

互联网基金对股票市场的影响

——基于大数据情绪指数的实证研究

许承明¹, 田婧倩²

(1. 南京晓庄学院, 江苏 南京 211171; 2. 南京财经大学 金融学院, 江苏 南京 210046)

[摘 要] 投资者在进行投资决策时易受到自身情绪的影响, 并且投资者行为是影响金融市场间波动溢出的直接原因。运用文本挖掘技术对新浪微博 2014 年 4 月至 2016 年 7 月的博文进行文本分析和随机森林-主成分分析并构建微博大数据投资者情绪指数, 根据投资者情绪指数研究互联网基金市场对股票市场的影响, 结果表明互联网基金市场对股票市场具有波动溢出效应。

[关键词] 互联网基金市场; 大数据情绪指数; 文本挖掘; 波动溢出效应; 投资者行为; 投资者情绪指数; 互联网金融; 股票市场; 余额宝

[中图分类号] F830.91 **[文献标识码]** A **[文章编号]** 1672-8750(2016)06-0033-08

一、引言

社交网络的发展为人们获取并预测有效信息提供了新的思路。Bollen 等研究发现 Twitter 大数据情绪指数中后移 3 天的冷静指数与道琼斯工业平均指数 (DJIA) 能够很好地匹配, 并且考虑到 Twitter 情绪指数的 DJIA 预测模型准确率会得到提升^[1]。我国学者黄润鹏等借鉴了 Bollen 团队的研究思路, 发现加入微博情绪信息的股票预测模型能获得更高的准确率^[2]。行为金融学认为金融市场间具有波动溢出效应, 即不同金融市场的波动存在相互联系, 其中投资者行为是影响金融市场波动溢出的直接原因。以基金市场与股票市场为例, 当两个金融市场上存在共同信息时, 投资者会马上做出反应, 改变投资决策, 因此投资者情绪指数可以作为金融市场间波动溢出效应的代理变量。

余额宝作为开放式货币基金的一种, 依托于互联网发展, 不仅具备普通货币基金购买后可赎回的特点, 还可以进行消费支付和实时转出, 其资金流动性更强。据此, 天弘基金推出了“余额宝入市意愿情绪指数”(下文简称“入市情绪指数”)这一衡量余额宝用户投资股市情绪的指标。同时, 新浪微博作为国内使用人数较多的社交网络平台, 其影响力不仅体现在新闻舆论方面, 而且在财经领域也得到进一步的延伸。因此, 新浪微博所反映的余额宝用户情绪可以作为衡量互联网基金市场投资者情绪的代理变量。在投资者情绪度量方法上, 郭晓菲等通过实证研究对比发现, 相较于岭回归-主成分分析法, 随机森林-主成分分析法构建的投资者情绪指数与金融资产价格趋势的相关性更高, 即选用后者作为情绪指数构建方法更有效^[3]。

本文运用文本挖掘技术对新浪微博 2014 年 4 月至 2016 年 7 月以“余额宝”为关键词的原创博文

[收稿日期] 2016-06-28

[基金项目] 国家自然科学基金项目 (71373114)

[作者简介] 许承明 (1960—), 男, 安徽广德人, 南京晓庄学院院长, 教授, 博士生导师, 主要研究方向为金融学; 田婧倩 (1991—), 女, 江苏南京人, 南京财经大学金融学院硕士生, 主要研究方向为互联网金融。

进行文本分析,并利用随机森林确定核心关键词,统计核心关键词的百度搜索指数,最终运用主成分分析法构建微博大数据投资者情绪指数(下文简称“微博情绪指数”)。与以往常用 SPSS 进行主成分分析不同的是,本文将采用 R 语言构建投资者情绪指数,这是因为 R 语言作为免费开源的大数据处理工具能够将不同数据集、工具和软件包结合在一起,相比于其他分析软件更具优势。最后,本文通过微博情绪指数和入市情绪指数的比较分析,从投资者行为角度出发判断互联网基金对股票市场的影响,并提出大数据背景下构建投资者情绪指数的政策建议。

