

社会互动对股票市场的影响

——基于新浪财经博客的实证分析

杨晓兰 高媚 朱淋

(浙江大学经济学院, 浙江 杭州 310027)

摘要: 本文以2013年12月~2015年11月新浪财经博客共14万多条博文为研究对象, 利用计算机文本挖掘技术定量描述投资者基于博客进行社会互动的程度及其情绪倾向, 检验社会互动对股票市场的影响。实证研究结果表明博主的互动诉求、博主的影响力、互动覆盖范围和互动程度对当期沪深300指数的收益率和成分股成交量有显著正向影响。以2015年6月股指最高点为分界线将样本分为上涨期和下跌期, 我们发现社会互动对收益率的影响在指数处于下跌期时程度更大。除了沪深300指数之外, 社会互动对上证综合指数和创业板指数的收益率都存在显著正向影响。此外, 社会互动各指标对波动率均没有显著影响。

关键词: 社会互动; 财经博客; 股票市场; 文本挖掘

Abstract: In this paper, we study a total of more than 140,000 blogs posted on the web of Sina Finance and Economics from December 2013 to November 2015. The method of text mining is adopted to measure the degree and the sentiment of social interaction based on blogs. We investigate the impact of social interaction on the stock market. The empirical results show that the measures of social interaction have significant positive impacts on the return and trading volume of CSI 300. Dividing the sample period into bullish period and bearish period by June 2015 when the index reaching the highest point, we find that social interaction has a greater impact on return in bearish period than in bullish period. In addition to the CSI 300, the impacts of social interaction on the returns of SHCI and GEM are robust. Besides, there is no significant effect of social interaction on the stock volatility.

Key words: social interaction, blogs, stock market, text mining

作者简介: 杨晓兰, 女, 经济学博士, 浙江大学经济学院金融系副教授。高媚, 女, 浙江大学经济学院金融系博士生。朱淋, 女, 浙江大学经济学院金融系硕士生, 供职于花旗银行。

中图分类号: F830.9 **文献标识码:** A

引言

对于个体投资者而言, 股票投资是一项复杂的决策, 个体通常会面临知识匮乏和信息不完全的局限性。在这种情况下, 社会互动成为投资者交流信息、表达意见的一个重要方式。近年来有相当多的文献讨论了社会互动对金融决策行为的影响。Duflo和Saez(2003)^[4]的研究证实社会互动对个人购买养老金产品的决策行为产生影响; Hong等(2004)^[6]和李涛(2006)^{[15] [16]}的研究都表明社会互动会显著推动居民参与股市投资; Karlan(2005)^[7]则

提供了通过社会互动影响个体信贷决策的证据。

随着互联网的飞速发展, 投资者不仅可以在网络上获取信息, 而且可以主动发布信息, 与其他投资者互相交流。博客、微博、论坛等网络社区已经成为投资者之间进行互动的重要平台。互联网时代, 投资者社会互动的规模、频率、速度都出现了变革性的提升。基于互联网的社会互动对证券市场的发展是一把双刃剑。一方面, 互联网平台具有及时、覆盖面广的特征, 投资者之间的社会互动有助于提高市场信息的透明度, 由此提高证券市场的信息有效性; 另一方面, 与传统媒体的权威

性相比较,互联网平台具有匿名性、草根性的特点,极易成为谣言传播和主观情绪宣泄的主要渠道,投资者之间的社会互动会导致市场价格产生过度反应,从而影响市场的稳定性。研究基于互联网的投资者社会互动对股票市场的影响,在理论上有利于探索互联网时代资产定价的新特征,在实践中有助于为促进网络媒体规范运作和金融市场稳定发展提供政策依据。

本文以网络互动平台——新浪财经博客为例探究社会互动对股票市场的影响。与已有文献相比,本文的创新点体现在以下几个方面。第一,从研究对象上来看,与已有文献采用的调查和实验方法不同,我们用财经博客作为刻画投资者社会互动的媒介。虽然当前信息传播和交流的媒介形式多样,但财经博客在描述投资者社会互动上具有一定的优势。与股吧等投资者论坛相比,财经博客表达的是专业人士的观点见解,股票投资本身是一种专业行为,市场上大多数散户都缺乏相关专业知识,财经博客能够较好地刻画普通投资者如何通过专业博主之间的交流来获取相关信息;与报纸新闻或者电视新闻相比,报纸、电视仅仅只是单向的信息传播平台,具体的互动影响范围、互动的程度都无法观察,以博客为研究对象,可以通过阅读量、评论/转发/喜欢/收藏数或者博主的人气等角度来刻画互动的特征。第二,从研究方法来看,我们采用文本挖掘技术获取新浪财经博客网页上的相关信息,不仅统计博客阅读量等数值型数据,更进一步分析博客文本包含的情绪倾向,据此构造的变量比较全面地揭示了社会互动的程度和属性。第三,从研究的视角来看,已有文献对社会互动的研究,较多地分析了社会互动对个体决策行为的影响,例如对股票投资参与度、对个人信贷决策和购买养老金产品决策的影响,本文的研究则检验了社会互动对股票市场收益率、波动率和成交量的影响,揭示社会互动对股票市场产生的宏观效应。

