

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

杨晓兰 沈翰彬 祝宇

(浙江大学经济学院, 浙江杭州 310027)

摘 要: 本文以投资者在东方财富网股吧针对创业板上市公司发表的 90 多万条帖子为研究对象, 通过 IP 地址识别, 构建本地关注指标; 并利用计算机文本挖掘技术, 提取网络发帖所体现的情绪倾向, 构建投资者情绪指标。实证结果显示, 本地关注对股票收益率的影响取决于投资者情绪, 当投资者持积极情绪时, 本地关注对股票收益率有显著的正向影响; 当投资者持消极情绪时, 该影响显著为负。在积极情绪和消极情绪下本地关注对股票交易量都有显著正向影响, 但积极情绪下的影响程度比消极情绪下更大。此外, 本地关注与投资者情绪的交叉效应在上市公司样本数量较多的北京、广东、江苏、上海、浙江这五个区域都存在, 并与整体样本的特征基本一致, 但在样本数量较少的其他区域并不完全一致。

关键词: 本地偏好; 投资者关注度; 投资者情绪; 股票收益率; 网络论坛

JEL 分类号: G01, G02, G12 **文献标识码:** A **文章编号:** 1002-7246(2016)12-0143-16

一、引言

根据马科维茨的投资组合理论, 完全理性的投资者可以通过股票配置的全球化实现最优投资组合。然而, 现有的研究显示不仅在国际金融市场上投资者会过多关注并超额配置本国公司的股票 (French and Poterba, 1991), 而且在本国的股票市场上, 投资者也会更多地关注和配置离自己比较近的上市公司 (Coval and Moskowitz, 1999; Seasholes and Zhu, 2010; 赵静梅等, 2012; Huang et al., 2016)。这种现象被称为投资者的本地偏好 (local bias)。

现有文献研究的“本地偏好”具体有两种表现形式, 一是过度配置, 指在资产配置上

收稿日期: 2014-07-03

作者简介: 杨晓兰, 通讯作者, 经济学博士, 副教授, 浙江大学社会科学基础研究平台, 浙江大学经济学院, Email: yxlan@zju.edu.cn.

沈翰彬, 经济学硕士, 杭州城投管理集团有限公司, Email: kgjr107@163.com.

祝宇, 博士研究生, 浙江大学经济学院, Email: ross.zhu@outlook.com.

* 本文感谢国家自然科学基金青年项目“基于文本挖掘的网络信息与股票市场关联机制研究”(71303210) 和浙江大学金融研究院 (AFR) 的资助。感谢匿名审稿人的宝贵意见。文责自负。

投资者更多地购买本地股票;二是过度关注,投资者更多地关注本地股票。由于市场上存在成千上万的投资者,他们的资产配置情况和交易动机往往无法直接观测,如何度量本地偏好,如何检验本地偏好与股票市场的联动性在已有文献中并没有给出清晰的答案。更为关键的是,从本地关注的角度来看,现有理论和实证文献仅局限于分析本地关注的程度,尚未细分关注的属性,也就是没有区分“好的”关注与“坏的”关注。事实上,当投资者对股价走势持悲观态度时,对上市公司的关注度越高,例如对公司丑闻等负面消息的关注越高,则可能意味着会对股价产生负面冲击,这与正面关注对股价的影响是存在差异的。

本文以我国最大的网络财经论坛——东方财富网股吧为研究对象,研究本地关注和投资者情绪的交叉效应,即检验在消极情绪和积极情绪两种不同的情况下,投资者对本地股票的过度关注对股票价格以及成交量产生的影响。东方财富网股吧是一个有关股票信息交流的互动社区,按照我国深沪两市所有股票的名称分成了若干子社区,论坛的发帖表达了投资者对特定股票的意见、情绪,每一条发帖数据还包含了时间、IP 地址等信息。本文利用股吧的发帖数据构建两个重要指标:第一是通过挖掘发帖的文本信息,构建文本所体现的投资者情绪指数;第二是通过统计发帖人的 IP 地址,判断发帖人所属区域,构建衡量投资者本地偏好的指标——本地关注指数。通过检验本地关注指数、投资者情绪指数及其二者交叉对股票收益率的影响,本文为解释投资者本地偏好现象及其效应提供了直接的证据。

与现有文献相比,本文的创新性体现在以下几个方面:第一,我们利用文本挖掘技术提取投资者情绪指标,利用 IP 地址判断投资者所属区域,首次区分了消极情绪与积极情绪两种不同情形下,本地关注对股票市场产生的效应;第二,本文的研究样本是代码为 300001-300356 的 355 只全部创业板股票^①,公司所在地覆盖了我国大部分省级区域,由此我们不仅能够检验本地关注对股票市场的影响在整体样本层面上呈现的效应,而且还能检验和比较该效应在各区域体现出的差异。

本文的研究主要包括五个部分,第二部分是文献综述,第三部分提出理论假设,第四部分是研究方法 with 样本基本情况介绍,第五部分是实证研究结果,最后是结论。

二、相关文献回顾

本文的研究主题是投资者本地关注以及投资者情绪对股票市场的交叉影响。与本文密切相关的文献来自投资者关注度、本地偏好、投资者情绪这三个研究领域。

投资者关注度研究以个体的有限理性为假设前提。Peng and Xiong(2006)构建了一个“注意力分配模型”研究投资者关注度的分配与资产定价的动态关系。模型显示由于投资者的关注度是有限的,投资者只能更多地关注于市场和行业的信息,而不是个股信息,从而导致股票价格相互联动。Barber and Odean(2008)和 Seasholes and Wu(2007)发