二、文献综述

(一) 投资者情绪的传统度量

如何对投资者情绪进行度量一直是行为金融学研究的难题。当前对投资者情绪测度的文献成果较多,传统的投资者情绪度量方法主要有客观类指标、主观类指标两种。

1. 客观类指标

客观类指标是指将客观变量作为投资者情绪的度量指标,如封闭式基金折价率、股票换手率、IPO 数目、IPO 平均首日回报率、权益资本发行与债务资本发行的比率等指标以及建立在这些指标之上的综合类复合指标。DeLong 等构建了噪声交易者长期生存模型——DSSW 模型进行研究后认为封闭式基金折价率可以反映噪声交易者的情绪变化,他还从微观行为角度解释了股票的市场价格相对于其基础价值持续偏离的原因^[4]。许承明等以封闭式基金的平均折价率来衡量投资者情绪,从实证角度对我国封闭式基金的市场表现进行了分析。他们研究发现,投资者信心随着股票市场收益率的增加而增加,流入基金市场的资金则会减少,基金价格的上升幅度低于股票市场的上升幅度,进而使基金的折价增加^[5]。Brown 则认为成交量或流动性也可作为衡量公众投资情绪的有效指标^[6]。除此之外,IPO 的发行量也是衡量投资者情绪的一个常用指标。Baker 等运用主成分分析法,选取封闭式基金折价率、NYSE 股票换手率、IPO 数目、IPO 平均首日回报率、权益资本发行与债务资本发行的比率和股利溢价等 6 个变量建立了衡量市场层面投资者情绪的综合指标,研究结果表明估值不确定性高、套利困难的股票收益受投资者情绪的影响较大,估值不确定性较低、易被套利的股票收益则不易受投资者情绪的影响^[7]。该指标是消除了经济周期因素并综合了客观类指标和主观类指标的综合投资者情绪指数。易志高等改进了 Baker 等的指数构建方法,加入了封闭式基金折价、IPO 数量等 6 个变量,构建了一个能较好测度中国股票市场投资者情绪的综合指数^[8]。邢欣羿等借鉴 Baker 等构建投资者情绪指标的思想,用换手率、IPO 数目、IPO 首日回报率和波动溢价等 4 个变量来构建中国股市的投资者情绪指标,再运用主成分分析法构建综合指标研究噪声交易与股票特性、股票期望收益率的关系^[9]。蒋先玲等通过构建基于中国国情的投资者情绪综合指数考查了投资者情绪综合指数与中国 IPO 上市首日收益率之间的关系,研究表明,投资者情绪对 IPO 上市首日收益率有显著的正向影响,同时,IPO 上市首日收益率的波动也会对投资者情绪波动有显著的冲击^[10]。马若微、张娜以上海证券交易所 A 股公司的面板数据为基础编制了投资情绪指数 SENT,并通过实证研究发现 SENT 情绪指数与上证 A 股月度指数之间存在双向格兰杰因果关系^[11]。

虽然客观类指标在投资者情绪指数的构建上被国内外学者长期运用,但它仍具有一定的局限性:一是由于利用这类指标构建的投资者情绪指数都是通过统计数据事后得出的,因此很难对金融市场当前和未来的投资者情绪进行有效的分析和预测;二是不同的研究者选取了不同的客观类指标,具有较强的主观性,因此所构建出的投资者情绪指数不能客观全面地反映真实的市场投资者情绪;三是国内现有的研究主要采用单一的客观类指标衡量投资者情绪,构建综合类复合型投资者情绪指数的研究还比较少。

2. 主观类指标

主观类指标一般是由各大机构根据自己建立的评价体系,通过问卷、多空调查形式对金融市场参

与者进行统计得出的用于衡量投资者情绪的指标。国外比较有代表性的主观类指标有美国个体协会指数和投资者智能指数。Fisher 等对美国个体投资者协会构建的个体投资者协会指数能否反映投资者情绪变化进行了研究,结果显示这个指标是预测标准普尔 500 指数的反向指标。同时,他们采用 Chartcraft 公司编制的投资者智能指数作为中型投资者情绪指标进行研究,发现该指标与标准普尔 500 指数的收益率关系并不显著^[12]。目前,我国比较有代表性的投资者情绪指数有巨潮投资者信心指数、耶鲁-CCER 股市投资者信心指数、《股市动态分析》好淡指数等。薛斐研究发现,在衡量公众的投资情绪方面,消费者信心指数比封闭式基金折价率更好^[13]。陆江川、陈军以好淡指数作为投资者情绪的代理变量,发现周投资者情绪对股票横截面收益有正向影响;月投资者情绪仅对小市值股票横截面收益有负向影响;周极端乐观和极端悲观情绪对股票横截面收益影响具有非对称性;市场对极端悲观情绪反应过度,对极端乐观情绪反应不足^[14]。