文献综述

与本文主题相关的文献包括两个方面,一是社会互动;二是以微博、博客为代表的网络互动平台对股票市场的影响。

一、社会互动的涵义及其度量方法

社会互动是来自社会学的概念。巴克(1984)^[12]认为

社会互动是一方或多方的反应,取决于或依赖于另一方所说所做的程度,社会情景则随着这种程度而变化。Durlauf和Ioannides(2010)^[5]将社会互动的内涵从社会学范畴拓展到了经济学范畴,他们认为社会互动是指个体之间的相互依赖性,在这些相互依赖性作用之下,一个兼具社会和经济行为特性的人,其偏好、信念以及其所面临的预算约束都受到其他人的特征与选择的直接影响。

近年来,经济学的研究逐步开始讨论社会互动对经济行为的影响。社会互动这一概念的引入也给经济学研究带来了新的视角。Duflo和Saez(2003)^[4]采用实验的方法研究了社会互动对个人做出购买养老金产品决策的影响,结论表明这种影响是显著的。Hong等(2004)^[6]在控制了财富水平、种族、受教育程度和风险忍受程度后发现,那些经常与邻居互动交流、出席教会的人参与股市投资的可能性更大,并且这种社会互动的影响在那些股市参与度更高的地区更加显著。Hong等(2004)^[6]以及Karlan(2005)^[7]还分别提供了通过社会互动获得信息而影响基金经理持股行为和通过社会互动交流感受而影响个体信贷决策的证据。李涛(2006a)^[15]研究了我国社会互动与股市参与的关系,通过问卷调查采用“居民在2004年春节期间以各种形式给亲属、朋友以及认识的人拜年的总人数”和“居民对自己在社会上人际交往程度的主观评价”这两个指标来衡量社会互动,结果表明社会互动推动了居民参与股市。李涛(2006b)^[16]采用问卷的方式对2005年我国12个城市的居民进行了抽样调查,详尽调查了居民的13种常见个人投资项目(包括银行存款、外汇、股票等),对于个体以及个体周围的人是否投资于这13个项目分别设立虚拟变量,并对“个体做投资决策时是否依赖来自朋友的信息”、“个体投资过程中与朋友交流的重要性判断”等信息进行量化,结果表明社会互动推动了个体对这些投资项目的参与。朱光伟等(2014)^[18]利用中国家庭金融调查2011年数据,发现关系对家庭股市参与的正向促进的影响渠道之一是社会互动。他们按社区的股市参与率将社区划分为高、中、低参与率社区并设置为虚拟变量,发现高参与率社区和关系的交叉项显著为正,而低参与率社区和关系的交叉项显著为负,这意味着在股市参与率更高的社区,社会互动推动家庭参与股市的作用会更大。郭士祺和梁平汉(2014)^[14]研究社会互动对家庭股市

参与的影响,他们得到的结论与朱光伟等(2014)^[18]类似,并认为传统的社会互动正被网络信息渠道部分替代。

二、微博、博客对股票市场的影响

关于微博、博客影响的研究,由传媒学领域拓展开来的主要包括市场营销、金融学等领域。在市场营销领域,Mishne和Glance(2006)^[9]研究了博客情绪和电影票房之间的关系,结果表明无论是电影上映之前还是之后,博客对电影的讨论与电影的票房收入有显著正相关性,博客积极的情绪是电影票房成功的有效预测指标。Liu等(2007)^[8]构建了一个利用博客情绪信息和商品过去的销售表现来预测商品未来销售情况的模型。以电影票房收入为例,该模型的准确性和有效性均得到了验证。

金融学领域的研究方向主要是微博、博客对股票市场的影响。Zhang等(2011)^[11]分析每条Twitter包含的情绪,分为积极情绪和消极情绪,用包含情绪的Twitter数量占总Twitter数量的比例作为一个自变量指标,发现这个比例与道琼斯指数、纳斯达克指数和标普500指数均呈显著负相关,但与芝加哥期权交易所波动率指数显著正相关。类似地,Bollen等(2011)^[12]基于Twitter,使用谷歌提供的情绪挖掘工具OpinionFinder(OF)和Google-Profile of Mood States (GPOMS)来度量文本情绪。OF能实时量化当前网络上积极情绪和消极情绪的强度;GPOMS则能将情绪细分为六类(平静(calm),警惕(alert),确信(sure),至关重要(vital),和蔼(kind)和高兴(happy))。实证结果表明积极和消极情绪的比例仅对滞后一天的道琼斯指数变化产生显著影响,而六个维度的情绪序列中仅calm对滞后2到6天的道琼斯指数变化情况影响显著。

De Choudhury等(2008)^[3]专门针对4家美国科技类公司,分析了2007年1~11月之间关于这4家公司博客的特征,包括博客数量、评论数量、评论平均长度和回复时间、不同的评论回复者的角色等,结果表明量化这些博客的特征之后,它们与公司股价每天的变化率显著相关,在股价变动方向的预测上有87%的准确性,在股价变动规模上有78%的准确性。Ruiz等(2012)^[10]不仅关注股价的变动还研究了交易量的变动,发现Twitter自身特征与股票交易量之间的关系比与股价之间的关系更加显著。

岑维等(2014)^[13]通过深交所“互动易”平台这一类似微博网站的问答数据,将深市上市公司某月被投资者

提问的问题数作为投资者关注度的衡量指标,发现关注度越高,股票收益越高,股票波动率风险和流动性风险越低,表明深交所“互动易”这一投资者和上市公司之间的互动平台可以显著的降低信息不对称、稳定市场和保护投资者权益。张继勋和韩冬梅(2015)^[19]根据深交所“互动易”平台统计数据,通过实验向被试提供上市公司管理层回复及时性和明确程度的信息,及时回复组被试被告知管理层一般24小时内回复,不及时回复组被试被告知管理层一般在一两天回复,有时在一周之后,而管理层回复明确程度则是通过向被试提供在互动平台上以一问一答的形式出现的五组公司管理层回复投资者的实例。实验结果表明管理层回复越及时,回复明确程度越高,投资者的投资可能性越高。

