

平台中的评论信息便有了现实意义。

45 由于网络信息质量参差不齐,获取有价值的信息变得十分困难,直到文本挖掘这一技术手段的出现才弥补了这一空缺。文本挖掘能够对大量的非结构化文本通过算法以及文字处理技术,对文本进行分类,方便学者们能够更快速的获取有用信息。如何从股票评论中获取有价值的信息,并以此了解股市变动情况,为股市参与者和监管者提供合理建议,是本文研究的重点。

1 文献综述

1.1 投资者情绪定义

50 Jeremy C. Stein (1996)^[1]认为,投资者在市场中基于自己对市场中金融资产的预判进行交易,但最终结果并不符合自己的预判,而是与资产现实价值之间有一定差距,这种差距就是由于投资者情绪的作用。Long (1990)^[2]认为投资者情绪是投资者在市场活动中通过自身经验的积累和对市场逐渐形成一定认知后从情感角度出发所做的一种判断。Brown 和 Cliff (2004)^[3]将投资者情绪定义为对股票市场未来趋势的预测偏差,或者是对未来走势的
55 过度积极或消极的态度,投机者会根据历史信息和已知事实对股票市场做出预判和预期。王美今 (2004)^[4]对于投资者情绪的定义和 Long 类似,投资者情绪是投资者在情绪影响下做出的投资决策。投资者情绪产生的原因是投资者的预期与现实存在差异,产生心理和认知上的偏差,从而对预期收益产生错误判断,即市场中完全理性投资者和有限理性或非理性投资者对资产未来收益和走势的判断间的偏差。张颖、粟坤全、袁野菁菁 (2013)^[5]将非理性情绪
60 归结为投资者因自身对市场信息的认知和错误预判。

可以看出,对于投资者情绪的定义是多种多样的,研究角度不同描述则不同,但总结起来主要是由于金融市场的不确定性,投资者情绪在自身基础上,如所受教育、性格、风险偏好等,根据已知事实对股票市场价格走势的一种预测和判断,是一种主观性的评价。投资者在进行投资活动时,受心理因素影响,基于既定事实对金融市场的预期与传统金融理论的“理性人假设”的偏离。
65

1.2 投资者情绪度量

在研究投资者情绪对股市影响时,如何度量投资者情绪也成为研究者们讨论的焦点,从不同角度出发,将投资者情绪的度量分为三类。

(1) 直接投资者情绪指标

70 直接投资者情绪指标主要是通过发放问卷或对话了解投资者对未来股市价格走势的预期。这种测量方法简单直接,但成本较高。薛斐 (2005)^[6]在研究中,将投资者情绪用封闭式基金折价率和消费者信心指数代替,并比较两种方法的差异,结果证明前者对于情绪的解释力和关联性优于后者。陈军,陆江川 (2013)^[7]在研究中选择了“央视看盘”指数以及“好淡指数”测量了投资者情绪,并与股票市场收益率做了实证研究得出两者具有很强的相关性的
75 结论。

(2) 间接投资者情绪指标

间接投资者情绪指标是指可以从金融市场的历史交易数据中获得的一些代理变量,这些代理变量通过整理分析在一定程度上能够反映出投资者情绪的变化,这种度量方法有很强的

事后性, 较为客观。Seo, Kim (2015)^[8]以期权隐含信息波动率作为情绪的代理指标, 研究发现它对投资者情绪有一定的预测作用。韩立岩 (2007)^[9]在研究中将投资者情绪的度量用新增开户数代替, 在研究中发现, 新增开户数越多, 一般来讲投资者情绪也会越高涨。