主观类指标由于其问卷调查的形式具有较强的主观性,无法准确地反映投资者内心的投资决策,进而影响构建出的投资者情绪指数的真实性,因此,国内外利用主观类指标对投资者情绪进行测度的研究都相对较少。

(二) 大数据背景下投资者情绪的度量

近年来,大数据分析在互联网金融研究领域日渐兴起,涉及的研究方法主要是关键词搜索量分析和文本语言分析。

1. 关键词搜索量分析

关键词搜索量分析借鉴了大数据在其他领域的应用,以谷歌流感趋势为例,该团队首先提出利用大数据来预测流感发病率,他们通过分析数十亿搜索中 45 个与流感相关的关键词,可以比美国疾控中心提前两周预报 2007—2008 年冬季流感的发病率。

Da 等利用 2004—2011 年网络搜索量构建了一个金融经济搜索指数(FEARS)以衡量投资者情绪,他们发现该指数能预测短期回报逆转、暂时性增加的波动性以及从股票型基金流出并流入债券型基金的共同基金^[15]。邵新建等以百度搜索量作为刻画个体乐观情绪的强度指标,研究拟上市公司通过媒体公关活动影响其证券发行定价的内在机制,结果显示,在短期内参与证券发行的投资者人数、证券需求量和证券发行价格的上调幅度都随着乐观情绪同向变动^[16]。曾建光利用百度大数据中的关键词搜索量来度量互联网金融的网络安全风险感知,通过构建回归模型探究互联网金融特有风险的定价问题^[17]。郭晓菲等结合网民的搜索行为,利用网络爬虫技术从东方财富网等社交媒体中抓取关键词,运用主成分法构造了投资者情绪指数。同时,他们基于我国股票市场收盘价和投资者情绪的历史数据,利用 BP 神经网络模型对上证综指的变动趋势和短期内的收盘价进行研究和预测,发现基于网络搜索大数据所构建的投资者情绪指数与上证综指高度相关^[3]。孟雪井等运用文本挖掘技术对中国知网 CSSCI 期刊与新浪微博话题信息进行文本分析,并结合百度关键词推荐系统,在综合三大词库的基础上确定最能反映我国投资者行为的关键词词库,并利用因子分析法构建了沪市投资者情绪指数^[18]。

综上所述,互联网时代背景下,投资者搜索行为所构成的关键词搜索量能够广泛覆盖到各类投资者,能较客观地通过海量数据反映投资者情绪。对关键词搜索量的统计测量目前主要来源于各大搜索引擎、期刊数据库等,由于这类数据基本都可以从公开渠道中获得,因此利用关键词搜索量来衡量投资者情绪具有较强的数据可得性优势。所以,关键词搜索量可以作为构建投资者大数据情绪指数的代理变量之一。