^① 代码 300060 没有对应的股票。

现关注度对股票价格有显著的正向影响。

最早提出地区因素会对投资者行为产生影响这一观点的学者是 French and Poterba (1991)。他们针对美国、英国、日本、法国和德国这五个国家的投资者的研究发现,绝大多数投资者(包括机构投资者)将大部分资金用于购买本国资产。Coval and Moskowitz (1999)的研究以美国基金经理的交易信息作为样本,实证研究发现绝大多数机构投资者会超额配置那些公司总部与他们更加接近的公司的股票。Huang et al. (2016)利用东方财富网股吧论坛的发帖数据,用 IP 识别方法判断发帖人所在地,证实了本地关注现象的存在,检验了公司特征、区域经济发展程度等因素对本地关注的影响。

已有文献证实了本地偏好的存在性,但是检验本地偏好是否会对股票收益率产生影响存在较大的困难。这是因为研究者无法观察到在公司层面,一个上市公司股票的购买者有多少来自本地,有多少来自外地,从而难以建立公司股票被本地投资者持有的比例与股票收益率之间的关系。董大勇和肖作平(2011)以东方财富网股吧发帖人的 IP 地址计算股票的本地偏好,他们的实证研究结果表明股票论坛中投资者以更大的概率参与本地股票信息的交流,信息交流的本地投资者比例越大,股票价格越高。

投资者情绪对金融资产价格的影响是近年来兴起的行为金融学研究的一个重要主题。与本文研究主题相关的系列文献利用网络媒体挖掘投资者情绪,并检验投资者情绪对资产定价的影响。例如,Antweiler and Frank(2004)运用朴素贝叶斯算法将来自雅虎财经的 150 万条帖子分成看空、看多、持平(包括噪音)三类,据此构建了看涨情绪指标。研究结果发现情绪指标对股价收益有显著正向影响。Bollen et al. (2010)利用情绪分析工具对微博(Twitter)进行文本分析,得出微博文本体现的情绪指标,发现情绪指标对美国道琼斯指数的波动率具有预测作用。金雪军等(2013)以东方财富网股吧为研究对象,通过文本挖掘方法提取了投资者情绪指数,验证了投资者情绪对股票收益率的正向影响。

综上所述,现有投资者关注度以及本地偏好现象的文献认为在个体注意力有限、区域股票供给有限等假设前提下,本地关注度越高,基于关注度的购买行为增加,从而股票收益率越高;来自网络论坛投资者情绪领域的文献则清晰地揭示出投资者对股票的关注、讨论具有一定的情绪倾向,投资者体现出来的积极情绪对股价有正面影响,消极情绪则负向影响股价。现有文献较少从关注度和情绪的交叉作用来分析本地偏好现象,这也正是本文的创新之处。

三、研究假设

本文提出的理论假设是投资者本地关注与投资者情绪会对股票市场产生交叉影响。该理论假设主要基于 Peng and Xiong(2006)的注意力分配模型,从信息处理的角度出发,该模型认为当投资者将注意力越分配到特定公司之上,该公司的股价会更加表现出过度反应的特点。因此,在投资者情绪乐观的情况下,本地关注程度越高,投资者会更积极地参与购买本地股票,导致股票价格的正向过度反应;而在投资者情绪悲观的情况下,本地

关注度越高,已经拥有本地股票的投资者抛售股票的可能性越大,股票价格将表现出负向的过度反应。据此我们提出本文的假设 1。

假设 1: 投资者本地关注对股票价格的影响取决于投资者情绪,在投资者对股票持积极情绪时,本地关注正向影响股价;持消极情绪时,本地关注则负向影响股价。

从交易量的角度来看,积极情绪下的本地关注会引发基于关注度的购买行为,带来成交量的上升。消极情绪下,一方面,本地关注会刺激股票持有者卖出股票,带来成交量的上升;另一方面,随着股票价格的下跌,在处置效应的影响下,部分投资者也不愿意卖出已经亏损的股票。总体上,消极情绪下,本地关注对成交量的正面影响较为有限。据此,本文提出待检验的假设 2。

假设 2: 在积极情绪下本地关注对股票交易量有正向影响;相对于积极情绪,消极情绪下,本地关注对股票交易量产生的正向效应较小。

在我国,区域之间的经济发展程度和结构存在一定差异,区域内上市公司的稀缺性各不相同,投资者的数量、投资者的风险偏好、信念以及收入水平等因素存在异质性,本地关注对股价产生的效应也可能呈现差异性。据此,我们提出假设 3。

假设 3: 本地关注对股票价格及成交量的影响存在区域差异。

四、样本与数据处理方法

本文研究的基本思路是通过对东方财富网股吧发帖的 IP 地址识别,构建投资者本地关注指标,并利用对帖子文本内容的情绪分析,构建投资者情绪指数以衡量投资者对每只样本公司在每个交易日的情绪倾向。

(一) 数据来源

本文以 356 只在深圳交易所创业板上市的公司(股票代码 300001—300356)为样本,样本期间为 2012 年 9 月 18 日至 2013 年 10 月 15 日。我们根据上市公司注册地所在的省级行政区域来划分本地投资者。例如,浙江的投资者对于任何注册地在浙江的上市公司来说都属于本地投资者。

本文选择创业板股票为研究样本。创业板股票具有这样几个特征:第一,公司规模较小,知名度不高;第二,许多公司的主营业务属于新兴产业;第三,公司的日常生产经营活动大多集中于公司所在地附近。这三个特征使得区域之外的投资者对上市公司的了解程度和了解意愿都比较低,而本地投资者对这些本地公司的了解会高于外地公司,自然也就会对本地公司的股票多加关注,并影响他们的交易决策。例如,相对于中国工商银行、中石化等家喻户晓的大规模知名企业而言,创业板股票的本地化特征更为明显。