(3) 复合投资者情绪指标

通过研究发现, 为更准地衡量投资者心情, 在研究中只选择一种指标是远远不够的, 结果也会有一定的偏差。为了更准确的衡量情绪波动, 学者们开始将投资者情绪代理变量用多个指标有机结合。M. Baker, J. Wurgler(2006)^[10]为研究投资者情绪与股票盈利间的联系, 研究通过换手率等六个指标构建并剔除了宏观经济原因, 并通过主成份分析法构建了投资者情绪变量模型, 并在此基础上作进一步研究。Han 和 Li (2017)^[11]运用主成分分析法将换手率等指标构建综合投资者情绪指数。结果表明, 投资者情绪对股票市场的未来变化具有较强的预测作用。张征超和刘宇宸妃 (2018)^[12]构建的复合投资者情绪指数, 在变量的选取上, 以投资者信心指数、交易量等五项指标为变量, 并得出了与 Han、Li 两人一致的结论。

1.3 投资者情绪与股市收益

由于技术限制, 前期的研究对于投资者情绪的表达通常采用网络数值型变量, 其中最经典的是 DSSW 模型。该模型认为, 在可获利益有限的背景下, 投资者情绪能够对股票价格产生一定影响, 这为噪声交易者模型的建立奠定了基础。被后续许多学者广泛使用。杨秋平 (2015)^[13]在研究中采用格兰杰因果关系检验出投资者情绪与股市收益之间存在因果关系, 后又通过相关关系检验, 证明两者在长期看是负相关关系, 在短期内是正相关关系, 且当投资者对股市表现持积极态度时对股市收益的影响要比持消极态度时的影响更强烈。

随着文本挖掘技术的发展, 学者们开始将可以直接反映投资者情绪的文本信息作为投资者情绪的代理指标来研究股市问题。经典文献例如 Antweiler(2004)^[14]通过挖掘雅虎财经网站中上市公司的 150 万条评论数据, 利用朴素贝叶斯算法将文本信息分为看涨、看跌、中立三种类别, 发现所构建的看涨指数对于股价波动有一定的预测作用。

Bollen 和 Mao(2010)^[15]通过谷歌文本挖掘工具对 Twitter 上的发言进行采集并做出情感倾向判断, 利用 Granger 检验投资者情绪和股票收益的因果关系, 结果发现投资者情绪对次日的股市预测结果的准确性近 90%。房文敏 (2016)^[16]运用文本挖掘方式提取网络论坛中股票相关评论, 采用支持向量机的情感分类和基于语义情绪的分析方法对投资者情绪稳健性进行了验证, 发现前者方法优于后者; 投资者情绪能够影响股价波动, 两者之间存在格兰杰因果关系, 且在走势上呈现正相关关系。张伟 (2015)^[17]通过关键词搜索对微博文本进行挖掘, 构建投资者情绪代理变量, 研究与股市收益率之间的关系, 结果发现, 两者之间存在很强的相关关系。丁小虎 (2016)^[18]通过抓取东方财富股吧数据, 通过建立情绪跟踪指数和情绪收敛指数, 来衡量情绪与股市价格波动之间的关系。研究结果表明, 投资者情绪跟踪指数和投资者情绪收敛指数与股票收益率正相关, 前者对股票收益率波动具有较好的预测和解释能力; 情绪收敛指数的偏离度越大, 股票的周转率越高, 股价的波动也就越大。

1.4 文献评述

当前学术界对于投资者情绪与股市收益的研究已经十分透彻, 就投资者情绪构建而言, 已经不再是利用单一的直接或间接变量构建, 而是更多的利用复合指标构建情绪代理变量, 文本挖掘技术应用的也越来越广泛; 在研究投资者情绪与股市收益方面, 研究者更倾向于采用格兰杰因果检验、脉冲响应等时间序列方法进行分析。

120 在研究中对于投资者情绪的分析不管采用什么方法,最后都会归结为投资者情绪与股市收益具有相关性,甚至投资者情绪还可以预测股市未来波动上在这方面的研究是比较全面的。但是在研究内容上还是存在一定的缺陷的,比如在投资者情绪的度量上不是很全面,因此本文在文本挖掘分析投资者情感的基础上,建立包含情感值的复合投资者情绪指标以及不包含情感值的综合投资者情绪指标,对比两种方法对股市的解释力,通过实证研究探讨投资者情绪与股市收益之间的关系,并得出相关结论。