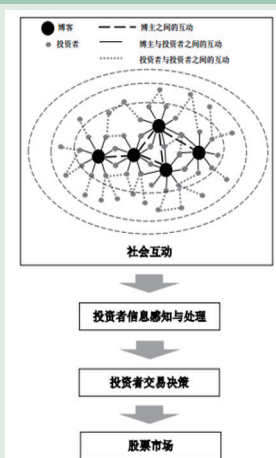
研究背景和理论假设

从互动的主体来看,目前国内典型的财经互动平台可以分为三种类型。第一类是上市公司与投资者之间的互动平台,比如深交所的“互动易”以及各上市公司的官方微博、微博,是网络时代提高上市公司信息透明度重要途径;第二类是投资者之间的互动平台,以东方财富网股吧为典型代表,股吧提供了双向信息交互方式,已成为中小投资者分享投资经验的重要平台,但也具有鲜明的草根特点,其中的信息表现出非理性情绪多、虚假信息多、无关信息多等特点(证券市场导报,2012)^[17]。第三类是财经名家或者专业投资者与普通投资者的互动平台,包括新浪、搜狐以及和讯的博客,专业人士作为博主发表财经评论,并与其他博主、普通投资者展开互动¹。本文的研究对象是第三类互动平台。

新浪博客自2005年上线,采用以明星吸引大量草根博客的策略,逐渐形成了分布广泛、基础庞大而稳定的用户群。根据Alexa中国官方网站2016年2月的及时数据,新浪博客的人均页面浏览量为3.69,而其主要竞争对手网易博客、搜狐博客的人均页面浏览量分别为1.35和1.67,新浪博客的知名度和活跃度在同类网络平台中处于领先地位。新浪财经博客汇集了财经界众多知名学者和专业投资者,每日刊发众多财经评论和观点。鉴于新浪财经博客具有较大的知名度和影响力,本文将其作为研究对象。

图1显示了基于博客的社会网络关系。根据社会网

图1 博客的社会网络关系与互动传播模式



络的分析方法，在一部分局部网络中，一些行动者具有“中心性”，他们居于中心地位并往往与他者有多种关联。我们认为博主便是这样一个在局部网络中处于中心地位的角色。基于博客的社会互动包含了三个层次：首先，不同博主之间存在相互关注、访问的关系；第二，投资者通过阅读博客、转载或评论等与博主产生直接交流关系；第三，不同的投资者之间存在信息沟通。从社会互动的内涵来看，以上三层关系均属于社会互动的范畴。本文所要研究的社会互动指上述的第二层关系，即投资者与博主的互动。

博主与投资者之间的互动对投资决策的影响可以通过以下两个渠道来实现。首先，财经博客的博主是财经界名人或者至少是具有一定财经专业知识的专业人士，大多数散户投资者都是缺乏相关专业知识的，因此，财经博客自然成为投资者十分倚重的信息源和决策基础，他们所传递的信息也往往被投资者当做具有权威性的意见来认知、传播，从而影响投资者的信息结构以及对信息的解读。其次，投资者在评论/转发博客时会对接接收的信息进一步深化，或者通过评论来表达自己的相反意见，评论/转发这一系列行为又会影响投资者的判断和决策。以上两个方面表明，互动会影响投资者对信息的感知和处理，并最终影响投资行为，也就是说，处于中心地位的博主以这样一种发散式网状互动模式，一层一层地对投资者产生直接、间接影响，从而通过投资者的交易行为持续作用于股票市场。

根据以上分析，本文提出的理论假设是以博客为平台的社会互动模式会对我国股票市场产生影响。

研究设计

一、样本选取与数据来源

我们选取新浪财经博客中与股票市场联系最直接、关联度最高的四个板块“独家看市”、“博客看市”、“板块/个股”以及“证券市场”作为博客样本。时间跨度确定为2013年12月18日~2015年11月27日，包含了股票指数逐步进入上升阶段，价格不断膨胀，随后猛烈下跌的周期。我们通过文本挖掘来获取每篇博客的文本信息、博客的阅读数、评论/转发/喜欢/收藏数、博主的关注人气等内容，并通过文本分类来获取每篇博客所传递的信号或者情绪，由此构建以日为单位的反映社会互动的各项指标。

股票市场的变量包括市场指数的收益率、波动率和成交量。数据均来自Wind数据库，收益率选用沪深300指数当日涨跌幅率，波动率选用沪深300指数近26周波动率，按照26周窗口滚动计算，计算公式为 $\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_x)^2}{n-1}}$ ，其中 x_i 是区间内对数收益率， μ_x 为平均收益率。

二、变量构造说明

为了获得新浪财经博客的相关数据，我们需要采用文本挖掘技术来摘取网页上的信息。使用Gooseeker网络爬虫软件扫描博客网站的底层代码，通过html标签来抓取原始数据。在原始数据清洗与整理之后得到1262条博主数据和142579条博文数据。博主数据的变量包括博主ID、博主主页、博主昵称、关注人气等；博文数据的主要变量是博客ID、博客标题、博客发表时间、阅读数/评论数/喜欢数/收藏数/转发数等。我们统计了关注度最高的10位博主信息，如表1所示。