本文的数据分为两部分,一部分是有关本地偏好和投资者情绪的网络论坛数据,另一部分为公司股票的交易数据。第一部分网络文本数据来自于我国最大的有关股票的网络互动社区——东方财富网股吧。我们编写网络爬虫程序抓取了从 2012 年 9 月 18 日直至 2013 年 10 月 15 日为止有关样本股票股吧发帖的全部网页文本信息。此外,我们剔除

了回帖、所有周末和节假日、个股停牌日、个股交易数据不全等特定情况下的帖子,最终获得90多万条主贴的文本内容。第二部分市场交易数据则来自于国泰安CSMAR数据库。

(二) 数据处理与变量描述

本地关注指标是本文的核心解释变量,来源于对股吧发帖IP地址的识别。目前在股吧上的发帖分为两种形式,一种是发帖人注册后以注册名进行发帖,第二种是匿名发帖。只有在匿名发帖的情况下,发帖人的IP地址才会显示。为了对发帖人IP地址进行识别,在统计本地关注指标时我们去掉所有以注册名发帖的帖子,仅保留所有匿名帖子,共计50.8747万条^②。我们将将这些帖子的IP地址数据与IP地址数据库的信息相匹配,可以识别出该IP地址来自于何地,从而判断出该条发帖是否为本地投资者所发布的信息。若为本地投资者发帖,则记 $N_{i,t,m}=1$;否则记 $N_{i,t,m}=0$,其中 i 表示特定某上市公司股票, m 表示这是当天的第 m 条帖子, t 表示发帖日期。在确定某一帖子的发帖人归属地以后,我们将某只股票 i 第 t 日本地发帖人的发帖比例 $Local_ratio$ 作为该股票当日的本地关注的代理指标,该指标的具体计算方法为:

$$Local_ratio_{i,t} = \sum_{m=1}^M N_{i,t,m} / M_{i,t} \quad (1)$$

其中 $M_{i,t}$ 表示股票 i 在第 t 日的发帖总数。

由于各省人口数量 and 经济发展程度存在差异,分配到的IP地址也存在差异, $Local_ratio$ 高可能来源于某地的网民数量规模较大。为了对此进行校正,我们获取了中国各省份的IPV4地址的分配比例 $IPV4_Ratio$ 作为校正标准,以消除不同省份之间因为IPV4地址的数量不同带来的本地发帖比例的差异。调整后的本地关注指标 Ln_Local 计算方法如下:

$$Ln_Local_{i,t} = \ln(1 + Local_ratio / IPV4_Ratio_{prov}) \quad (2)$$

其中 $prov$ 表示股票 i 在第 t 日的公司注册地所属省级区域。

投资者情绪指标也是本文的重要变量。我们借助文本挖掘工具Weka中的KNN算法将全部90多万条帖子(包括匿名和非匿名的帖子)照其所体现的情绪分成三类“积极”、“中立/噪音”、“消极”。用计算机进行文本分类的主要步骤包括:第一、从全体样本中随机抽出2000条帖子作为“训练数据集”进行人工分类。由三名被试独立进行人工分类,最后取三人的众数作为结果,若三人的分类结果都不一样,则该帖归为中立/噪音类。第二,计算机通过学习训练数据集的结果建立起自己的分类模型。我们使用数据挖掘开源包Weka,在将多种算法进行对比后最终选择正确率最高的KNN算法,并选参数 $K=5$ (K 为相邻样本数)。第三,我们将训练得到的分类模型反过来应用于训练数据集上进行验证,发现计算机对帖子分类的正确率为72%。实施这三个步骤的结果是我们实现了对每一条帖子的情绪进行判断,即“积极”、“中立/噪音”、“消极”这三类。

^② 在本文收集的全部90多万条帖子中,匿名帖子为50多万条,比例超过一半。由于非匿名帖子没有IP地址,我们在计算本地关注指标时不得不剔除所有非匿名帖子,该指标能够准确反映本地投资者发帖占全部发帖比例的假设前提是各个区域之间投资者匿名发帖与非匿名发帖的比例没有显著差异。

我们参考 Antweiler and Frank(2004) 的处理方法,将 t 日某只股票 i 论坛中所有的情绪为“积极”的帖子的数量记为 $M_{buy\ i\ t}$,所有的“消极”帖子的数量记为 $M_{sell\ i\ t}$ 。 t 日股票 i 的情绪指数变量 Sentiment 计算公式如下:

$$Sentiment_{i\ t} = \ln [(1 + M_{buy\ i\ t}) / (1 + M_{sell\ i\ t})] \quad (3)$$

从 Sentiment 的计算公式可以得知,当一天中某只股票积极情绪的帖子数量等于消极情绪的帖子数量时, Sentiment 为零;当积极情绪帖子数量高于消极情绪帖子数量时,该值大于零,反之则小于零。为了区分在积极情绪和消极情绪两种不同的情况下本地关注对股票市场的影响,我们构建了如下两个交叉变量:

$$Cross_negative = Ln_local * Negative_S \quad (4)$$

$$Cross_positive = Ln_local * Positive_S \quad (5)$$

其中,虚拟变量 Negative_S 代表消极情绪,即当情绪指数小于零时,该虚拟变量为 1;类似地, Positive_S 是代表积极情绪的虚拟变量。当这两个虚拟变量都为零时,则代表着中性情绪。消极情绪、积极情绪与本地关注程度的交叉变量——Cross_negative 和 Cross_positive 一方面反映了本地关注度的程度大小,另一方面也分别代表了关注度的情绪属性。

