

基金市场上投资者情绪变动及其影响研究

——基于新浪财经基金评论的统计视角

常 静,闫丽瑞,王 镭

(山西财经大学,山西太原 030006)

摘要: 该文通过文本挖掘和情感分析的技术对新浪财经基金评论量化得到基金市场投资者情绪指标,并结合上证基金指数的收益率,交易量,通过构建 VAR 模型探究基金市场投资者情绪与收益之间的相关性,以此验证投资者情绪的可靠性。由于基金市场是股市的间接反映,投资者情绪会影响基金市场收益率,而基金市场收益率不是情绪的格兰杰原因。情绪可以反映市场波动,情绪与交易量互为格兰杰因果。脉冲响应结果表明,投资者情绪会对基金市场的收益率产生短期的影响,对交易量的影响却是长期的。

关键词: 文本挖掘;基金市场;投资者情绪

中图分类号: F830.9 **文献标识码:** A **文章编号:** 1671-2404(2018)84-0028-07

1 背景

传统金融理论认为市场是有效的,但是在现实情况中,传统金融学的假设过于苛刻,理论也难以起到指导作用。学者们开始寻求突破,这种背景下产生了行为金融理论。行为金融学将心理学尤其是行为科学的理论融入到金融学之中,近年来迅速发展并取得了一定的研究成果。投资者情绪理论最早由 *Delong* 提出,之后受到广大学者的关注。

随着互联网的普及,新浪等平台会定期发布市场相关消息,而且越来越规范化。股吧、微博等社交平台的网络舆论中,蕴含着非常有价值的投资者情绪,这些情绪可用于分析和理解当前中国金融市场表现,为投资者和国家相关部门提供决策支持。目前已有不少学者通过微博评论,东方财富网获取投资者情绪,研究情绪与股市之间的关系。采用计算机自动地分析这些主观性文本表达的情感,也称为目前学术界研究的一个热点,这个方向就是文本情感分析。

基金市场作为中国主体市场,通过研究基金市场收益与情绪的相关关系,基金经理选股时是否受到情绪的影响。本文通过对投资者行为及心理状态进行文本挖掘,研究基金市场与投资者情绪的内在

影响机制。利用这种相关关系对市场未来进行预测,以期对投资者理性决策、促进市场健康运行有一些启发。

2 文献综述

2.1 投资者情绪的定义

Delong 等(1990)认为情绪以及有限套利的存在,会导致资产价格经常偏离正常水平,因此,投资者情绪是影响资本市场定价的系统性风险。*Baker* 和 *Wurgler*(2006)将情绪定义为投资者基于对资产未来现金流和投资风险的预期而形成的一种信念,但这一信念并不能完全反映当前已有的事实。*Brown* 和 *cliff*(2004)把情绪视为投资者对股票市场总体乐观或悲观。*黄德龙*等(2009)认为投资者情绪包含虚拟经济和实体经济两部分。分别是投资者的投机性需求和对上市公司盈利前景的总体乐观程度。易志高等对情绪的认识是由于未来不确定性的存在,投资者对于资产所能创造的未来现金流大小及风险变化,当前很难做出准确的评估,只能对其形成一种预期或信念。而这种预期或信念除了与资产基本面相关外,还与投资者自身的教育经历、投资经验和知识、社会背景、所拥有信息、个性和偏好等密切相关,是投资者对资产未来价值的一种“主客观”综合评估。因此,对于同一资产,不同的人会有不同的信念,即称之为“情绪”。

2.2 投资者情绪的测度

投资者情绪测量是研究情绪问题的一个基本内

收稿日期:2017-08-15

作者简介:常静,硕士研究生,主要从事行为金融等方面的研究;闫丽瑞,博士,副教授,主要从事金融风险管理等方面的研究;王镭,硕士研究生,主要从事行为金融等方面的研究。

容。目前情绪度量的方法主要有以下几种:

一是直接指标法。国外主要是投资者智慧指数和美国个体投资者协会指数。中国现有的调查数据有央视看盘,好淡指数等。*Brown*和*Cliff*(2005)用投资者智慧作为情绪代理变量,证明了定价误差与情绪成正比。王美今和孙建军(2004)基于央视看盘指数构建情绪指标,构建了理论模型表明投资者情绪是影响收益的系统性因子。

二是复合指标法。*Baker*和*wurgler*(2006)将封闭式基金折价率、股票换手率、*IPO*上市首日平均收益率、新股发行数量、股利溢价等在剔除宏观因素影响的基础上进行主成分分析,得到综合投资者情绪指标。易志高采用封闭式基金折价,*IPO*数量,上市首日收益,消费者信心指数,新增投资者开户数等6个单项指标构建,研究发现投资者情绪与股市走势基本保持一致的变动趋势。

三是文本挖掘法。目前这种方法越来越受到广泛关注。网络发帖内容直观地刻画投资者的心理活动,反映他们对市场的预期,可以作为衡量投资者情绪的代理变量。此外,信息技术的发展使得网络评论获取更加方便。*Kim*(2014)运用python爬取雅虎财经数据,利用词典法和贝叶斯两种方法构建情绪指数,结果发现情绪对美国股市的影响不显著。易洪波等(2015)对东方财富网主题发帖内容进行词频统计来构建情绪指数,结果表明投资者多空双方情绪对交易市场成交量和收益率存在非对称影响。空方情绪对成交量的影响明显,多方情绪对收益率的影响更明显。

2.3 文本挖掘

文本挖掘是以文本型信息源作为挖掘对象,利用定量计算和定性分析的办法,从中寻找信息的结构、模型、模式等隐含的具有潜在价值的新知识的过程。由于文本类型数据的快速增长,文本挖掘的重要性也日益增加。

文本挖掘的任务:

文本预处理:文本预处理的目的是使得文本数据结构化,从而方便利用现有的数据挖掘技术对其进行分析。文本预处理的过程包括分词、特征表示、特征提取等一系列操作。

文本分类:文本分类是指按照预先定义好的类别,为文档集中的每个文档确定一个或者几个类别。文本分类是一个有监督的机器学习过程,一般

分为训练,分类,和测试三个阶段。通过对一个训练文档集进行人工标注类标签的方式构建分类模型,然后计算机基于这种学习得到的关系模式自动将一篇文档归入所属类别。

文本聚类:与文本分类不同的是,文本聚类没有预先定义好的类别,是一种典型的无监督的机器学习方法。文本聚类的目标是将文档集合分成若干个簇,使得同一簇内文档内容的相似度尽可能地大,而不同簇之间的相似度尽可能地小,并为每个簇给出一个主题摘要。

关联规则指的是寻找同一事物中不同变量同时出现的规律性。关联规则最初用于发现数据库中项或属性间的关系,在文本分析中指的是从文档集合中找出不同词语之间的有意义的关系。

趋势预测是指通过对已有文档的分析,推测出特定数据在将来某个时刻的取值情况。如果考虑文本数据的时间维度,那么文本数据可以看作一类特殊的时序数据,能够反映一段时间内文本主题的变化情况,因此用来预测下一时间段文本主题的发展变化趋势。

在研究广度上,目前文本挖掘研究涉及到计算机,企业经济,工程学,医学信息学,新闻传播学等多个学科,同时,情报分析、专利分析、情感分析、舆情分析、热点发现、知识管理等成为国内外学者关注的热点话题。在技术深度方面,国内研究局限于文本分类、文本聚类等基本文本挖掘技术,而国外论文还广泛涉及了自然语言处理、机器学习等在计算机领域运用较广的研究方法。