我们以日为单位构建社会互动变量。变量构建步骤

表1 关注度最高的10位博主信息

博主ID	博主昵称	博主主页	关注人气
1216826604	wu2198	http://blog.sina.com.cn/u/1216826604	406482
1092672395	郎咸平	http://blog.sina.com.cn/u/1092672395	391305
1233227211	叶檀	http://blog.sina.com.cn/u/1233227211	374367
1284139322	凯恩斯	http://blog.sina.com.cn/u/1284139322	297865
1236135807	空空道人	http://blog.sina.com.cn/u/1236135807	284456
1458594614	淘金客	http://blog.sina.com.cn/u/1458594614	274412
1249424622	但斌	http://blog.sina.com.cn/u/1249424622	271987
1279916282	揭幕者	http://blog.sina.com.cn/u/1279916282	268361
1246151574	康斯坦丁	http://blog.sina.com.cn/u/1246151574	260584
1514551012	股道班长——短线王国	http://blog.sina.com.cn/u/1514551012	238765

如下：

第一步，按如下公式计算每日的博客数量、阅读数、评论/转发/喜欢/收藏数以及高人气博主所占比例。

$$\text{每日的阅读数} = \text{每日所有博客的阅读数之和} / \text{每日的博客数量} \quad (1)$$

$$\text{每日的评论/转发/喜欢/收藏数} = \text{每日所有博客的评论、转发、喜欢、收藏数之和} / \text{每天的博客数量} \quad (2)$$

每日高人气博主所占比例的计算方法是将1262位博主按照关注人气排序，选取关注人气排名前100的博主。我们将这100位博主定义为高人气博主，然后统计每天发表博客的博主中这些高人气博主所占比例。

第二步，利用文本分类方法，确定每日的情绪指数。

142579篇博客按照它所传递的信号被分为“积极”、“中立”和“消极”三类。由于博客正文篇幅过长，这里我们仅针对博客标题进行分类²。分类的过程主要如下：首先，从全体博客样本中随机抽取一小部分作为“训练数据集”进行人工分类。为了保证训练数据集分类客观、有代表性，我们从全体样本中抽取2000条博客，邀请三位金融学研究生共同进行人工分类：若认为博客标题传递的是积极信号或情绪则分类为“积极”；若博客标题传递的是消极信号或情绪则分类为“消极”；若是中性信号则分类为“中立”。取三人的众数作为最终结果，若三人的分类各异，则把该条博客归为“中立”。然后，计算机通过学习训练数据集的结果建立计算机自己的分类模型。我们使用应用最广泛的数据挖掘开源包Weka，将多种算法进行比较后选择正确率最高的KNN算法³，并选取参数K=5。最后，将训练得到的分类模型再反过来应用在训练数据集上进行验证，发现计算机对博客分类的正确率超过90%，这说明分类方法较为准确地判断了博客传递的信号⁴。

在对142579篇博客进行情绪分类之后，我们借鉴Antweiler和Frank(2004)^[1]的方法构建博客情绪指数。假设一天中传递“积极”信号的所有博客数量设为 $M_t^{positive}$ ，一天中传递“消极”信号的所有博客数量设为 $M_t^{negative}$ ，则每日的博客情绪指数构造如下：

$$E_t = Ln \left[\frac{1 + M_t^{positive}}{1 + M_t^{negative}} \right] \quad (3)$$

可以看出，该指数越大，表明当天博客整体传递的信号或情绪越积极，反之，越消极。

第三步，构造包含情绪特征的互动变量。

互动的主体是博主与投资者，从博主的角度，一方面我们需要考察众多博主的互动诉求如何，即每天有多少篇博客被发表；另一方面则是博主的影响力，在所有博客中，其博主为高人气博主的比例如何。从投资者的角度，一方面可以通过博客的阅读数来衡量互动覆盖范围的大小，另一方面可以通过博客的评论/转发/喜欢/收藏数来衡量互动程度的高低。除了上述衡量互动程度的指标之外，我们构造的互动变量还需要包含博客所传递的情绪倾向，也就是这种社会互动的属性是消极、积极还是中性。

结合第一步中计算出的每日的博客数量、阅读数、评论/转发/喜欢/收藏数和高人气博主所占比例，以及第二步中构建的每日博客情绪指数，将它们分别相乘取乘积，便可以确定最终的每日包含情绪的社会互动变量，如表2所示。

实证结果与分析

一、变量描述性统计

在我们的样本期内，共有475个交易日。表3是本文所使用的每日博客数据的统计情况。

475个交易日中，博客数量平均值为276，发表最多的一天达448篇。阅读数最多的一天平均每篇博客被阅读了34086次，阅读数最少的一天则是1754次。评论/转发/喜欢/收藏数中，平均值最大的是喜欢数76.57，表明投资者使用最多的一种互动形式是点击“喜欢”，就像在微博、微信等社交平台上点赞一样，由于最为方便快捷而

表2 社会互动变量的构造

变量名称	含义	计算方法
博主的互动诉求EN	包含情绪的博客数量	博客数量*情绪指数
博主的影响力EI	包含情绪的高人气博主比例	高人气博主所占比例*情绪指数
互动的覆盖范围ER	包含情绪的阅读数	阅读数*情绪指数
互动的程度EL	包含情绪的评论/转发/喜欢/收藏数	评论/转发/喜欢/收藏数*情绪指数