2 投资者情绪指标综合构建

125 2.1 投资者情绪指标的选择

2.1.1 指标种类选择

130 当前投资者情绪的代理变量有直接、间接和复合指标三种表示方法。本文在研究中采用投资者在网络论坛中的帖子信息这一直接代理变量来反映投资者情绪指数,原因在于,另外两种投资者情绪的度量方法所采用的数据都是经过历史交易数据整合获得的,一方面,它不能很直接的体现投资者情绪的变化,另一方面,因为是对历史数据的整理,在研究中难免会与实际有所偏差。由前文可以得知,直接指标一般是对投资者进行问卷调查来得知投资者对股市行情的变化的预测,由此计算出投资者情绪指标,但其存在时间成本大、数据不易统计等缺点。

135 本文所采用的方法是依托互联网大数据背景,通过文本挖掘网页信息整理分析得出,弥补了传统的市场调查成本高昂等缺点。通过文本挖掘方式得到的情绪指标有以下优点:其一,随着互联网技术的不断发展和普及,以及网民和股民高度重合的特点,投资者也更愿意通过网络平台分享自己的投资问题,因此通过网络数据挖掘分析可以研究投资者情绪对股市收益在一段时间内的影响,是传统的市场调查所不具备的;其二,股民在网络平台上自由发言,短时间内就可以获得大量的样本信息,由此得到的投资者情绪指标的指示性更强;其三,投资者情绪指标的包容性更强,投资者所表达的观点受环境、心理、他人发言等多种因素影响,网络信息并不一定全部是真实的,通过对全部文本的处理所得到的数据更符合实际中的股票市场。

140 2.1.2 数据来源

145 目前股市参与者主要在微博、微信公众号、财经论坛和股吧等网络平台获取股票信息以及发表言论。比较来看,微博和微信公众号只能被少量关注的人获取到相关信息,传播范围窄速度慢,而财经论坛和股吧参与者众多,传播信息速度快,东方财富股吧论坛目前发帖量达 206549423 条,最早可追溯到 2002 年 4 月 16 日,单帖最高浏览量达 480 万,可见其影响范围之大。东方财富股吧最具股吧代表性,其可追溯时间长、能够获得的样本大,所得结论更准确;板块以及个股均分别创建了股吧,信息搜索方便;数据更加全面,可以获得标题、发帖时间、阅读量等重要信息。因此本文选用财经论坛和股吧中具有代表性的东方财富股吧(<http://guba.eastmoney.com/>)为对象,利用 python 编程对股吧评论进行文本挖掘,提取其中的帖子评论做投资者情绪变量。当所需要采集的数据信息是以数万计时利用人工方式时十分费时费力且不现实的,而网络爬虫技术可以在较短时间内获取到所需信息,python、R 语言等软件均可实现网络爬虫,考虑到便捷性和易操作性,本文采用 python 软件进行爬取。

155 2.1.3 数据选取

本文所选择的研究数据是东方财富股吧论坛中中证 100 指数十大权重股的个股评论数据，原因有以下几点：

160 首先，中证 100 指数的成分股均取自于沪深 300 指数中按规模排列最大的前 100 支股票，中证 100 指数覆盖了沪深两板块的核心蓝筹股，相比于单一的上交所或深交所的股票指数来说，中证 100 指数对我国 A 股市场有影响力的股票的表现能够更好的反映出来。

165 其次，股吧中的发帖内容包括标题、最后更新时间、浏览量、评论数等，包含信息十分全面，加上时间因素，想要完整采集中证 100 指数所有成分股评论并储存所需信息需要十分巨大的人力、时间资源，综合考虑后本文选取中证 100 指数成分股中十大权重股的发帖数据作为本文的研究对象，2020 年 10 月 1 日到 2021 年 3 月 31 日之间，半年时间内帖子数量共计 373284 条。