2. 文本语言分析

在文本语言分析方面,目前研究比较多的是对社交网络大数据进行文本语言分析,通过对文本信息的量化探寻其对金融市场的影响。

Barberis 等提出了投资者情绪(BSV)模型,认为当投资者使用公开的信息预测未来现金流时会产生

生系统性误差,从而造成股票价格偏离其基本价值,导致泡沫的产生^[19]。游家兴、吴静运用文本分析法构建了衡量媒体情绪指数的综合评价指标体系,并对媒体情绪与资产误定价两者的关系进行了分析。研究发现,新闻报道所传达的媒体情绪高涨或低落时,股票价格会偏离其基本价值水平,并且乐观的媒体情绪更易使价格偏离基本价值,导致股价泡沫^[20]。Chen 等对美国最大的投资社交网站 Seeking Alpha 和道琼斯新闻服务网站 DJNS 上的报道和评论进行了文本分析并构建了回归模型分析报道评论负面率与股票超额收益之间的关系,结果显示,这些文章和评论能预言股票价格走势,甚至比华尔街分析师的预测还要准确^[21]。汪昌云、武佳薇利用 IPO 公司在上市前不同时间段内主流财经媒体报道中的正面词汇与负面词汇数据,构建“媒体语气”这一度量公司层面投资者情绪的代理变量,以媒体关于上市公司报道中的正面和负面词语占总词汇的比例来度量媒体语气,从个股层面分析投资者情绪对 IPO 抑价率的影响^[22]。尹海员等利用我国财经权威媒体对股票市场的共 847 篇报道文本,通过情绪分类词频统计构建媒体报道衡量指标。在投资者情绪的构建方面,他借鉴了 Baker 等的综合指标构建方法,实证结果显示媒体报道的乐观倾向、市场指数收益率对情绪产生显著正向推动作用,并且牛市环境下媒体报道的乐观倾向对情绪的影响比熊市环境更为显著^[23]。

可以看出,对社交网络、媒体报道的文本语言分析可以较为全面地反映互联网时代投资者情绪的变化。其中,社交网络中的海量博文及评论信息成为研究者利用互联网平台获取准确衡量投资者情绪信息的重要渠道。大数据技术能够将这些文本语言量化,使社交媒体的文本语言成为构建投资者大数据情绪指数的另一个有效代理变量。

三、理论分析与研究假设

传统金融理论认为市场有效假说(EMH)的前提之一是所有市场参与者能够同时接受信息,并且所有市场参与者是理性和追求效用最大化的。然而,在现实中这一条件很难成立,所有投资者不仅无法同时获取全部有效信息,而且投资者在进行投资决策时会受到心理因素的影响。行为金融学强调了心理因素在投资决策中的作用,认为在金融市场中投资者会在信息不确定的情况下受到他人决策行为或舆论信息的影响,如产生“羊群效应”等,而这些情绪变化影响了投资者投资决策的合理性。当今社交网络的蓬勃发展为投资者提供了信息高效交换的平台,解决了部分信息不对称问题。尤其对于近年来在我国发展迅猛的互联网基金行业而言,投资者年龄呈现年轻化的特征,这与社交网络的主要使用人群年龄特征相一致,所以社交网络在一定程度上可以反映互联网金融投资者情绪的变化。

波动溢出是指可以被投资者观察到的金融市场间的信息传导现象。市场有效假说认为,在强有效市场上,任何与金融市场有关的信息会在最短时间内同时被所有金融市场消化,金融市场间并不存在溢出效应。然而,在现实金融市场上是存在信息跨市场传导的,市场间是存在波动溢出效应和“金融传染”现象的。在处理信息方面,互联网技术提供的社交网络平台能及时地、直接地反映各个金融市场上的投资者情绪,投资者可以根据所获取的信息迅速进行投资决策。一般来说,风险规避者会选择投资基金,风险爱好者则会选择投资股票。但对于余额宝这类以中小个人投资者为主要用户对象的互联网基金而言,其投资者容易受到网络舆情影响,因此在股票市场上可能会出现与基金市场相一致的投资意愿。

综上,本文提出如下假设:微博情绪指数对余额宝用户入市意愿有预测作用,即互联网基金市场对股票市场具有波动溢出效应。

四、实证研究与结果分析

(一) 微博大数据投资者情绪指数的构建

1. 核心关键词的确定:随机森林法

我们运用八爪鱼采集器对新浪 2014 年 4 月 1 日到 2016 年 7 月 31 日以“余额宝”为关键词的微

博原创博文进行文本采集,并将采集到的博文用 ROST 新闻分析软件进行词性识别和词频统计,在分析过程中过滤无意义词语(如“我们”、“以下”、“com”等),得到包含 102 个词语的过滤后有效词表。