情感分析又称观点挖掘,一般应用文本分类技术对带有情感色彩的主观性文本进行处理,主要是对媒体评论和产品评论的情感分析。*Ceron*通过对*Twitter*上关于政治领导人和选举的帖子进行情感分析,发现社会化媒体可以有效地监测舆情和预测民意。叶强等提出了一种针对中文互联网客户评论的主观情感自动判别方法,消费者据此可以了解其他用户的态度倾向,做出更好的购买决策。

2.4 投资者情绪与金融市场

目前大部分文献研究投资者情绪与股票市场收益率之间存在互动关系。如蒋玉梅和王明照(2010)研究表明情绪与短期市场收益正相关,与长期市场收益负相关,投资者情绪对股票收益产生系统性影响。杨阳和万迪昉(2010)实证分析结果表

明中国投资者情绪对股票市场的收益无显著影响,而股票收益却对投资者情绪的作用显著:牛市阶段股票收益与情绪正相关,而熊市阶段则为负相关。刘丽文和王镇(2016)发现基于公司特征、交易特征、所属行业这三个维度划分的各种类型股票组合在情绪乐观期的收益均大于在情绪悲观期的收益,且各类型股票组合的收益在不同情绪期下也差异较大,说明投资者情绪的确在很大程度上影响了股票组合的收益。当然,也有学者认为可以用情绪来解释股市异象。吕东锴和蒋先玲(2013)尝试用情绪来解释 IPO 溢价,即 IPO 发行价与 IPO 溢价之间存在显著负相关关系。

采用 Tobit 模型发现 IPO 发行价越低,个人投资者持股比例越高,机构投资者持股比例越低。正是由于低发行价股票投资以个人投资者为主,个人投资者易受投资者情绪影响,高估股票价值,增加非理性投资需求,才导致了“发行价溢价异象”产生。汪昌云和武佳薇(2015)采用文本分析技术构建媒体语气作为情绪的代理变量,研究结果表明,负面媒体语气能够更好地解释 IPO 抑价率、首日换手率以及超募比例。ZX Ni, DZ Wang, WJ Xue(2015)研究发现投资者情绪在1月至24月对股市的影响都是显著的。短期内为正向影响,6个月及以后为负向关系,体现了长期反转效应和短期动量效应。中国股民有明显的认知偏差和投机倾向。

有少许文献开始探讨投资者情绪对期货市场的影响,如陈志毅(2017)采用主成分构建期货市场情绪指标,研究发现投资者情绪、沪深300指数和沪深300指数期货价格三者相互影响,投资者情绪增强了沪深300指数期货价格对沪深300指数的影响,削弱了沪深300指数对沪深300指数期货价格的影响。

由于目前尚没有研究基金市场与投资者情绪的文献,所以本文利用行为金融学的相关理论研究基金市场情绪与收益间联动关系。主要贡献在于1)利用网络评论的内容作为投资者情绪的数据,然后情感分析对评论本文进行分析,最终对基金市场情绪指标进行度量,2)利用 VAR 模型,研究基金市场情绪与收益率之间的相互关系。

3 数据选取及处理

研究的样本区间为2009年4月1日至2017年4

月28日,共1965个交易日,研究对象为基金市场,日交易量和日收益率数据来源于 wind 资讯。

文本挖掘构建情绪流程:

3.1 网络爬虫

网络爬虫是按照一定的规则,自动地抓取万维网信息的程序或者脚本。通过分析互联网的组织结构,web 页面的超链接关系结构及 web 文档的内部结构从 url 目录路径结构中对文本进行挖掘。新浪财经网页市场划分明确完整,而且布局规整,便于文本爬取及分析。因此文本内容来源于新浪财经基金市场的基金评论版块。采用 R 软件的 rvest 包,XML 包,编写程序从网页源代码中批量获取文本数据。