本文所采用的主要核心变量及其计算方法如表 1 所示。

表 1 变量汇总表

变量名	变量描述
Ret	个股日收益率。用考虑现金红利再投资的个股日收益率 dretwd 对数化处理: $Ret = \ln(1 + dretwd)$
Ln_Trading volume	个股成交量。个股某日的成交量取自然对数。
Ln_local	本地关注指标。每一天本地投资者的发帖数量占该股票当天发帖数量的百分比为 Local_ratio;再经过各省市 IPV4 地址段比例调节后,并取自然对数得到。
Sentiment	投资者情绪指标。由股票 i 在交易日 t 看涨帖子数量 $M_{buy\ i\ t}$ 和看跌帖子数量 $M_{sell\ i\ t}$ 决定。
Cross_negative	本地关注与投资者消极情绪虚拟变量的交叉项指标。
Cross_positive	本地关注与投资者积极情绪虚拟变量的交叉项指标。
Market_return	市场收益指标。创业板指数某交易日的收益率。
Market_trading	市场成交量指标。创业板市场某交易日的总成交量取自然对数

五、实证检验

(一) 描述性统计

本文的样本公司分布在全国 27 个省级行政区域。从上市公司的区域分布来看,主要

集中在广东、北京、江苏、浙江和上海这五个区域，其他大部分区域拥有的上市公司数量都在 10 家以下。表 2 显示了上市公司数量超过 10 家的样本区域的基本情况。

表 2 部分区域样本公司数量与帖子数量

省份	样本公司数量	发帖总量	平均每家公司发帖量
北京	53	86858	1 638.83
福建	11	11420	1 038.18
广东	76	109821	1 445.01
湖北	11	15372	1 397.45
湖南	11	13669	1 242.64
江苏	42	56132	1 336.48
山东	18	23078	1 282.11
上海	28	33252	1 187.57
浙江	36	46381	1 288.36

为了展示每个省级区域投资者本地关注的基本情况，利用每家上市公司在每个交易日的本地关注数据，我们汇总了每个区域的平均本地发帖比例，以及经过该省份 IPV4 地址调整后的本地关注指标。图 1 显示了各区域本地关注程度的差异，图中区域颜色的深浅对应着“本地发帖比例/该省份 IPV4 地址比例”的大小，该指标越大表示对本地股票发帖的比例高于本地实际 IP 地址比例的倍数越大，也就是本地关注的程度越显著。图中显示，本地关注程度最高的三个区域是贵州、江西和天津，本地发帖比例均超过 IP 地址比例的三倍。本地关注程度高与这些区域上市公司数量少有密切的关系。上市公司在本区域具有稀缺性，因而引发本地更多的关注，也就是本地投资者以更大的比例讨论本地股票。

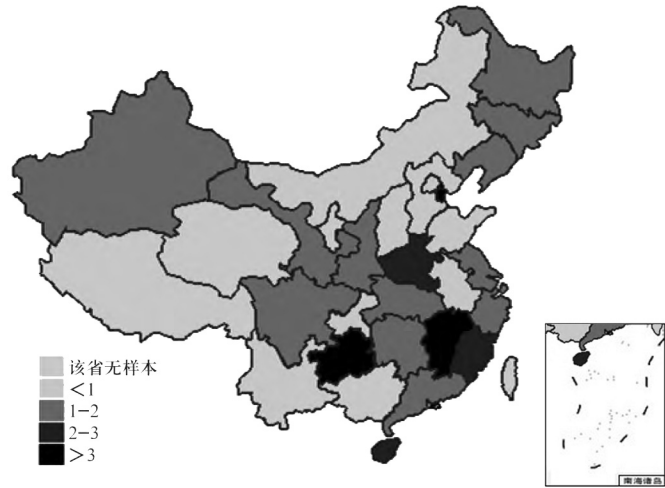


图 1 各区域本地关注程度

表 3 显示了主要变量的统计性表述。在全部样本中,情绪指数 Sentiment 的均值是 0.3178,具体分布情况是有 17953 个观测值等于零,30226 个观测值大于零,12146 个观测值小于零。这表明在样本期间积极情绪占据的比例最大。表 4 是主要变量的相关系数表。从我们的核心变量股票收益率、股票成交量、本地关注、投资者情绪来看,四个变量两两之间都存在显著的正向关系。本地关注、投资者情绪对股票收益率和成交量究竟产生怎样的影响还需要通过回归方程来做进一步检验。

表 3 变量描述性统计

变量名	样本量	均值	标准差	最小值	最大值
Ret	60325	0.0016	0.0310	-0.1059	0.0962
Ln_Trading volume	60325	14.4423	1.0324	10.0395	18.5665
Ln_local	60325	0.3603	0.7560	0	5.4305
Sentiment	60325	0.3178	0.7653	-2.9444	3.9318

表 4 相关系数表

	Ret	Ln_local	Sentiment	Ln_Trading volume
Ret	1.0000			
Ln_local	0.0092*	1.0000		
Sentiment	0.2520*	0.0159*	1.0000	
Ln_Trading volume	0.1236*	0.0425*	0.1481*	1.0000

注: * 表示在 5% 的水平下显著。

(二) 投资者本地关注、投资者情绪对股票价格以及成交量的解释能力

为了检验本地关注、投资者情绪对股票市场的影响,我们借鉴 Antweiler and Frank (2004) 的方法分别构建了以个股收益率和成交量为因变量的回归方程,其中收益率方程控制了创业板指数收益率,交易量方程控制了市场成交量。方程的估计方法采用面板数据固定效应模型,并以样本公司聚类进行标准差的估计。