表3 每日博客原始数据描述性统计

变量	样本数量	平均值	标准差	最大值	最小值
博客数量	475	276.35	88.19	448	4
高人气博主比例	475	0.20	0.06	0.35	0
阅读数	475	11276.73	5730.81	34086	1754.37
评论数	475	9.23	12.80	203.83	0.95
转发数	475	4.75	4.68	56	0.92
喜欢数	475	76.57	37.92	206.09	13.09
收藏数	475	1.25	1.01	13.8	0.20

成为了投资者最喜欢的互动形式。一些博客关闭了评论功能或者禁止转载，但我们难以分辨究竟是博主关闭了这种功能还是没有人评论和转发，两种情况混在一起，会使得我们统计出来的数据偏低，因此在下文的分析中我们仅使用其中的“喜欢数”来作进一步的分析。高人气博主所占比例的均值为0.20，并且大多比例都介于0.15和0.30之间，较为稳定。

关于博客情绪指数，根据上文的构建方法可知，情绪指数大于0，说明当天的博客整体传递出积极信号，数值越大，积极信号越强烈；情绪指数等于0，传递中性信号；情绪指数小于0，则整体传递消极信号。本文的475个样本中，213天的数值小于0，平均值为-0.27；等于0的天数为16天；其余246天的情绪指数均大于0，平均值为0.29。各互动变量和股票市场变量的描述性统计，如表4所示。

为了能够直观地观察互动变量与股票市场的关系，我们选取其中的几组变量绘制成如图所示的变化曲线。图2的左轴是博主的互动诉求——包含情绪的博客数量，右轴是沪深300指数收益率。总体来看，两者呈现出一定的联动关系，收益率的上涨往往伴随着博主的互动诉求的增加。

表4 主要变量描述性统计

变量	样本数	平均值	标准差	最大值	最小值
博主的互动诉求EN	475	12.59	87.40	279.00	-254.52
博主的影响力EI	475	0.0000392	0.07	0.29	-0.30
互动的覆盖范围ER	475	234.64	4512.944	42097.58	-20372.67
互动的程度EL	475	1.79	27.74	173.19	-151.06
沪深300指数收益率Return	475	0.0011	0.0196	0.0671	-0.0875
沪深300指数波动率Volatility	475	0.2612	0.1511	0.7167	0.0861
沪深300指数成交量Volume(亿)	475	204.4903	151.5281	686.4	35.7

图2 博主的互动诉求与沪深300指数收益率

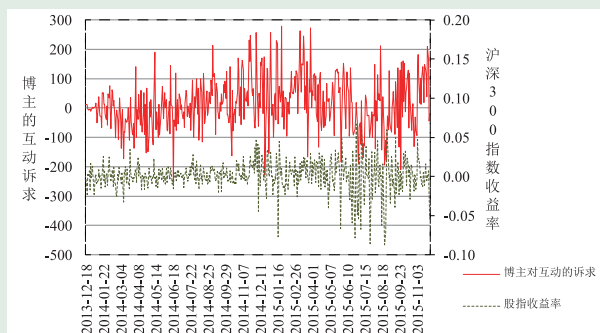


图3 博客阅读量与沪深300指数成交量

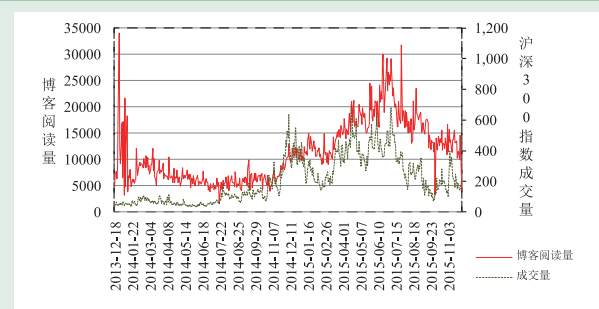


图3显示了博客阅读量与沪深300指数成分股成交量随时间变化的趋势。左轴代表博客阅读量，右轴代表成交量。成交量在2014年10月底开始陡增，随后的阶段博客阅读量与成交量存在明显的正向动态关系⁵，尤其是在2015年3~6月股市成交量迅猛增加的时期，投资者热情高涨，博客阅读量也显著上升。在2015年6~7月股灾发生的阶段，成交量迅速萎缩，而博客阅读量曾出现急剧上升的情况，这意味着市场的高度不确定性引发了大量投资者的关注，人们不愿意相信牛市终结了。至2015年10月成交量和博客阅读量双双跌落到低点，这表明投资者对市场已经失去信心。

二、社会互动对股票市场的影响

本文中股票市场的相关变量是以时间序列的形式呈现的，因此我们采用GJR-GARCH(1,1)模型来分析社会互动对股票市场的影响。

我们使用Stata计量软件。首先，对各个互动变量和股票市场变量序列进行ADF平稳性检验，结果显示，所有互动变量均为平稳时间序列，沪深300指数收益率为平稳序列(收益率单位为%)，而波动率和成交量取其一阶差分Dvolatility和Dvolume后平稳。接着，我们检验是否存在ARCH效应。以沪深300指数收益率和包含情绪的博客数量为例，先建立简单的线性模型：

$$Return_t = \lambda_0 + \lambda_1 EN_t + \varepsilon_t \quad (4)$$

BG检验的p值为0.0409，在5%显著性水平下拒绝“无自相关”的原假设，即认为存在线性的自相关关系，因而修改上述简单线性模型为：

$$Return_t = \lambda_0 + \lambda_1 EN_t + \lambda_2 Return_{t-1} + \lambda_3 Return_{t-2} + \varepsilon_t \quad (5)$$