3.2 文本预处理

文本预处理的目的是使得文本数据结构化,从而方便利用现有的数据挖掘技术对其进行分析。文本预处理的过程包括分词、特征表示、特征提取等一系列操作。词汇是文本情感分析的基础,篇章的情感分析要由词汇来实现,因此需要对句子进行分词。在此过程中,使用 jiebaR 包对中文文本进行分词。同时为了简化操作,运用 tm 包对语料库进行清洗,去除不必要的标点符号,数字及英文字母。

3.3 情感词抽取

情感词又称极性词、评价词语,特指带有情感倾向性质的词。通常情况下,情感词有褒义和贬义两类倾向性。情感词抽取和判别既是词汇级情感分析的基础工作,更是句子级和篇章级情感分析的基础。基于词所代表的金融情感色彩及专业投资者对市场的看法,人工筛选建立了积极(市场看多)、消极(市场看空)情绪词典。词典法具有获取情感词全面,准确的优点。

3.4 构建市场情绪(频数)

对每个交易日构建一个相对应的情绪指标。分别对对网页爬取下来的每日文本中积极,消极情绪单词进行频数统计,计算每日情绪得分。

采用文本挖掘的方法批量获取新浪财经基金评论内容,对该文本数据进行量化分析,构建情绪指标。人工筛选法建立金融情绪词典,分为积极情感词典和消极情感词典。积极情感词典包含296个关键词,消极情感词典包含342个关键词。

在金融情感词典的基础上,对每日文本内容的单词进行匹配并统计频数,以此来构建情绪指标。部分词典内容如下:

表 1 部分金融情感词典

积极情感词典	暴增,利好,扭转颓势,跳升,狂飙,看多,攀升,上扬
消极情感词典	暴跌,低迷,跌停,连降,卖空,陷入困境,骤减,亏损额

根据情绪词典对挖掘文本进行频数统计,以此

$$\text{构建情绪指数: } Sent_t = \frac{fre_{pos,t} - fre_{neg,t}}{fre_{pos,t} + fre_{neg,t}}$$

其中 $fre_{pos,t}$ 表示当天文本中表示积极情感的单词的频数, $fre_{neg,t}$ 表示当天所有文本中代表消极情感的单词的频数。

基金市场的交易量和收益率数据来自于上证基金指数,日交易量和日收益率分别采用对数成交量和对数收益率来进行计算,即: $vol_t = \ln(volume_t)$

$$r_t = \ln(P_t) - \ln(P_{t-1})$$

其中 $volume_t$ 为 t 日上证基金指数的交易量, P_t 为 t 日上证基金指数的收盘价, P_{t-1} 为 $t-1$ 日上证基

金指数的收盘价。

4 实证检验

4.1 变量的描述性统计

表 2 变量的描述性统计

	均值	标准差	最大值	最小值
ret	0.0003	0.012	0.065	-0.076
sent	-0.018	0.25	1.00	-0.90
vol	9.161	0.25	10.3	8.376

从上表可以看出,我们把投资者情绪指数限定在 -1 至 $+1$ 这个区间内,投资者情绪均值为负值,表明中国资本市场牛短熊长,投资者情绪低落期长于高涨期,这与已有研究一致。情绪与交易量的标准误相同,进一步表明情绪可以反映出市场交易量的波动。