表 1 中证 100 指数权重股
Table 1 Weighted stocks of the CSI 100 Index

指数英文名称 Index Name(Eng.)	成分券代码 Constituent Code	成分券名称 Constituent Name	权重(%) Weight(%)
CSI 100	600519	贵州茅台	7.84
CSI 100	601318	中国平安	6.62
CSI 100	600036	招商银行	4.91
CSI 100	000858	五粮液	4.04
CSI 100	000333	美的集团	3.14
CSI 100	601166	兴业银行	2.72
CSI 100	600276	恒瑞医药	2.67
CSI 100	000651	格力电器	2.35
CSI 100	601888	中国中免	2.32
CSI 100	600887	伊利股份	1.89

2.2 数据采集及处理

170 2.2.1 数据采集

175 在抓取文本信息方面，本文采用 Python 语言作为爬虫工具，利用 Xpath 获取特定标签所储存的信息，找到网页规律后构造循环爬取网页源代码，通过 requests 库获得 reponse 返回的总结果，并利用 Xpath 解析有用的信息，将抓取结果保存到 Excel 中。目前大多数财经网站都使用了防爬虫策略，为了顺利采集需要的数据信息，本文在编程过程中加入模拟用户操作行为，增加睡眠时间进行间歇性采集，防止网站将爬取定义为机器爬虫。

2.2.2 数据处理

180 在采集完所需要的数据以后，需要对数据进行去停用词、分词的操作，本文通过 jieba 库对采集到的信息进行处理，jieba 分词原理是将一个复杂的文本拆分成若干个可能的词语，这些词语经过重新组合又会成为原文本，共有三种分词模式可供选择，本文采用精确模式即可完成分词操作，图 1 为部分词云图。

195 2.3 基于主成分分析法的投资者情绪指标

2.3.1 综合投资者情绪变量

本文选取的变量能够代表我国投资者的主观情绪，数据来源均为国泰安数据库，所用数据包括封闭式基金折价率、IPO 数量、新增开户数和消费者信心指数。

200 封闭式基金折价率（DCEF）：这种封闭式投资基金的折价率主要是一种泛指在基金发行前设定一个基金的投资发起人就已经设定好了该基金的发行数量和总额,当基金在发行中达到此投资数额时就可能会对其进行一个封闭,一定的时间内不会再接受新投资。张丹,廖士光（2009）[19]在研究中指出，封闭式基金能够间接的反映出投资者情绪，并且在一定程度上能够反映出投资者情绪与股市的关系。封闭式基金折价率越高，说明投资者越不看好股市走向，对股市是持消极态度的，相反封闭式基金折价率越低时，说明投资者看好股市走向，对股市是持积极态度的。当投资者的情绪比较高时，就会使得股市收益率提高，对收益波动产生影响，进而产生风险溢价，说明由封闭式基金折价率变动带来的投资者情绪的变动是能够影响股市收益波动和收益率变动的。

205 IPO 个数（IPON）：IPO 的数量可以直接反映投资者对股市的需求和投资热情，热季投资者的情绪高昂,对于股市持有信心,冷季时投资者的情绪低落,对股市持消极态度。

210 新增开户数（NIA）：新开户数能有效反映投资者对股市的态度和参与程度,投资者对于股市抱持积极态度的同时会逐渐增加市场中新增的开户人数,反之投资者对于股市抱持消极态度时会使得市场中的新增开户数减少，投资者会对市场持观望态度。

215 消费者信心指数（CCI）：消费者自信心指数有效地反映了我国消费者对当前和未来市场经济状况变化的态度，能够有效提前预测到经济走势和消费趋向，投资者信心指数上升时说明投资者更愿意投资，对股市参与度也会变高，反之消费者信心指数下降时说明投资者更愿意将资金拿在手中，对股市的参与度会减弱。