我们对以上 102 个初始关键词 2014 年 4 月 1 日至 2016 年 7 月 31 日的周平均百度搜索指数(剔除包含节假日的周数据)进行统计,发现有 13 个词语未被百度指数收录在统计范围内,因此投资者情绪指数的候选关键词由剩余 89 个关键词组成。利用 R3.3.1 软件中的 randomForest 包分析 89 个候选关键词对余额宝收益率的重要性,运用随机森林法计算出 89 个关键词共 9078 个数据对余额宝收益率的关键性衡量值 MeanDecreaseAccuracy 和 MeanDecreaseGini。其中,MeanDecreaseAccuracy 用于衡量把一个变量的取值变为随机数后随机森林预测准确性的降低程度;MeanDecreaseGini 则通过基尼指数计算每个变量对分类树每个节点上观测值异质性的影响来比较变量的重要性。以上两个值越大,表示该变量的重要性越大。本文选取以上两项指标均大于 1 的关键词作为核心关键词(见表 1),这 48 个核心关键词将作为构建投资者情绪指数的变量。

表 1 候选关键词表

候选关键词	候选关键词	候选关键词
现金	消费	安全
余额宝	互联网	影响
淘宝	密码	支付宝
信用卡	经济	小金库
手机	支付	基金
银行卡	一年	P2P
存钱	利率	财富
红包	货币基金	数据
创新	传统	理财产品
赚钱	块钱	选择
京东	平均	阿里巴巴
银行	未来	首次
股市	学生	一天
转账	准备金	股票
央行	货币	今日
利息	每日	服务

2. 微博余额宝情绪指数的构建:主成分分析法

接下来,我们利用 R3.3.1 软件中的 psych 包构建微博情绪指数(Wdx),这是因为 psych 包提供的函数比 R 基础安装包中的主成分分析 PCA 函数输出结果更直观。首先,计算相关关系矩阵,通过计算变量的相关系数数据集替代原始数据集。然后,进行主成分个数的判断,常见的判断主成分个数的方法是特征值法。其中,Kaiser-Harris 准则选择特征值大于 1 的作为主成分,Cattell 碎石图则选择将图形上弯曲变化最大之处以上作为主成分,而平行分析法则可以通过多次模拟依据随机数列矩阵提取特征值来判断主成分个数。psych 包中 fa.parallel() 函数可以同时用以上三种方法综合判断主成分个数,代码生成图 1。

从图 1 中可以看出,Kaiser-Harris 准则建议选择 10 个主成分,而碎石检验和根据 100 个随机数据矩阵进行的平行分析建议选择 6 个主成分。因此,选取 6 个主成分可以保留数据集的大部分信息,我们构建该投资者情绪指数的表达式为:

$$RC = \frac{\lambda_1 RC_1 + \cdots + \lambda_6 RC_6}{\sum_{i=1}^6 \lambda_i} \quad (1)$$

其中, λ 为各主成分的方差贡献率。

利用方差极大旋转的主成分分析能让主成分分析结果更具解释性。它对载荷阵的列进行去噪音,让每个成分只有一组有限的变量解释,结果显示 6 个主成分分别解释了微博情绪指数 18%、16%、13%、11%、7%、5% 的方差,共解释了其 70% 的方差,说明以上 6 个主成分能较完整地解释微博情绪指数。由此,基于随机森林-主成分分析法的微博情绪指数表达式为:

$$RC = (18RC_1 + 16RC_2 + 13RC_3 + 11RC_4 + 7RC_5 + 5RC_6) / 70 \quad (2)$$

最后,利用 R 中的主成分得分和微博情绪指数表达式得出 2014 年 4 月 1 日至 2016 年 7 月 31 日

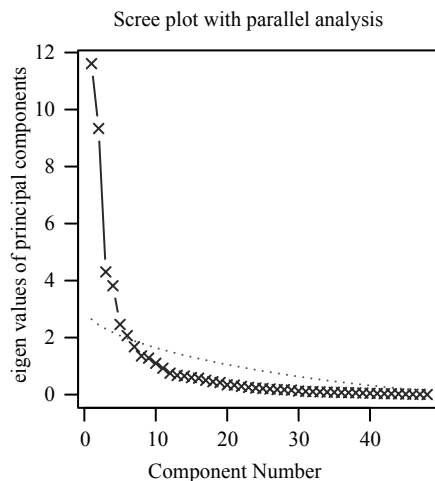


图 1 判断微博情绪指数数据集所需的主成分数

的微博余额宝投资者情绪周平均指数^①。

(二) 微博情绪指数与余额宝投资者入市意愿情绪指数关系分析

入市情绪指数是天弘基金在2015年11月推出的全球首支大数据指数,是天弘基金通过对余额宝资金流入股市的数据进行挖掘来刻画散户入市意愿的指数。该指数反映的是余额宝用户参与股市的意愿,代表市场上中小投资者的股市投资意愿。朱振等研究发现入市情绪指数与各证券市场指数收益率之间存在互相影响的关系,入市情绪指数可以作为衡量股市收益率波动的代理变量之一^[24]。