表 5 显示了股票收益率回归结果。在控制市场因素影响的情况下,回归方程(1)表明本地关注对股票收益率的正向影响在 1% 的水平下显著,本地关注的系数为 0.000451。本地关注系数绝对值较小,其原因在于被解释变量——收益率的数值很小,均值仅为 0.0016。从影响力来看,如果本地关注指标提高一个单位,收益率将提高 0.000451,相对于均值为 0.0016 的收益率而言,提高的幅度为 28.13%。从回归方程(2)来看,投资者情绪指数对股价的正向影响在 1% 的水平下显著,这支持了已有文献的结论,即在互联网时代,通过网络论坛文本挖掘得到的投资者情绪指数是影响股票定价的一个不可忽视的因

素。回归方程(3)引入了本地关注指数、本地关注与积极情绪的交叉变量以及消极情绪的交叉变量^③。根据方程估计结果,两个交叉变量的系数都在1%的水平下显著。当Cross_positive等于1,且Cross_negative为零时,是投资者持有积极情绪的情形,此时本地关注指标的总效应为0.0027(−0.000633+0.00335);当Cross_negative为1,且Cross_positive为零时,代表投资者持消极情绪,本地关注指标的总效应约为−0.0037(−0.000633−0.310)。回归方程(3)得出的结论支持了本文提出的假设1,即投资者关注度对股价收益率的影响取决于投资者情绪。在积极情绪下本地关注越高,股价收益率越高,而消极情绪下,本地关注则具有负向效应。比较而言,消极情绪时本地关注对股价产生的负面冲击程度略高于积极情绪下产生的正向冲击。

表 5 个股收益率回归结果

	(1)	(2)	(3)
Market_return	1.032 *** (91.98)	0.993 *** (93.32)	1.026 *** (92.70)
Ln_local	0.000451 *** (3.23)		−0.000633 *** (−3.25)
Sentiment		0.00734 *** (35.08)	
Cross_positive			0.00335 *** (11.32)
Cross_negative			−0.00310 *** (−9.12)
常数	−0.000553 *** (−9.48)	−0.00263 *** (−34.52)	−0.000555 *** (−9.39)
样本量	60325	60325	60325
R ²	0.373	0.404	0.377

注: 括号中是 t 值, **代表 5% 显著, ***代表 1% 显著。

表 6 显示了对成交量的回归结果。回归方程(4)中,本地关注指数的系数在1%的水平下显著为正,这意味着在不区分情绪倾向的情况下,在总体样本层面,本地关注度对成交量产生正向影响。回归方程(5)表明投资者情绪指数对成交量也有显著的正向影响,投资者对某只股票持有的意见越乐观,股票成交量相应更高。回归方程(6)展示了积极情绪和消极情绪分别与本地关注交叉作用影响股票成交量的结果,在投资者持有积极情绪时(Cross_positive=1,且Cross_negative=0),本地关注每提高一个单位,对成交量的影

③ 回归方程(3)同时引入了本地关注指标Ln_local、交叉变量Cross_negative、Cross_positive。对于大部分上市公司而言,Ln_local在每个交易日都是各不相同的,Cross_negative、Cross_positive、Ln_local之间不存在完全的线性关系。为了检验回归方程(3)是否存在多重共线性,我们计算了各自变量的方差膨胀因子(VIF),Ln_local、Cross_negative、Cross_positive的VIF分别是2.88、1.59、2.43,均低于10,即可认为三个自变量的同时引入没有导致多重共线性。

响为约为 $0.103(-0.0296 + 0.133)$ 个单位;在消极情绪时($Cross_negative = 1$,且 $Cross_positive = 0$) ,本地关注上升带来的成交量上升的幅度为 $0.053(-0.0296 + 0.0822)$ 。这一结果显示本地关注对成交量的影响在积极情绪和消极情绪下都是正向的 ,但是影响程度存在差异 ,积极情绪下本地关注引发的成交量上升幅度比消极情绪下更高 ,即本文提出的假设 2 成立。

表 6 个股成交量回归结果

	(4)	(5)	(6)
Market_trading	0.118 *** (38.65)	0.116 *** (38.07)	0.117 *** (38.57)
Ln_local	0.0526 *** (7.66)		-0.0296 *** (-3.88)
Sentiment		0.0903 *** (19.27)	
Cross_positive			0.133 *** (13.91)
Cross_negative			0.0822 *** (8.24)
常数	13.21 *** (415.44)	13.20 *** (416.35)	13.21 *** (416.66)
样本量	60325	60325	60325
R ²	0.368	0.373	0.372

注: 括号中是 t 值 , **代表 5% 显著 , ***代表 1% 显著。

表 5 回归方程(3)和表 6 回归方程(6)显示出在投资者情绪中性的情况下($Cross_positive = 0$,且 $Cross_negative = 0$) ,本地关注对收益率和成交量都有显著负面影响。这一结果可以用 Fang and Peress(2009)提出的“未被报道溢价(no media premium)”来解释。该理论认为 ,在信息不完全的市场上 ,由于投资者并不知晓所有的股票 ,关注度低的股票不能被充分套利 ,反而具有更高的收益率。也就是说本地投资者的过度关注会增加股票套利活动 ,导致相应股票的收益率下降。在区分投资者情绪的情况下 ,我们的实证结果显示 ,如果投资者对股票持有明显的乐观或者悲观的情绪 ,本地关注度相应会给股价及成交量带来正面或者负面的冲击;而在投资者对股票持中性态度时 ,本地关注度低的股票反而有更高的收益率 ,体现出“未被报道溢价”。