BG检验的p值为0.2564，无法拒绝“无自相关”的原假设，即认为不存在线性的自相关关系。然后对该线性模型做ARCH效应检验：采用最小二乘法OLS估计原方

程,进而对残差平方序列 $\{\varepsilon_t^2\}$ 进行Q检验,结果显示存在ARCH效应。

最后,我们建立GJR-GARCH(1,1)模型,具体形式如下:

$$\text{均值方程: } \text{Return}_t = \lambda_0 + \lambda_1 \text{EN}_t + \lambda_2 \text{Return}_{t-1} + \lambda_3 \text{Return}_{t-2} + \varepsilon_t \quad (6)$$

$$\text{方差方程: } \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1' \sigma_{t-1}^2 + \beta_1 d_{t-1} \varepsilon_{t-1}^2 \quad (7)$$

其中, λ_1 表示包含情绪的博客数量对沪深300指数收益率的影响,衡量了博主的互动诉求是否会影响到市场指数的收益率。 α_1 、 γ_1' 和 β_1 则分别是ARCH项、GARCH项和TARCH的系数,其中, d_{t-1} 为虚拟变量,当随机扰动项 ε_{t-1} 小于0(表示负冲击)时, d_{t-1} 等于1,反之等于0,因此,若 β_1 大于0,表示负的冲击对收益率的影响要大于同等强度的正向冲击,小于0则表示正向冲击对收益率的影响要大于同等强度的负冲击。这里,我们需要重点考察的是系数 λ_1 。

对于其他几组变量,采用上述同样的方法检验ARCH效应并构建模型。

表5、6、7和8显示了四个互动变量分别与沪深300指数收益率、波动率和成交量回归的结果(其中收益率的单

表5 博主对互动的诉求EN对股票市场的影响

	变量	Return	Dvolatility	Dvolume
均值方程	EN	0.0065*** (8.26)	-0.0009 (-0.81)	0.0345** (2.54)
	因变量滞后一期	-0.0135 (-0.28)		-0.2627*** (-4.87)
	因变量滞后二期	-0.0716* (-1.65)		-0.1775*** (-3.30)
	因变量滞后三期			-0.1745*** (-3.37)
	因变量滞后四期			-0.0686 (-1.20)
	因变量滞后五期			-0.1342** (-2.51)
	因变量滞后六期			-0.1159** (-2.29)
	因变量滞后七期			-0.1305** (-2.49)
	因变量滞后八期			-0.0903** (-2.07)
方差方程	ARCH(1)	0.0872*** (4.87)	0.3184*** (3.21)	0.0975** (2.20)
	GARCH(1)	0.9135*** (66.76)	0.3416** (2.05)	0.8398*** (36.76)
	TARCH(1)		-0.2828*** (-2.78)	0.1806*** (3.02)
	Log likelihood	-854.1176	-931.7603	-2311.967
	Wald chi2	70.87***	0.65	43.19***

注:1.系数下边括号里的值为z统计量,***、**、*分别表示1%、5%和10%水平上显著,下同;2.经检验,该组结果中以收益率Return为因变量的回归,其TARCH系数不显著,因此我们将模型简化为GARCH(1,1),下述表格中若TARCH项系数空缺表明对模型进行了简化;3.若因变量存在自相关,则需要AR(p)的基础上再构建GARCH模型,经检验三个因变量中收益率和成交量需要引入自回归项,分别滞后2期和8期,使得均值方程不存在自相关性。

位取%)。

表5的结果显示,包含情绪的博客数量EN对沪深300指数收益率在1%显著性水平上具有正向影响、对成交量在5%显著性水平上具有正向影响,但对于波动率的影响不显著。这表明财经博客的博主作为互动的主体,他们的互动诉求越强烈,包含情绪的博客数量越大,那么这种互动效应就越能影响到市场指数收益率和成交量的提高。

从表6可以看出,包含情绪的高人气博主所占比例与沪深300指数收益率与成交量都呈显著正相关关系,而与波动率关系不显著。说明互动主体的影响力会在一定程度上增加市场收益、扩大成交量。

表7和表8的结果显示,互动覆盖范围与沪深300指数收益率和成交量均显著正相关,与波动率负相关但是不显著;互动程度同样与收益率和成交量正相关并且显著,同时与波动率负相关但不显著。我们知道,互动覆盖范围和互动程度是分别由阅读数和喜欢数与博客情绪指数的乘积项来衡量的,对于同样的阅读数或喜欢数,若情绪越积极,或者对于同样的情绪指数,若阅读数或喜欢数越高,互动的效应就会越大,这种效应带来了股市收益率和成交量的增加。

此外,表5、6、7和8的结果中,波动率方程中的

表6 博主的影响力EI对股票市场的影响

	变量	Return	Dvolatility	Dvolume
均值方程	EI	5.3455*** (5.54)	-0.1759 (-0.10)	28.9874** (2.07)
	因变量滞后一期	0.0074 (0.14)		-0.2649*** (-4.89)
	因变量滞后二期	-0.0498 (-1.09)		-0.1792*** (-3.31)
	因变量滞后三期			-0.1797*** (-3.45)
	因变量滞后四期			-0.0732 (-1.27)
	因变量滞后五期			-0.1373** (-2.58)
	因变量滞后六期			-0.1197** (-2.33)
	因变量滞后七期			-0.1310** (-2.50)
	因变量滞后八期			-0.0891** (-2.07)
方差方程	ARCH(1)	0.0833*** (5.48)	0.3741*** (3.32)	0.0989** (2.27)
	GARCH(1)	0.9169*** (77.77)	0.3459** (2.37)	0.8392*** (36.82)
	TARCH(1)		-0.3339*** (-2.94)	0.1789*** (3.07)
	Log likelihood	-874.2592	-932.0398	-2312.487
	Wald chi2	33.25***	0.01	39.33***