2.3.2 综合投资者情绪指标构建

表 3 KMO和巴特利特检验

Table 3 KMO and Bartlett tests

KMO 取样适切性量数。		.593
	近似卡方	34.407
巴特利特球形度检验	自由度	6
	显著性	.000

220

表 4 总方差解释

Table 4 Total variance interpretation

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	1.846	46.151	46.151	1.846	46.151	46.151
2	1.147	28.679	74.830	1.147	28.679	74.830
3	.656	16.410	91.240	.656	16.410	91.240
4	.350	8.760	100.000			

225

表 5 成分矩阵^a

Table 5 Composition matrix A

	成分		
	1	2	3
封闭式基金折价率	-.120	.939	-.234
IPO 个数	.837	.317	-.161
新增开户数	.755	.165	.601
消费者信心指数	-.748	.371	.463

表 6 成分得分系数矩阵

Table 6 Component score coefficient matrix

	成分		
	1	2	3
封闭式基金折价率	-.065	.819	-.357
IPO 个数	.453	.276	-.246
新增开户数	.409	.144	.915
消费者信心指数	-.405	.324	.706

230

通过 KMO 和巴特利检验，可以看出 KMO 接近 0.6，显著性 0.000，表明适合做主成分分析；在主成分分析中，选择前三个作为主成分，对于原始数据的解释力达 91.24%，根据各特征值大小可以构建出综合投资者情绪指数 $sen_{综合}$ ：

$$sen_{综合} = \frac{1.846F_1 + 1.147F_2 + 0.656F_3}{3.649} \quad (1)$$

2.4 复合投资者情绪指数

235

复合投资者情绪指数的构建是将主成分分析构建的综合投资者情绪变量与文本挖掘计算的情绪值相结合，采用加权平均法构建复合投资者情绪指数。在进行权重系数的确定时，选择了 CRITIC 权重法确定 α 和 β 系数的值。CRITIC 权重法主要是通过利用数据的波动性和相关关系的情况有效的确定出数据的权重。首先，对数据进行归一化处理消除量纲上的影响；其次，对处理后的数据进行权重计算，最终结果如表 7 所示。

240

表 7 CRITIC 权重计算结果

Table 7 weight calculation results of CRITIC

项	指标变异性	指标冲突性	信息量	权重
文本挖掘计算的情感值	4.618	1.148	5.303	44.88%
综合投资者情绪变量	5.672	1.148	6.514	55.12%

首先，对数据进行归一化处理消除量纲上的影响；其次，对处理后的数据进行权重计算，最终结果如式 (2) 所示：

$$SI_i = \alpha sen_{综合} + 0.4488sen_i \quad (2)$$

245 **3 实证研究**

3.1 机理分析及研究假设

3.1.1 机理分析

250 在传统金融学理论中,假设投资者在做投资决策时都是完全理性的,会选择对自己自身利益最大化的决策,但这一假设并不能合理解释市场中的金融异象问题。现实生活中,投资者的情绪不仅直接受到了市场环境的影响,同时会直接受到各种心理因素的影响,投资者自身对于认知的偏差和羊群行为的影响会直接产生投资者情绪,进而影响投资者的投资行为,投资者是有限理性的个体,不完全理性投资者的这种投资行为将被视为噪声交易。De Long 等人在研究投资者行为时提出了一个噪声交易模型,即 DDSW 模型。该模型将理性投资者和非理性投资者的资产分为两类:有风险的资产和不具备风险的资产,噪声交易者在市场中的投资行为会使得资产价值偏离其内在价值,造成市场波动。

255 我国股票市场的异常波动也有其独特的原因:首先,我国股市是在经济体制转轨这一特定背景下产生的,政府干预明显,由于政府在市场中的多重身份定位会导致股价大幅波动;其次,噪声交易者的投资行为以及金融市场中双方所获得信息的不对称会使得股市发生异常波动,理性投资者在进行交易决策时已经掌握了市场中的大部分信息,而为了自身利益最大化,股市开盘时就会进行大量的买卖,理性投资者往往会尽可能地选择隐藏自己的交易行为,交易虽然发生了但是股市并没有明显的变动,在此之后的股市买卖大多是因为噪声交易者的从众行为引起的股市交易量和价格的大幅波动;最后,我国对于股市的规范还不够完整,相关法律法规并不健全,在股市中存在做庄、寻租等行为,也会导致股票市场的异常波动。