入市情绪指数从2014年4月1日起开始进行计算统计,我们对2014年4月1日至2016年7月31日入市意愿指数周平均值进行标准化处理,并比较它与本文构建的微博情绪指数间的关系,具体见图2所示。

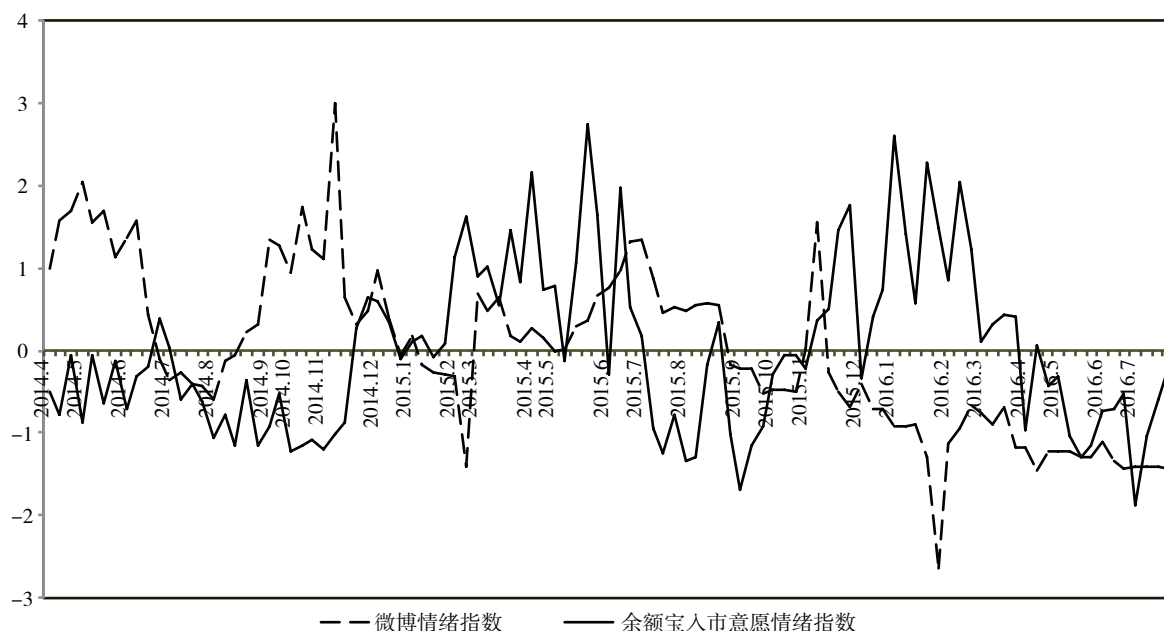


图2 微博情绪指数与余额宝入市意愿情绪指数走势图

从图2中可以看出,两个情绪指数整体具有较一致的波动趋势,虚线代表的微博情绪指数在具有上涨趋势时能对实线代表的入市情绪指数进行至少提前1个月的预测(如2014年6月至12月)。然而当市场出现震动股市出现下跌时,社交网络舆情发酵,舆论悲观,此时微博情绪指数会剧烈下跌并且在股市再次上涨时出现1周左右的时滞(如2015年2月至8月)。2015年8月底,央行通过降准降息释放大量流动性,市场资金面不再紧张,互联网货币基金面临收益率下降的必然趋势,因此,余额宝用户在2015年8月底将大量资金投资到股票市场,短期内余额宝用户相比于货币基金更看好股票市场,微博情绪指数也处于整体下跌趋势。综上,代表互联网基金投资者情绪的微博情绪指数可以对互联网基金投资者转投股市意愿的入市意愿情绪进行预判,即互联网基金投资者情绪可以提前反映股票市场上散户投资者情绪,本文假设成立。

五、研究结论及政策建议

本文结合文本挖掘和搜索指数构建微博余额宝情绪指数,并探讨其对余额宝用户入市情绪指数的影响,得到如下结论:

^①因篇幅所限,此处不再列出,读者如有需要,可向笔者索取。

第一,以社交网络文本语言为数据集构建的情绪指数能够有效反映互联网基金投资者情绪的波动情况。这有以下两点原因:一是现代社交网络的高速发展加快了信息传播速度,降低了信息传播成本,在一定程度上解决了信息不对称问题;二是以新浪微博为代表的社交网络使用人群年龄集中在24~33岁,而以余额宝为代表的互联网基金投资人群年龄集中在18~35岁,两者受众对象在年龄范围上有交集,因此社交网络情绪指数能够准确反映互联网基金投资者情绪。

第二,微博情绪指数对余额宝用户入市意愿有预测作用,即互联网基金市场对股票市场具有波动溢出效应。这是由于投资互联网货币基金的大多是市场上散户投资者,他们易受网络舆情影响并且资金有限,当基金市场上收益率下跌时,他们会选择转出资金立刻投资股市以获得更大的收益;当基金市场上收益率较高时,中小投资者则会考虑将所获收益转投至基金市场以规避风险。正因为互联网基金用户面对不同市场情况时风险偏好的变化,微博情绪指数与入市情绪指数在整体上波动趋势相似。

可以看出,利用文本挖掘等大数据分析构建情绪指数能有效反映互联网基金市场上投资者情绪的变化,也能有效地对股票市场的投资情绪进行预判。然而,当前我国投资者大数据情绪指数暂时只有利用客观数据进行构建的“余额宝入市意愿情绪指数”,尚未有机构发布利用文本分析和关键词搜索量等间接衡量投资者情绪的情绪指数。鉴于我国当前的投资者大数据情绪指数发展状况,针对如何推进运用大数据技术对投资者情绪进行指数构建与实际运用的问题,我们提出以下建议:

第一,提升大数据分析思维能力,拓展大数据指数构建方法。大数据时代带来了全新的思维方式,强调分析对象从随机样本向全体数据的转变、数据精确性向数据混杂性的转变以及追寻变量间因果关系向追寻相关关系的转变。在金融学研究中运用大数据思维,有助于研究者在一定程度上突破传统数据分析存在的局限。因此,提升大数据分析思维能力可以从根本上拓展投资者情绪指数构建的思路,有助于构建客观全面的投资者大数据情绪指数。

第二,建设第三方大数据分析机构,保证大数据指数的客观性。当前,最具代表性的投资者大数据情绪指数是“余额宝入市意愿情绪指数”。但值得注意的是,该指数是由天弘基金编制,而余额宝作为互联网货币基金,对应的正是天弘增利宝基金。天弘基金作为余额宝的基金管理公司,拥有余额宝投资者的海量交易数据,一定意义上垄断了该互联网金融产品的大数据。另外,由于数据来源和算法的不透明,大数据情绪指数的客观性也有待考证。因此,应加快建立第三方大数据分析机构,防止数据垄断,保证大数据来源和情绪指数算法的透明性,确保大数据情绪指数的客观真实性。

第三,建立大数据安全保护规则,处理好投资者隐私与大数据预测间的关系。构建投资者大数据情绪指数的最终目的是利用该情绪指数对金融市场的发展趋势进行预判,但在利用投资者交易信息、文本信息等数据进行情绪指数构建的同时,大数据运用所带来的投资者个人隐私安全问题也随之而来。欧盟委员会在2012年通过颁布《一般数据保护条例》,重新建立法律监管框架来应对大数据带来的挑战。我国应借鉴欧盟的经验,加强对互联网金融的监管,尽快出台大数据安全保护规则,积极应对大数据时代给投资者隐私带来威胁的问题。

参考文献:

- [1] BOLLEN J, MAO H, ZENG X. Twitter mood predicts the stock market[J]. Journal of Computational Science, 2011, 2: 1-8.
- [2] 黄润鹏, 左文明, 毕凌燕. 基于微博情绪信息的股票市场预测[J]. 管理工程学报, 2015(1): 47-52.
- [3] 郭晓菲, 李清军, 潘子颖, 等. 基于投资者情绪指数的上证综指预测研究[C]//中国统计教育学会. 2015年(第四届)全国大学生统计建模大赛论文. 北京:中国统计教育学会, 2015: 24.
- [4] DELONG J B, SHLEIFER A, SUMMERS L H, et al. Noise trader risk in financial markets[J]. Journal of Political Economy, 1990, 98: 703-738.
- [5] 许承明, 陈百助. 投资者情绪对封闭式基金折价的影响——对我国封闭式基金折价变动的实证[J]. 产业经济研