TARCH(1)项均显著为负，而成交量方程中的TARCH(1)项均显著为正，表明对波动率来说，正向冲击的影响比同等强度的负向冲击的影响更大，即好消息对波动率的波动性的影响大于坏消息；而对成交量来说，负向冲击的

表7 互动覆盖范围ER对股票市场的影响

	变量	Return	Dvolatility	Dvolume
均值方程	ER	0.0001*** (17.26)	-0.0000 (-0.73)	0.0007* (1.68)
	因变量滞后一期	0.0088 (0.18)		-0.2576*** (-4.82)
	因变量滞后二期	-0.0440 (-1.03)		-0.1735*** (-3.21)
				-0.1742*** (-3.36)
				-0.0653 (-1.13)
				-0.1361** (-2.50)
				-0.1193** (-2.31)
				-0.1327** (-2.48)
				-0.0960** (-2.18)
方差方程	ARCH(1)	0.0829*** (5.32)	0.3161*** (3.15)	0.0855** (2.00)
	GARCH(1)	0.9161*** (75.23)	0.3535** (2.11)	0.8495*** (38.26)
	TARCH(1)		-0.2843*** (-2.78)	0.1778*** (3.16)
	Log likelihood	-872.7834	-931.7318	-2323.133
	Wald chi2	101.82***	0.53	39.64***

表8 互动程度EL对股票市场的影响

	变量	Return	Dvolatility	Dvolume
均值方程	EL	0.0210*** (8.47)	-0.0035 (-0.70)	0.0995* (1.76)
	因变量滞后一期	0.0031 (0.06)		-0.2590*** (-4.84)
	因变量滞后二期	-0.0511 (-1.19)		-0.1746*** (-3.23)
				-0.1734*** (-3.31)
				-0.0628 (-1.10)
				-0.1346** (-2.48)
				-0.1182** (-2.29)
				-0.1300** (-2.44)
				-0.0922** (-2.09)
方差方程	ARCH(1)	0.0873*** (5.02)	0.3251*** (3.19)	0.0879** (2.04)
	GARCH(1)	0.9121*** (66.74)	0.3536** (2.15)	0.8486*** (38.47)
	TARCH(1)		-0.2936*** (-2.83)	0.1744*** (3.08)
	Log likelihood	-867.0564	-931.8067	-2312.998
	Wald chi2	75.79***	0.49	40.08***

影响比同等强度的正向冲击的影响更大，即坏消息对成交量的波动性的影响大于好消息。

三、稳健性检验

为了进一步考察社会互动对股票市场的影响同时对上述结果进行稳健性检验，我们按照沪深300指数点位分组回归、更换因变量为上证指数和创业板指数再进行回归。

根据2014~2015年沪深300指数的具体情况来看，2015年6月8日到达最高5335.12点，此前指数呈明显上涨期，而此后呈明显下跌期。因此，我们以2015年6月8日为节点，将总体样本分为两个部分分别对收益率进行回归：5335.12点之前，359个交易日，为上涨期；5335.12点之后，116个交易日，为下跌期。回归结果如表9所示(此处仅报告均值方程中社会互动变量的系数情况)。

上文总体样本的回归的四个互动变量均显著，此处分组的结果与总体样本保持一致。再进一步观察回归结果，下跌期样本的系数均大于上涨期样本。例如，对于互动程度EL，EL每增加(减少)一单位，下跌期样本对指数收益率的影响是低指数点位样本的1.87倍。由此可见，在我国股市2015年6月以后出现的暴跌阶段，面对股市高度的不确定性，投资者之间社会互动对股票市场产生了更为显著的影响。

表10是以上证指数收益率和创业板指数收益率为因变量的估计结果。我们将上文的沪深300指数的结果同时列出进行比较。可以看出，四组因变量回归的结果不论是互动变量的系数方向还是显著性水平均保持一致，只是系数大小不同。总体来看，社会互动对创业板指数的影响略大一些，这可能是处于成长阶段的创业板更容易受到社会互动这种外部因素的作用。

表9 互动变量对沪深300指数收益率影响的分组回归结果(均值方程)

变量	上涨期(5335.12点之前)	下跌期(5335.12点之后)
EN	0.0056*** (7.60)	0.0108*** (4.15)
EI	4.7251*** (5.24)	15.9632*** (3.98)
ER	0.00009*** (7.69)	0.00027*** (5.44)
EL	0.0174*** (7.25)	0.0326*** (4.62)

表10 互动变量对沪深300指数、上证指数、创业板指数收益率的影响(均值方程结果)

变量	沪深300指数收益率	上证指数收益率	创业板指数收益率
EN	0.0065*** (8.26)	0.0063*** (9.21)	0.0063*** (7.10)
EI	5.3455*** (5.54)	4.9118*** (6.14)	5.3973*** (4.84)
ER	0.0001*** (17.26)	0.0001*** (9.99)	0.0001*** (8.37)
EL	0.0210*** (8.47)	0.0197*** (8.60)	0.0208*** (7.54)