3.1.2 研究假设

265 H1: 与综合投资者情绪指数相比,加入文本挖掘的复合投资者情绪指数对股市收益具有更强的预测和解释能力。

H2: 复合投资者情绪指数与股市收益正相关。当投资者情绪高涨时,股市收益率上升;当投资者情绪低落时,股市收益率就会下降。

3.2 综合投资者情绪与符合投资者情绪指标比较研究

270 表 8 相关性分析结果

Table 8 Correlation analysis results

	综合投资者情绪指数	复合投资者情绪指数	中证 100 指数
综合投资者情绪指数	1		
复合投资者情绪指数		1	
中证 100 指数	.598**	.818**	1

275 由表 8 可以看出,综合投资者情绪指数与中证 100 指数相关性为 0.598,显著性 $0.000 < 0.01$,综合投资者情绪指数与股市收益呈正相关。复合投资者情绪指数与中证 100 指数相关性为 0.818,显著性 $0.000 < 0.01$,说明包含文本挖掘得到的复合投资者情绪指数与股市收益的相关性要强于只含有间接指标的综合投资者情绪指数,对股市收益的解释要更强。

3.3 中证 100 指数与复合投资者情绪指数实证研究

3.3.1 平稳性检验

利用 ADF 对复合投资者情绪指数和中证 100 指数收益率序列进行检验，原始假设是存在单位根。检验结果如下所示：

280

表 9 复合投资者情绪指数 ADF 检验结果

Table 9 Composite investor sentiment index ADF test results

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.174722	0.0018
Test critical values: 1% level	-3.568308	
5% level	-2.921175	
10% level	-2.598551	

表 10 中证 100 指数收益率 ADF 检验结果

Table 10 ADF test results of CSI 100 index yield

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.524375	0.0000
Test critical values: 1% level	-3.568308	
5% level	-2.921175	
10% level	-2.598551	

285

由表 9、10 所示，复合投资者情绪指数在三种检验水平下的统计量为-3.568308、-2.921175、-2.598551，ADF 检验的统计量 t 为-4.174722，P 值为 0.0018<0.05；中证 100 指数收益率序列在三种检验水平下的统计量分别为-3.568308、-2.921175、-2.598551，ADF 检验统计量为-6.524375，P 值 0.0000<0.05，说明拒绝原假设，在 95% 的显著性水平上两序列都是平稳的。

3.3.2 格兰杰因果检验

290

表 11 格兰杰因果检验结果

Table 11 Granger causality test results

滞后阶数	格兰杰因果	F 值	P 值	结论
1	SR does not Granger Cause SI	2.08411	0.1555	SR 不是 SI 的 Granger 原因
	SI does not Granger Cause SR	0.58376	0.4487	SI 不是 SR 的 Granger 原因
2	SR does not Granger Cause SI	1.59935	0.0135	SR 是 SI 的 Granger 原因
	SI does not Granger Cause SR	0.49044	0.6157	SI 不是 SR 的 Granger 原因

3	SR does not Granger Cause SI	4.55798	0.0076	SR 是 SI 的 Granger 原因
	SI does not Granger Cause SR	0.48367	0.6955	SI 不是 SR 的 Granger 原因
4	SR does not Granger Cause SI	3.49626	0.0159	SR 是 SI 的 Granger 原因
	SI does not Granger Cause SR	1.81174	0.1467	SI 不是 SR 的 Granger 原因

由表 11 可知，当滞后阶数选择 1 阶时，中证 100 指数收益率与复合投资者情绪指数不存在格兰杰因果关系，当滞后阶数为 2 阶及以后时，中证 100 指数收益率是复合投资者情绪指数的原因，接下来构建复合投资者情绪指数与中证 100 指数的向量自回归模型。