- 究,2003(6):1-14.
- [6] BROWN G W, CLIFF M T. Investor sentiment and asset valuation[J]. The Journal of Business, 2005, 78:405-440.
- [7] BAKER M, WURLER J. Investor sentiment and the cross-section of stock returns[J]. The Journal of Finance, 2006, 61:1645-1680.
- [8] 易志高,茅宁. 中国股市投资者情绪测量研究:CICSI 的构建[J]. 金融研究,2009(11):174-184.
- [9] 邢欣羿,罗妍,石松. 噪音交易对股票收益率的影响——基于个股层面的研究[J]. 技术经济,2015(12):116-124.
- [10] 蒋先玲,张斯琪. 投资者情绪对中国 IPO 首日收益率影响的实证分析[J]. 经济问题,2015(6):51-56.
- [11] 马若微,张娜. 我国股票市场投资者情绪 SENT 指数的构建——基于上证 A 股公司的面板数据[J]. 中央财经大学学报,2015(7):42-49.
- [12] FISHER K L, STATMAN M. Investor sentiment and stock returns[J]. Financial Analysts Journal, 2000, 56:16-23.
- [13] 薛斐. 基于情绪的投资者行为研究[D]. 上海:复旦大学,2005.
- [14] 陈军,陆江川. 基于 DSSW 模型投资者情绪与股价指数关系研究[J]. 预测,2010(4):53-57.
- [15] DA Z, ENGELBERG J, GAO P. The sum of all fears investor sentiment and asset prices[J]. Review of Financial Studies, 2015, 28:1-32.
- [16] 邵新建,薛熠,江萍,等. 投资者情绪、承销商定价与 IPO 新股回报率[J]. 金融研究,2013(4):127-141.
- [17] 曾建光. 网络安全风险感知与互联网金融的资产定价[J]. 经济研究,2015(7):131-145.
- [18] 孟雪井,孟祥兰,胡杨洋. 基于文本挖掘和百度指数的投资者情绪指数研究[J]. 宏观经济研究,2016(1):144-153.
- [19] BARBERIS N, SHLEIFER A, VISHNY R. A model of investor sentiment[J]. Journal of financial economics, 1998, 49:307-343.
- [20] 游家兴,吴静. 沉默的螺旋:媒体情绪与资产误定价[J]. 经济研究,2012(7):141-152.
- [21] 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, 27:1367-1403.
- [22] 汪昌云,武佳薇. 媒体语气、投资者情绪与 IPO 定价[J]. 金融研究,2015(9):174-189.
- [23] 尹海员,王盼盼. 媒体报道、市场收益与投资者情绪——基于新闻分析视角[J]. 软科学,2015(7):136-139.
- [24] 朱振,蒋文璐. 余额宝情绪指数与中国股票市场间相互影响的实证分析[J]. 中国物价,2016(5):36-39.

[责任编辑:黄 燕]

The Influence of Internet Fund on the Stock Market: An Empirical Study Based on Big Data Sentiment Index

XU Chengming¹, TIAN Jingqian²

(1. Nanjing Xiaozhuang University, Nanjing 211171, China;

2. Nanjing University of Finance and Economics, Nanjing 210046, China)

Abstract: Behavioral finance believes that investors are vulnerable to the impact of their own emotions when making investment decisions, and investors' behavior is the direct cause of the volatility spillover between financial markets. From the perspective of the sentiment index, this paper makes a research on the influence of Internet fund market on stock market, on the basis of text mining technology and random forest-principal component analysis method on Sina Micro-blog to build big data investor sentiment index from April 2014 to July 2016. The results show that the Internet fund market has volatility spillover effect on the stock market.

Key Words: Internet fund; big data sentiment index; text mining; volatility spillover effect; investors behavior; investors sentiment index; Internet financing; stock market; Yu' E Bao