结论

本文以新浪财经博客作为社会互动的方式,实证检验社会互动对股票市场的影响。从互动的主体——博主和投资者两个角度出发,我们利用文本挖掘技术分析了2013年12月~2015年11月期间博客文本的情绪倾向,构造了四个变量来描述和衡量投资者的社会互动。具体包括博主的互动诉求、博主的影响力、互动覆盖范围和互动程度。实证结果显示博主的互动诉求、博主的影响力、互动覆盖范围和互动程度对市场指数的收益率和成交量存在正向影响。样本期间我国股票市场发生了剧烈的震荡,股指呈现倒U型走势。以2015年6月股指达到顶点为分界线,我们发现社会互动对收益率的影响程度在指数处于下跌期时比上升期时更大。与沪深300指数和上证

指数相比,创业板指数收益率受到互动的的影响更大。此外,博主的互动诉求、博主的影响力、互动覆盖范围和互动程度对波动率均没有显著影响。

本文的结论为网络互动对股票市场产生的影响提供了证据。基于上述结论,我们认为在互联网时代,监管当局应该规范网络互动平台,杜绝不实谣言,并积极引导投资者理性、健康地参与社会互动,在长期内营造和谐的、有助于推动经济发展与改革的社会互动氛围。而对于财经博客的博主来说,作为一名财经专业人士,应该理性表达自己的观点态度,不能为了吸引读者眼球、彰显个性、提升知名度而刻意夸大说法;此外,还应该避免传递盲目乐观或者过分悲观的信号及情绪,以此引导投资者行为趋于理性,为市场的发展产生积极的作用。 ■

[基金项目:国家自然科学基金青年项目(项目编号:71303210)]

注释

1. 近年来兴起的雪球网兼具第二类和第三类平台的特征,即有实名认证的知名专业人士,也有普通投资者。
2. 通常情况下,投资者在浏览博客时,正文部分不一定会看完,或者仅仅浏览博客列表,那么此时投资者接收到的信息几乎完全来自博客标题,因此我们仅针对标题分类也具有现实合理性。
3. KNN算法的基本原理是:如果一个样本与K个最相似的样本中的大多数属于某一个类别,则该样本也属于这个类别。例如,在K=5时,有某篇待分类的博客,我们找出它在训练数据集

中特征最相近的5篇博客,如果其中有3篇博客都是积极类,则该篇博客也被分类为积极。

4. 实际上,计算机对博客分类的错误大多来自将一部分传递积极或消极信号的博客归为了中立类型,而混淆积极和消极的错误是很少出现的。具体结果显示:训练数据集中被人工分类为“消极”的博客,其中3.12%被错误地归为了积极;而人工分类为“积极”的博客,其中0.50%被错误地归为了消极。

5. 如果剔除2013年12月26日博客阅读量的异常点,阅读量高峰均出现在2015年6月之后。

参考文献:

- [1] Antweiler W, Frank M Z. Is all that talk just noise? The information content of internet stock message boards[J]. The Journal of Finance, 2004, 59(3): 1259-1294.
- [2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market[J]. Journal of Computational Science, 2011, 2(1): 1-8.
- [3] De Choudhury M, Sundaram H, John A, et al. Can blog communication dynamics be correlated with stock market activity?[C]. Proceedings of the nineteenth ACM conference on Hypertext and hypermedia. ACM, 2008: 55-60.
- [4] Duflo E, Saez E. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment*[J]. The Quarterly journal of economics, 2003, 118(3): 815-842.
- [5] Durlauf S N, Ioannides Y M. Social Interactions[J]. Annual Review of Economics, 2010, 2(1): 451-478.
- [6] Hong H, Kubik J D, Stein J C. Social interaction and stock-market participation[J]. Journal of finance, 2004, 59(1): 137-163.
- [7] Karlan D S. Using experimental economics to measure social capital and predict financial decisions[J]. American Economic Review, 2005: 1688-1699.
- [8] Liu Y, Huang X, An A, et al. ARSA: a sentiment-aware model for predicting sales performance using blogs[C]. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 607-614.
- [9] Mishne G, Glance N S. Predicting Movie Sales from Blogger Sentiment[C]. AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. 2006: 155-158.
- [10] Ruiz E J, Hristidis V, Castillo C, et al. Correlating financial time series with micro-blogging activity[C]. Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012: 513-522.
- [11] Zhang X, Fuehres H, Gloor P A. Predicting stock market indicators through twitter “I hope it is not as bad as I fear” [J]. Procedia-Social and Behavioral Sciences, 2011, (26): 55-62.
- [12] 巴克. 社会心理学[M]. 天津: 南开大学出版社, 1984.
- [13] 岑维, 李士好, 童娜琼. 投资者关注度对股票收益与风险的影响——基于深市“互动易”平台数据的实证研究[J]. 证券市场导报, 2014, (07): 40-47.
- [14] 郭士祺, 梁平汉. 社会互动、信息渠道与家庭股市参与——基于2011年中国家庭金融调查的实证研究[J]. 经济研究, 2014, (S1): 116-131.
- [15] 李涛. 社会互动, 信任与股市参与[J]. 经济研究, 2006, (01): 34-45.
- [16] 李涛. 社会互动与投资选择[J]. 经济研究, 2006, (08): 45-57.
- [17] 证券市场导报编辑部. “股吧”对监管的挑战及其应对[J]. 证券市场导报, 2012, (03): 1.
- [18] 朱光伟, 杜在超, 张林. 关系、股市参与和股市回报[J]. 经济研究, 2014, (11): 87-101.
- [19] 张继勋, 韩冬梅. 网络互动平台沟通中管理层回复的及时性、明确性与投资者投资决策——一项实验证据[J]. 管理评论, 2015, (10): 70-83.