旭 2008,《证券分析师利益冲突影响投资者利益吗——一个经验研究评述(1995 - 2007)》,《金融研究》第 7 期,第 170 ~ 183 页。
- [12]赵龙凯、陆子昱和王致远 2013,《众里寻“股”千百度》,《金融研究》第 4 期,第 183 ~ 195 页。
- [13]Agrawal, Anup, and Mark A. Chen. 2008. “Do Analyst Conflicts Matter? Evidence from Stock Recommendations”, *Journal of Law and Economics* 51(3) : 503 ~ 537.
- [14]Antweiler, Werner, and Murray Z. Frank. 2004. “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”, *Journal of Finance* , 59(3) : 1259 ~ 1295.
- [15]Barber, Brad M. , and Terrance Odean. 2008. “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors”, *Review of Financial Studies* , 21(2) : 785 ~ 818.
- [16]Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. “A Model of Investor Sentiment”, *Journal of Financial Economics* , 49(3) : 307 ~ 343.
- [17]Cao, H. Henry, Joshua D. Coval, and David Hirshleifer. 2002. “Sidelined Investors, Trading - Generated News, and Security Returns”, *Review of Financial Studies* , 15(2) : 615 ~ 548.
- [18]Das, Sanjiv R. , and Mike Y. Chen. 2007. “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web”, *Management Science* , 53(9) : 1375 ~ 1388.
- [19]De Long, J. B. , A. Shleifer, L. Summers, and R. Waldmann. 1990. “Positive Feedback Investment Strategies and Destabilizing Rational Speculation” *Journal of Finance* , 45(2) : 375 ~ 395.
- [20]Hirshleifer, Jack. 1977. “The Theory of Speculation under Alternative Regimes of Markets” *Journal of Finance* 32(4) : 975 ~ 999.
- [21]Huang, Y. , H. Qiu, and Z. Wu. 2016. “Local Bias of Investor Attention: Evidence from China’s Internet Stock Message Boards”, *Journal of Empirical Finance* 38 338 ~ 354.
- [22]Jackson, A R. . 2005. “Trade Generation, Reputation, and Sell - Side Analysts”, *Journal of Finance* , 60(2) : 673 ~ 717.
- [23]Kahneman, D. 1973. “Attention and Effort”, Englewood Cliffs, NJ: Prentice - Hall.

- [24]Milgrom , Paul , and Nancy Stokey. 1982. “Information ,Trade and Common Knowledge” , *Journal of Economic Theory* , 26(1) : 17 ~ 27.
- [25]Sabherwal , Sanjiv , Salil K. Sarkar , and Y. Zhang. 2011. “Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News” , *Journal of Business Finance & Accounting* , 38 (9) : 1209 ~ 1237.
- [26]Seasholes , Mark S. , and G. Wu. 2007 . “Predictable Behavior , Profits , and Attention” , *Journal of Empirical Finance* , 14 (5) : 590 ~ 610. Seasholes ,
- [27]Tetlock , Paul C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market” , *Journal of Finance* , 62(3) : 1139 ~ 1168.
- [28]Zhang , Y. and P. Swanson. 2010. “Are Day Traders Bias Free? —Evidence from Internet Stock Message Boards” , *Journal of Economics and Finance* , 34(1) : 96 ~ 112.

Analysis on the Information Content of China's Internet Stock Message Boards

DUAN Jiangjiao LIU Hongzhong ZENG Jianping

(School of Business ,University of Shanghai for Science and Technology;
School of Economics/School of Computer Science , Fudan University)

Abstract: In this paper , we select stock message board in Guba East – money , and extract sentiment from messages using computer text processing methods. We examine the information content of China's Internet stock message boards in the event of analysts' neutral recommendations. We find that contemporaneous stock returns are significantly positively related with sentiment. And the number of messages significantly negatively affects stock returns of the day and next two days. We also find that the number of messages significantly positively affects contemporaneous price volatility ,and it also affects the volatility of the day and next two days. Greater disagreement among the posted messages induces more trading volume on the next two days. Hence , this research can not only provide new insight into the mechanism through which stock messages boards affect market , but also provide helpful guidance for stock market regulators.

Key words: Investor Sentiment , Sentiment Disagreement , Stock Return , Price Volatility , Trading Volume

(责任编辑: 李景农) (校对: LN)