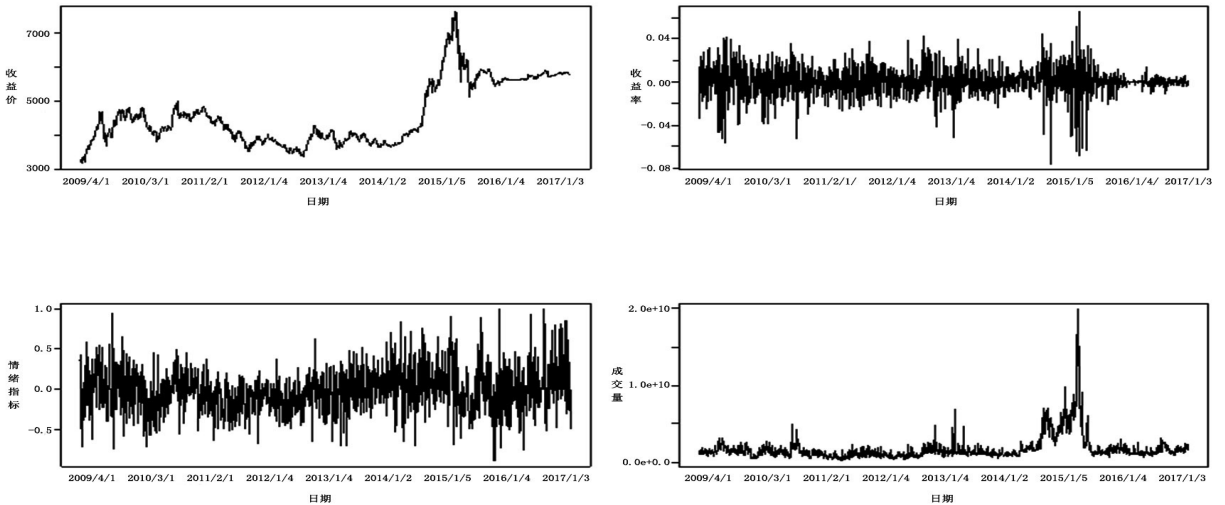


图 1 收盘价、收益率、投资者情绪及成交量序列

从收盘价与情绪指标走势图来看,在 2009 年 4 月至 2010 年 4 月基金市场收盘价上涨的同时,市场情绪呈现积极状态。4 月之后短暂下跌的过程中,情绪指标总体为负,但是有上升的趋势,市场在 2010 年 7 月底出现回升迹象,表明情绪对市场价格有起到先行指标的作用。在 2011 年 11 月至 2013 年 1 月股市持续下跌的过程,情绪不太明朗。伴随着 2014 年至 2015 年收盘价节节攀升,情绪呈现剧烈波

动。投资者情绪与收盘价总体保持一致。

收益率与情绪的一致关系主要表现在波动上。继 2009 年和 2010 年波动较大的收益率之后,2011 年收益率小幅波动,可以看到市场情绪是呈现消极状态。观察投资者情绪与交易量,可以明显看到在 2015 年 7、8 月份交易量呈现异常增长时,投资者情绪也存在异常波动,情绪可以反映出市场波动。

4.2 平稳性检验:收益率和情绪指标在 1% 的水平下平稳,在所研究样本期间,交易量有过短暂的剧烈波动,但是仍然在 5% 水平下显著。

表 3 变量的平稳性检验结果表

变量	ADF 值	P 值
ret	- 11.333	0.01
sent	- 6.9191	0.01
vol	- 3.6242	0.03034

4.3 Johansen 检验表明至少存在三个协整关系,投资者情绪,基金市场收益率与基金市场交易量间存在着长期关系

表 4 协整检验结果表

	test	10pct	5pct	1pct
$r \leq 2$	74.55	6.5	8.18	11.65
$r \leq 1$	511.75	15.66	17.95	23.52
$r = 0$	1400.97	28.71	31.52	37.22

4.4 格兰杰因果检验

表 5 格兰杰因果检验结果表

原假设	F 统计量	P 值
return 不是 sent 的格兰杰原因	0.0972	0.7553
sent 不是 return 的格兰杰原因	8.9487	0.002811
vol 不是 sent 的格兰杰原因	4.713	0.03006
sent 不是 vol 的格兰杰原因	15.337	9.30E - 05

首先查看变量间的 Granger 因果关系,为建模提供合理性。由上表可以看出,情绪指数是收益率的格兰杰原因,但是基金收益率变动对情绪没有显著影响。结果不同于股票市场收益率与情绪互为格兰杰因果,这可能是因为基金市场是股市的一个间接反映,投资者情绪可以影响市场收益率,基金市场机构投资者居多,收益率对情绪的影响不显著。情绪和交易量互为 Granger 因果,投资者情绪会影响交易量的变动,交易量大小也会影响市场情绪,交易量变动可以间接反映投资者情绪,情绪也可以反映交易量的波动。因果检验表明文本挖掘技术构建的情绪指标是可靠的,可以代表基金市场上投资者情绪。