目前有关投资者关注度研究的已有实证文献没有得到一致性的结论 ,有的支持关注度对收益率的正向影响 ,有的则认为关注度越高的股票 ,收益率反而越低。产生不一致结论的原因很多 ,例如实证方法的差异 ,指标构造的差异等。然而 ,其中可能存在的一个重要原因是已有文献没有区分关注度的属性 ,这导致如果样本期间投资者更多地呈现出积极情绪时 ,关注度对股价的正向效应得到证实 ,反之则是负面效应。本文的研究检验了本

地关注度与投资者情绪对股票市场的交叉影响,更为清楚地揭示出积极情绪与消极情绪下,本地关注度对股票收益率的影响具有不对称性,对成交量的影响也存在程度上的差异。

(三) 本地偏好、投资者情绪交叉效应的分区域检验

前文的研究证实了在整体层面上,本地偏好与投资者情绪的交叉作用对股票市场呈现显著的影响。本部分将进一步检验这种效应在我国不同的区域内是否依然成立。根据区域上市公司数量的多寡,我们将各区域分为两大类,一类是上市公司样本大于 20 的 5 个区域(北京、广东、江苏、上海、浙江),另一类是上市公司样本数量小于 20 的其他 22 个区域。划分这两类区域的出发点是因为整体样本呈现的交叉效应同时包含时间和个股两个维度,当我们把一个区域作为一个相对独立的市场时,只有当该区域上市公司样本数量超过一定规模才有可能体现这两个维度的效应。在一些区域内,上市公司数量很少,实证模型的结果在更大程度上是代表时间维度上的效应,即在这家上市公司本地关注度越高的那些交易日,收益率是否相应变化越大。我们将相继检验在大样本区域和小样本区域本地偏好与投资者情绪的交叉效应。

表 7 显示了利用五个大样本区域数据对收益率进行的回归的结果,回归方程同样采用了面板数据固定效应模型,并以上市公司为聚类计算标准差。从各区域的回归结果中我们得到这样一些结论:首先,各区域回归方程中包括本地偏好、本地偏好与积极情绪交叉、本地偏好与消极情绪交叉在内的各自变量的系数在符号上是一致的,这表明本地偏好在积极情绪下对收益率有正向影响,在消极情绪下对收益率有负面影响在各个区域都是存在的;其次,区域之间的差异性体现在消极情绪下本地偏好对收益率的负向影响程度上,从 Cross_negative 的系数上看,北京和广东的系数高于其他三个区域,且在 1% 的水平下显著,江苏和上海的显著性水平为 5%,而浙江为 10%。

本地偏好来源于投资者对本地公司的非理性热爱。从我们的结论可以推测在积极情绪的影响下,这五个区域投资者对本区域公司的非理性热爱程度接近,对股价产生类似的正向推动作用。然而,各区域投资者在消极情绪下对本地公司的抛售行为可能存在差异,差异产生的原因可能是对经济环境以及市场的长期走势存在异质性信念,也可能是风险态度存在差异。

表 7 收益率分区域回归结果(大样本区域)

	北京	广东	江苏	上海	浙江
Market_return	1.1020 *** (47.08)	1.0714 *** (43.13)	1.0015 *** (34.07)	0.9825 *** (22.24)	1.0277 *** (27.85)
Ln_local	-0.0014 (-1.13)	-0.0007 ** (-2.03)	-0.0007 (-1.56)	-0.0005 (-0.92)	-0.0015 *** (-2.87)
Cross_positive	0.0077 *** (4.11)	0.0050 *** (8.14)	0.0031 *** (3.92)	0.0028 *** (4.69)	0.0034 *** (4.4)

续表					
	北京	广东	江苏	上海	浙江
Cross_negative	-0.0061 *** (-3.18)	-0.0050 *** (-8.28)	-0.0018 ** (-2.4)	-0.0023 ** (-2.32)	-0.0017 * (-1.99)
常数	-0.0002 ** (-2.04)	-0.0006 *** (-4.2)	-0.0011 *** (-5.73)	-0.0001 (-0.52)	-0.0001 (-0.47)
样本量	8702	13447	7332	4426	6098
R ²	0.392	0.3945	0.3720	0.3751	0.3775

注: 括号中是 t 值, **代表 5% 显著, ***代表 1% 显著。

表 8 是五个区域对成交量进行分区域回归的结果,采用面板数据固定效应模型和公司聚类的标准差计算方法。表 8 得到的结论与表 7 的结论是相互印证的。我们可以看到,在消极情绪下,与北京、广东相比,在江苏、上海、浙江三个区域,关注度上升对成交量的影响显著性程度不高,影响系数也较小,也就是说负面情绪和本地关注度交叉作用没有触发非常明显的抛售行为。这与表 7 显示的这三个区域在投资者持负面情绪时,本地关注度上升导致收益率下降的程度不如北京、广东显著的结论是一致的。

表 8 成交量分区域回归结果(大样本区域)

	北京	广东	江苏	上海	浙江
Market_trading	0.1225 *** (16.92)	0.1255 *** (20.01)	0.1066 *** (14.19)	0.1290 *** (14.56)	0.1235 *** (11.69)
Ln_local	-0.0880 * (-1.82)	-0.0401 *** (-2.87)	-0.0195 (-1.22)	-0.0068 (-0.31)	-0.0164 (-0.84)
Cross_positive	0.3499 *** (6.06)	0.1809 *** (8.37)	0.1192 *** (4.97)	0.0990 *** (4.06)	0.1221 *** (5.37)
Cross_negative	0.1380 *** (2.27)	0.1267 *** (5.98)	0.0784 ** (2.66)	0.0495 * (2.03)	0.0557 ** (2.52)
常数	13.3115 *** (171.69)	13.1555 *** (203.72)	13.1777 *** (177.57)	13.0123 *** (139.39)	13.0283 *** (122.51)
样本量	8702	13447	7332	4426	6098
R ²	0.2797	0.3181	0.2298	0.3212	0.2516