295 **3.3.3 向量自回归模型**

表 12 滞后阶数的确定

Table 12 Determination of hysteresis order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	26.19200	NA	0.001170	-1.075200	-0.994904	-1.045266
1	33.91178	14.41025	0.000992	-1.240523	-0.999635*	-1.150723
2	35.58434	12.46690*	0.001102*	-1.137082*	-0.735601	-0.987414*
3	42.96606	4.973454	0.000951	-1.287380	-0.725308	-1.077845
4	46.82437	6.173296	0.000963	-1.281083	-0.558418	-1.011681

表 13 回归分析

Table 13 Regression analysis

300

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SR(-2)	6.543472	1.873743	3.492192	0.0011
C	6.522966	1.899055	-3.434849	0.0013

由表 12 可知最优滞后阶数为 2 阶，由表 13 可知，P 值为 0.0011 < 0.05，说明回归模型是较为有效的，模型为：

$$SI_{i,t} = 6.523 + 6.543SR_{i,(t-2)} + \varepsilon_{i,t} \quad (3)$$

305

该模型说明中证 100 指数收益率和复合投资者情绪指数之间存在明显的正向关系，滞后二期的中证 100 指数的收益率越高。

4 对策建议

310

随着互联网的不断发展，越来越多的人在网络上抒发自己的情感，互联网已经成为人们生活中不可或缺的一部分，由于网络的便捷性和信息传播的快速性，投资者在网络上发布的信息不仅能够影响自己的投资行为，也影响着浏览信息的其他投资者的投资行为，网络信息被进一步扩散，对更多的投资者造成影响，就有可能对股票市场产生影响，造成股市的波动。

根据以上研究, 本文分别对股市投资者、金融市场监管者提出一些可行性建议。

4.1 针对投资者

315 一方面, 在进行投资活动时要注意到整体投资者情绪对于股票市场的影响。由于投资者非理性情绪会影响自身决策行为, 因此在进行投资活动时, 要提前考虑到投资者情绪对股市的影响, 合理利用股吧论坛中的评论信息, 减少不必要的资产损失, 提高自身收益, 保持理性态度, 不盲目跟风其他投资者的投资行为。

320 另一方面, 要多关注其他投资者的情绪变化。投资者情绪对股票收益率也会有一定影响, 一般而言, 投资者情绪是积极的, 股票收益率会增加; 投资者情绪是消极的, 股票收益率也会随之下降。市场中的投资者情绪在较长一段时间内保持高涨或低落时会加大投资风险, 因此投资者要关注市场中其他投资者的情绪变化。可以利用各大股吧论坛的言论对市场进行观察, 当负面评价比较多时, 要理性投资, 防范危险的发生; 当市场中大多数投资者对股市持积极态度时, 也要避免盲目跟风, 可以适当的选择买一些走势强劲的股票短期持有, 一般能够获得一些超额收益, 但要注意选择合适时机卖出股票。

325 4.2 针对金融市场监管者

一方面, 要提高投资者的专业素养, 注重对投资者的教育。投资者情绪对股市收益率的影响是十分明显的, 因此, 对于股民的专业知识素养就需要加强, 定期对投资者进行专业知识培训, 引导投资者合理对待投资者情绪, 能够有效使得股票市场健康发展, 使投资者合理看待各方面信息, 减少从众行为, 避免盲目自信, 理性投资。

330 另一方面, 要加强对金融市场的监管力度。金融市场监管者要注意到投资者情绪对股票市场产生的影响, 可以以各大股吧论坛为切入点观测投资者情绪的变化, 尤其是在发布新政策时候可以此来作为政策实施有效性的参考, 同时可以通过网络平台对股票市场中的重大变化做出回应, 缓解投资者情绪, 引导投资者理性投资, 避免投资者情绪异常波动引起股票市场的震荡。