4.5 脉冲响应函数

对投资者情绪、收益率和交易量做 VAR 模型,滞后阶数确定为 6。进一步分析该 VAR 模型的脉冲响应图。

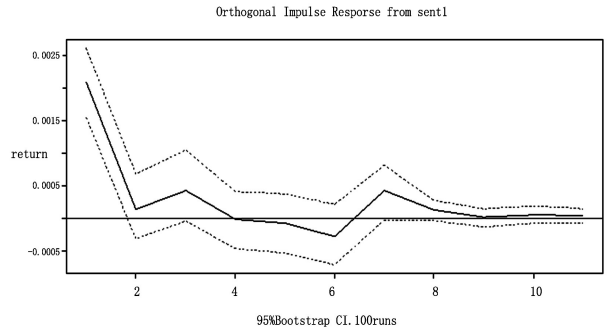


图 2 情绪受到冲击对收益的脉冲响应图

从图中可以看出,当基金市场的情绪受到一单位正向冲击时,收益在第一期产生最大的正向的响应,约为 0.0021;之后开始回落,最终趋近于零。这和行为金融的理论是相符的,当市场上的投资者情绪为积极的时候,人们更乐意投资,收益率为正,反之亦然。

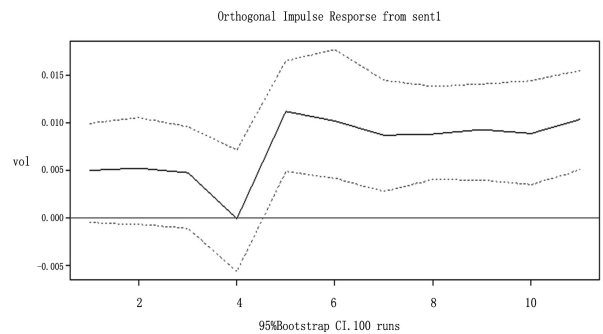


图 3 情绪受到冲击对交易量的脉冲响应图

从图中可以看出,当基金市场的情绪受到一单位正向冲击时,交易量产生正向的响应,在第五期达到最大,约为 0.012。这和行为金融的理论是相符的,当市场上的投资者情绪为积极的时候,人们更乐意投资,交易量随之变大,又由于“羊群效应”,投资者间存在跟风行为,因此这个正向的影响是持续的。当市场情绪为消极时,交易量也会减少。

5 总结与建议

基于网络评论分析的投资者情绪对基金市场的

成交量和收益率都有显著影响,投资者情绪是基金市场收益率的格兰杰原因,情绪受到一单位冲击后,对收益率会产生正的影响,两期后逐渐趋于零。投资者情绪与基金市场交易量互为格兰杰因果。情绪受到冲击后,交易量会产生长期正的响应。情绪对于交易量冲击的响应也是长期的。

本文证实了基金市场投资者情绪与基金市场收益率和交易量间的联动关系,结果支持行为金融理论,投资者非理性情绪实际上会影响资产的价格。个人投资者和机构投资者在参考网络评论进行投资时,应该了解情绪会对自身以及管理者投资策略的影响,由此改进自己的投资方式。监管者加强对网络论坛的监管,防止不法分子操纵舆论影响投资者,尤其是情绪的转变可能引起的市场泡沫及非理性繁荣。

中文文本挖掘一方面要重视基本理论与框架的研究,另一方面要重视针对汉语本身的研究,要研究中文文本的构成特点与特征提取机制,只有在对中文文本的分析技术取得进展后才能为后面的知识挖掘提供强有力的支持。此外,文本挖掘不应局限于分类、聚类等简单知识的发现,可以探索更丰富的知识挖掘方法,如语义规则发现、趋势分析、主体跟踪等。