注: 括号中是 t 值, **代表 5% 显著, ***代表 1% 显著。

按照与表 7、表 8 同样的方法我们对小样本区域数据进行了回归分析。表 9 汇总了收益率和成交量的回归结果(不包括云南省)。我们重点关注两个交叉变量的影响。从收益率回归结果来看,在 20 个区域中,与全国整体样本一致,有 15 个区域 Cross_positive 的系数为正,有 16 个区域 Cross_negative 的系数为负,这表明大部分区域本地偏好与投资

者情绪对股票市场的交叉影响方向与整体样本是一致的,但是大部分区域交叉变量的系数并不显著。类似地,成交量的回归结果显示,大部分区域两个交叉变量对成交量的影响与整体样本一致为正,但是并不显著。

与整体样本和大样本区域的结果相比较,我们认为部分小样本区域本地偏好效应与整体样本结果不一致或者结果不明显的一个可能原因是公司数目较少。关注度理论认为关注度对股票定价的影响在于投资者过度关注一些股票,而忽视了另一些股票。如果一个区域内上市公司数量很少,从时间维度来观察,每家上市公司得到的本地关注度可能是比较稳定的,因为没有足够多的本地股票供投资者选择关注一部分,并忽视另一部分,在这种情况下本地关注度变化与收益率之间的关系就很可能不显著^④。表 9 显示,在拥有 10 家以上上市公司的福建、湖北、湖南、山东这四个区域中,除了湖南省的交叉系数 Cross_positive 之外,其他交叉系数的符号与大样本区域的结果完全一致。这从一定程度上表明,样本数量越多,结果的一致性随之提升。

然而,表 9 的结果并没有表明上市公司样本数量的增加必然带来本地关注与投资者情绪交叉效应的显著性提升,例如只拥有 1 家上市公司的吉林省,交叉变量对收益率的影响是显著的,而拥有 18 家上市公司的山东省,该效应并不显著。这意味着除了样本数量之外,各区域投资者的风险态度、家庭收入、对区域经济的预期等因素都可能存在差异,这都将导致分区域检验的结果存在差异。与小样本区域相比,北京、广东、江苏、上海、浙江这五个大样本区域得到与整体样本较为一致的结论,其原因一方面区域内上市公司数目较多,另一方面是这五个区域的经济发展程度也比较接近。

表 9 收益率、成交量分区域回归结果(小样本区域)

省份	收益率			成交量			公司数量
	Ln_local	Cross_positive	Cross_negative	Ln_local	Cross_positive	Cross_negative	
贵州	0.0067	-0.0029	-0.0141 **	0.0599	-0.0160	-0.0469	1
黑龙	-0.0017	0.0042	-0.0034	0.0840	0.0563	-0.1570	1
吉林	-0.0425 ***	0.0487 ***	0.0469 ***	1.1105 ***	-1.0829 ***	-1.1673 ***	1
甘肃	-0.0043 ***	0.0021	-0.0012	-0.0867 ***	0.0926	0.0446	2
海南	0.0057	-0.0048	-0.0062	0.0721	-0.0366	-0.0367	2
山西	0.0018 **	0.0060	-0.0007	0.1385 **	0.0284	0.0635	2
江西	-0.0005	0.0038 *	0.0026	-0.0807	0.0775	0.1205	3

④ 一个极端的例子是云南省,在样本期间只有一家创业板上市公司,在 104 个交易日里本地关注度指标有 103 个为零,因此无法观察关注度变化对收益率和成交量的影响。表 9 的回归方程里并未包括云南省。

续表

省份	收益率			成交量			公司数量
	Ln_local	Cross_ positive	Cross_ negative	Ln_local	Cross_ positive	Cross_ negative	
内蒙	0.0126 ***	-0.0168 **	-0.0227 **	0.1691 *	-0.1917 **	-0.1335	3
新疆	-0.0011	0.0095	-0.0095	-0.0983	0.1745	0.2157	3
重庆	-0.0011	0.0028	-0.0105 ***	-0.0675	0.2152	0.2706 *	4
河北	-0.0009	0.0029	-0.0010	-0.1204 **	0.2193 ***	0.2806 ***	5
天津	-0.0004	0.0021 ***	-0.0029	0.0191	0.0745 *	0.0875	5
陕西	-0.0015	0.0024	0.0008	-0.0817 ***	0.0804	0.1356 ***	6
安徽	0.0029	-0.0009	0.0009	-0.0002	0.1003	-0.0454	7
四川	0.0018	0.0020	-0.0045	0.0096	0.0685	0.1055	7
河南	-0.0009	0.0025	-0.0038 *	-0.0182	0.1177 ***	0.0519	8
辽宁	-0.0007	0.0017	-0.0031	0.0024	-0.0320	-0.1266 *	8
福建	-0.0013	0.0031 **	-0.0014	-0.0307	0.1538 ***	0.1180 *	11
湖北	-0.0004	0.0019	-0.0004	-0.0420 **	0.1001 **	0.0290	11
湖南	0.0017	-0.0009	-0.0039 *	-0.0188	0.1007 *	0.0930	11
山东	-0.0003	0.0024	-0.0027	-0.0285	0.0893 **	0.0402	18