335 5 结论

本文给出了综合投资者情绪指数和复合投资者情绪指数的对比分析, 并对复合投资者情绪指数与中证 100 指数进行实证研究, 结果表明:

340 (1) 相对于综合投资者情绪指标来讲, 文本挖掘下的复合投资者情绪指标对股市具有更强的解释力和预测力。利用文本挖掘方式得到的投资者情绪倾向能够看出, 当投资者对金融市场的未来预期持乐观态度时, 股票收益会比较好; 当股市中大部分投资者对金融市场的未来预期持悲观态度时, 股市收益会比较差。

345 (2) 滞后 2 期中证 100 指数收益率与复合投资者情绪指数正相关, 股市收益越高, 未来投资者情绪也越积极。当股市收益较高时, 投资者对股市也会持积极态度, 更多的投资者选择投资股票, 又因为网络信息传播的广泛性, 辐射范围广, 进而使得参与进来的投资者的积极性有所增加, 会导致股票收益率的增加。反之, 当投资者对股市持消极态度时, 投资者倾向于选择卖出他们持有的大量股票。持观望态度的投资者会做出更加谨慎的投资策略。当这种悲观情绪蔓延到股市时, 股价就会下跌, 并会持续一段时间。

[参考文献] (References)

- 350 [1] Long J B D, Shleifer A, Summers L H, et al. Noise Trader Risk in Financial Markets[J]. Journal of Political Economy, 1997, 98(4): 703-738.
- [2] 杨秋平. 投资者情绪与股票收益率关系的实证研究[D].东北财经大学,2015.
- [3] Gregory W Brown, Michael T Cliff. Corrigendum to "Investor sentiment and the near-term stock market"[J]. Empirical Finance 11(2004)1-27[J]. Journal of Empirical Finance, 2004, 11(4): 627-628.
- 355 [4] 王美今,孙建军.中国股市收益、收益波动与投资者情绪[J].经济研究,2004(10):75-83.
- [5] 张颖,粟坤全,袁野菁菁.投资者情绪对风险市场价格的系统影响--中、美股票市场的比较[J].国际商务:对外经济贸易大学学报,2013(4):58-66.
- [6] 薛斐. 基于情绪的投资者行为研究[D].复旦大学,2005.
- [7] 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,2010,29(04):53-57.
- 360 [8] Seo S W, Kim J S. The information content of option-implied information for volatility forecasting with investor sentiment[J]. Journal of Banking & Finance, 2015, 50(7):106-120.
- [9] 韩立岩,伍燕然.投资者情绪与 IPOs 之谜--抑价或者溢价[J].管理世界,2007(03):51-61.
- [10] M. Baker, J. Wurgler. Investor Sentiment and Cross-section of Stock Returns[J], Journal of Finance,2006,61(4),1645-1680.
- 365 [11] X. Han, Y. Li. Can investor sentiment be a momentum time-series predictor? Evidence from China[J]. Journal of Empirical Finance,2017(42):212-239.
- [12] 张征超,刘宇宸妃.我国股票市场投资者情绪指数构建及有效性检验[J].商业经济研究, 2018(7):156-158.
- [13] 杨秋平. 投资者情绪与股票收益率关系的实证研究[D].东北财经大学,2015.
- [14] Antweiler,W., M.Z.Frank.Is all that talk just noise? The information content of Internet stock message boards[J].Journal of Finance,2004,59:1259-95.
- 370 [15] Johan Bollen, Huina Mao, 2010. Twitter Mood Predicts the stock market. Working paper, available at. <http://www.ccs.edu/home/amislove/twittermood>.
- [16] 房文敏. 基于文本挖掘的投资者情绪与股价变动的关联性分析[D].青岛大学,2016.
- [17] 张伟.基于微博文本挖掘的投资者情绪与股票市场表现研究[D].青岛:山东大学,2015.
- 375 [18] 丁小虎. 投资者网络情绪与股票价格波动[D].安徽财经大学,2016.