参考文献

- [1] Baker, Malcolm, Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns[J]. Journal of Finance, 2006, 61:1645-1680.
- [2] Brown, G.W., M.T.Cliff. Investor sentiment and the near-term stock market[J]. Journal of Empirical Finance, 2004, 11(1):1-27.
- [3] Brown, Gregory W, Michael T. Cliff, Investor Sentiment and Asset Valuation[J]. Journal of Business, 2005, 78(2):405-440.
- [4] Delong, J.B., Shleifer A, Summers L.H., Waldmann, R.J., Noise trader risk in financial markets[J]. Journal of Political Economy, 1990, 98(4):703-738.
- [5] SH KIM, D kim. Investor Sentiment from Internet Message Postings and Predictability of Stock Returns[J]. Journal of Economic Behavior and Organization, 2014, 107(PB):708-729.
- [6] ZX Ni, DZ Wang, WJ Xue. Investor sentiment and its nonlinear effect on stock returns—New evidence from the

Chinese stock market based on panel quantile[J]. Economic Modelling, 2015.

- [7] 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):75-93.
- [8] 易洪波,赖娟娟,董大勇.网络论坛不同投资者情绪对交易市场的影响—基于VAR模型的实证分析[J].财经论坛,2015(1):46-54.
- [9] 易志高,茅宁,汪丽.投资者情绪测量研究综述[J].金融评论,2010,02(3):113-121.
- [10] 黄德龙,文凤华,杨晓光.投资者情绪指数及中国股市的实证[J].系统科学与数学,2009,29(1):1-13.
- [11] 蒋玉梅,王明照.投资者情绪与股票收益:总体效应与横截面效应的实证研究[J].南开管理评论,2010,13(3):150-160.
- [12] 杨阳,万迪昉.不同市态下投资者情绪与股市收益,收益波动的异化现象——基于上证股市的实证关系[J].系统工程,2010(1):19-23.
- [13] 吕东锴,蒋先玲.IPO发行价溢价异象与投资者情绪研究[J].经济与管理,2013,27(1):32-36.
- [14] 刘丽文,王镇.投资者情绪对不同类型股票收益影响的实证研究[J].金融理论与实践,2016(2):90-97.
- [15] 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.
- [16] 陈志毅.基于投资者情绪的沪深300指数期货与指数价格关系[J].特区经济,2017(4):69-74.
- [17] 徐德金,张伦.文本挖掘用于社会科学研究:现状、问题与展望[J].科学与社会,2015.
- [18] Ceron A, et al. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preference with an application to Italy and France[J]. New Media & Society, 2014, 16(2):340-358.
- [19] 叶强,张紫琼,罗振雄.面向互联网评论情感分析的中文主观性自动判别方法研究[J].信息系统学报,2007,1(1):79-91.

Research on the Changes of Investor Sentiment in the Fund Market and Its Impact——from the statistical perspective based on Sina Finance Fund commentary

CHANG Jing, YAN Lirui, WANG Lei

(*Shanxi University of Finance and Economics, Taiyuan Shanxi Province 030006, China*)

Abstract: Using text mining and sentiment analysis techniques to quantify Sina financial fund comments to obtain investor market sentiment indicators of the fund market, combined with the yield and trading volume of the SSE fund index, the paper explores the correlation between investor sentiment and return in the fund market by constructing a VAR model to verify the reliability of investor sentiment. Since the fund market is an indirect reflection of the stock market, investor sentiment will affect the rate of return of the fund market, but it is not an emotional Granger cause. Emotions can reflect market fluctuations, and emotions and trading volume are Granger causality. The results of the impulse response show that investor sentiment will have a short-term impact on the yield of the fund market, but the impact on trading volume will be long-term.

Key words: text mining; fund market; investor sentiment