注: 回归方程与表 7、表 8 完全一致, 本表仅报告了三个主要自变量的回归结果。*代表 5% 显著, **代表 1% 显著。

六、结 论

本文利用东方财富网股吧对创业板股票在样本期的发帖, 进行 IP 地址识别, 构建了本地投资者关注度的指标, 反映在每个交易日针对每只股票的讨论有多大的比例来自上市公司注册地的本地投资者。利用计算机文本挖掘技术, 我们提取帖文所包含的情绪倾向, 构建了投资者情绪指数。通过实证检验, 我们发现投资者本地关注现象对股票的影响依赖于投资者情绪。这一结论意味着当投资者情绪低迷、市场看跌的时候, 本地偏好与投资者情绪的共同作用会使本地股票加速下跌; 而当投资者情绪高涨、市场看涨时, 这种共同作用又会助涨本地上市公司的股票。这一结论支持了 Peng and Xiong (2006) 的理论模型, 表明投资者对个股的关注程度越高, 股价越可能产生过度反应。此外, 通过分区域检验, 我们发现本地关注度对股票市场的影响在北京、广东、江苏、上海、浙江这五个区域都存在, 尤其是在积极情绪下, 本地关注对股票收益率和成交量都产生显著的正向影响, 但

消极情绪下本地关注产生的负效应在程度上存在一定的差异。在除此之外的其他小样本区域中,本地关注以及本地关注与投资者情绪的交叉效应并未全部得到验证。区域差异产生的可能原因是不同省份上市公司数量、经济发展程度存在差异以及投资者之间存在异质性,导致本地投资者的关注度影响交易行为的程度存在差异。

投资者对本地股票的过度关注、偏好是投资者存在的一种固有的心理偏差,源于投资者注意力的局限性以及对非本地股票存在模糊性厌恶等原因。本文的实证研究结果揭示了本地关注度对股票市场的影响依赖于投资者情绪。也就是说,本地关注对交易行为的影响在投资者情绪悲观和乐观两种不同状态下产生的效应并不相同。此外,本文对本地关注和投资者情绪变量的度量都来自网络论坛,这表明在网络时代,利用网络论坛数据挖掘投资者行为对于认识和预测我国股票市场有着重要的意义。

本文研究存在一定的局限性。例如,我们发现了本地关注对股票定价的影响在各区域存在差异,这种差异可能来源于投资者的异质性,但我们尚没有对各区域投资者的异质性进行验证。这也是未来进一步研究的方向。

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页。
- [2]金雪军, 祝宇和杨晓兰. 2013,《网络媒体对股票市场的影响——以东方财富网股吧为例的实证研究》,《新闻与传播研究》第12期,第36~51页。
- [3]赵静梅, 吴风云和罗梅. 2012,《投资决策中地区偏好与地区规避的实证研究》,《投资研究》第1期,第123~141页。
- [4]Antweiler, W. and Frank M. Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance*, 59(3): 1259~1294.
- [5]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.
- [6]Bollen, J, Mao HN, and Zeng XJ. 2010. "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2(1): 1~8.
- [7]Coval, J and Moskowitz T J. 1999. "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *The Journal of Finance*, 54(6): 2045~2073.
- [8]Fang, L., and J. Peress. 2009. "Media Coverage and the Cross Section of Stock Returns," *The Journal of Finance*, 64(5): 2023~2052.
- [9]Fisher, K. L., & Statman, M. 2003. "Consumer Confidence and Stock Returns," *The Journal of Portfolio Management*, 30(1): 115~127.
- [10]French, K. R., & Poterba, J. M. 1991. "Investor Diversification and International Equity Markets," *The American Economic Review* 81(2): 222~226.
- [11]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.
- [12]Lou, D. 2014. "Attracting Investor Attention through Advertising," *Review of Financial Studies*, 27(6): 1797~1829.
- [13]Peng, L., and W. Xiong. 2006. "Investor Attention, Overconfidence And Category Learning," *Journal of Financial*

- Economics* ,80(3) : 563 ~ 602.
- [14]Seasholes , M S and Zhu N. 2010. “Individual Investors and Local Bias ,” *The Journal of Finance* ,65(5) : 1987 ~ 2010.
- [15]Seasholes , M. S. , and Wu , G. 2007. “Predictable Behavior , Profits , and Attention ,” *Journal of Empirical Finance* , 14(5) : 590 ~ 610.
- [16]Solnik , B. , 2008. “Equity home bias and regret: an international equilibrium model ,” SSRN working paper.

The Effect of Local Bias in Investor Attention and Investor Sentiment on Stock Markets: Evidence from Online Forum

YANG Xiaolan SHEN Hanbin ZHU Yu

(Zhejiang University)

Abstract: In this paper , we test the cross effect of local bias and investor sentiment on stock markets. Text messages posted by investors are downloaded from an online stock forum in China. We build the Investor Sentiment Index based on the message content and the Local Bias Index based on their IP address. The empirical results show that the effect of Local Bias Index on stock market depends on the investor sentiment. As the investor sentiment is positive (negative) , there is a positive (negative) effect of local bias on the relevant stock return. Besides , local bias positively affect stock trading volume and this effect is greater as the investors are optimistic than as they are pessimistic. Besides , the cross effects of local bias and sentiment on stock market are similar across five different regions with more than 20 list firms each. However , these effects are not significant and consistent in regions with less list firms.

Key words: Local Bias , Attention , Investors Sentiment , Stock Return , Online Forum

(责任编辑: 林梦瑶) (校对: ZL)