



基于分位数 Granger 因果的网络情绪与股市收益关系研究

许启发^{1,2}, 伯仲璞¹, 蒋翠侠¹

1 合肥工业大学 管理学院, 合肥 230009

2 合肥工业大学 过程优化与智能决策教育部重点实验室, 合肥 230009

摘要:行为金融学理论认为,股票市场的价格变动除受宏观基本因素影响外,还在很大程度上受众多个体投资者或噪音交易者行为左右。中国股票市场拥有庞大的个人投资者群体,且股民群体与网民群体之间具有高度耦合性,使用网络情绪等信息能够探索中国股市收益变动基本规律。

为揭示个体投资者行为对股市收益的影响,以个体投资者情绪为视角,以网络环境中个体投资者的情绪信息为切入点,检验投资者情绪与股市收益的关联关系,评估网络情绪信息价值。使用中文文本情感分析方法,从新浪微博文本中提取出网络情绪时间序列;分别运用均值Granger因果和分位数Granger因果检验方法,探讨网络情绪波动与股市收益之间是否存在因果关系;将股票市场发展阶段进行细致划分,研究不同市场阶段下网络情绪波动与股市收益之间的因果关系。

对沪深300指数收益进行实证研究,结果表明,尽管在均值框架下网络情绪波动与股市收益之间因果关系并不明显,但基于分位数Granger因果分析却发现两者在极端分位点区间处存在广泛且显著的因果关系。数据显示,在40个因果关系检验中,分位数Granger因果检验的因果关系发现了23个显著的因果关系,发现率为57.5%,远高于均值Granger因果检验的7.5%。此外,股市收益受到网络情绪波动影响的程度和方式在不同市场阶段下有所不同。

研究结果具有一定的理论意义和应用价值。在一些特定分位点区间网络情绪波动对股市收益存在显著因果关系影响,这在特定条件下股市收益的可预测性提供了佐证。网络情绪能够预测股市收益的尾部(上尾或下尾)行为特征,可以为金融风险防范提供决策参考。研究结果为股票市场的定价、收益预测和波动率估计等相关研究提供了新的研究思路,也为网络情绪信息使用提供了新的方向。

关键词:分位数回归; Granger因果检验; 网络情绪; 股市收益; 微博数据

中图分类号:F830.91

文献标识码:A

doi:10.3969/j.issn.1672-0334.2017.03.013

文章编号:1672-0334(2017)03-0147-14

引言

股市收益变化及预测一直是学界和业界关注的

焦点。有效市场假说认为,在信息透明、充分竞争的市场环境下,无法利用公开的市场信息对股市收益

收稿日期:2016-08-20 **修返日期:**2017-03-21

基金项目:国家自然科学基金(71671056); 国家社会科学基金(15BJY008); 教育部人文社会科学研究规划基金(14YJA790015)

作者简介:许启发,管理学博士,合肥工业大学管理学院教授,研究方向为数量经济理论与方法、金融计量等,代表性学术成果为“Composite quantile regression neural network with applications”,发表在2017年第76卷《Expert Systems with Applications》, E-mail: xuqifa1975@126.com

伯仲璞,合肥工业大学管理学院硕士研究生,研究方向为金融计量等, E-mail: zp2576302465@163.com

蒋翠侠,管理学博士,合肥工业大学管理学院副教授,研究方向为金融时间序列分析和金融计量等,代表性学术成果为“基于LASSO分位数回归的对冲基金投资策略研究”,发表在2016年第3期《管理科学学报》, E-mail: jiangcx1973@163.com

进行预测。然而,现实市场环境理想状态相去甚远,一些研究发现,在一定时空范围内,股市收益具有一定的可预测性。NOFSINGER^[1]研究发现社会情绪与金融行为和金融市场之间存在相互作用;PRECHTER^[2]以人类社会行为的波动影响现象解释金融市场;还有以社会行为视角研究股票市场^[3]。这些研究都从行为金融的角度为股票市场的预测提供了可行的着眼点。行为金融将行为科学引入金融分析,旨在揭示金融市场的非理性行为和决策规律,已有的行为金融的相关研究从理论和实证两个层面论证了投资者情绪与股票市场之间具有显著的相关性。YAO et al.^[4]研究证明中国股票市场中存在明显的羊群效应,这一结论同样可以从情绪传递的角度解释;HUANG et al.^[5]构建了一种投资者情绪代理变量,并实证了投资者情绪对股票市场的预测能力;KIM et al.^[6]通过分析YAHOO金融信息内容测量投资者情绪,进而预测股票市场走势。

随着互联网的迅速普及和快速发展,一方面网络行为势必影响到金融市场运行机制;另一方面,网络行为在一定程度上也必然反映了金融市场的运行状况,网络行为中无疑蕴藏着金融市场变化的“蛛丝马迹”。JOSEPH et al.^[7]使用谷歌搜索数据预测股票市场的异常波动;张谊浩等^[8]尝试使用百度搜索数据预测股票市场波动,结果良好;洪涛等^[9]从中国网民在线搜索行为中挖掘出了用于预测房价的有用信息。此外,中国股票市场具有个人投资者群体庞大的特点,股民群体与网民群体之间呈现出高度耦合的特性,使用网络情绪等信息探讨中国股市收益变动基本规律具有可行性和必要性。

已有相关研究多采用基于均值回归的研究方法,适合于探讨在中心位置或正常环境下网络情绪对股市收益的影响,存在一定局限。本研究认为,中国网络情绪与股票市场之间确实存在某种联系,但这种联系未必表现在中心位置,更有可能存在于尾部极端分位区间处。本研究借助网络社交平台,通过情感分析方法,提取网络情绪信息;综合应用均值Granger和分位数Granger因果检验,探讨不同市场环境下网络情绪对股市收益的影响,并进一步刻画该影响存在的分位区间、影响程度和影响方式等,深入揭示网络情绪对股市收益的预测能力。

1 相关研究评述

在已有关于使用网络情绪数据分析股票市场的研究中,早期研究多以相关网络论坛或网站为情绪数据来源。GILBERT et al.^[10]从社交网站Live Journal上的2 000万文章中提取了焦虑、担忧和恐惧,在Granger因果分析框架下发现焦虑情绪的增长能够用于预测S&P 500指数的下降。还有其他更早期的研究,GRUHL et al.^[11]使用博客、媒体和网页中关于50万册书籍的相关贴文,预测书籍的销量排名;MISHNE et al.^[12]使用博客数据预测了电影的票房走势;LIU et al.^[13]借助PLSA模型提取博客中的情绪指标,进而成功预测了

产品未来销量;PENG et al.^[14]的研究表明股票市场受投资者情绪和顾客满意程度两大群体情绪的影响。

随着网络社交平台的兴起,研究者们逐渐将目光转向网络社交平台,国外相关研究中多以Twitter为主。ZHANG et al.^[15]收集了6个月内的Twitter数据,并从中随机抽选百分之一用于测量每日Twitter中的hope和fear值,研究发现情绪化Twitter的占比与Dow Jones、NASDAQ和S&P 500都表现出显著负相关,与VIX(CBOT Volatility Index)表现出显著正相关,从而证明Twitter中的情绪信息具有巨大价值。更进一步,BOLLEN et al.^[16]着眼于研究Twitter中影响股票市场的具体情绪种类,通过对Twitter文本进行自然语言处理,提取出Twitter中包含的情绪信息,根据提取方法的不同,共提取出7个不同维度情感时间序列(分别为polarity、calm、alert、sure、vital、kind、happy),并将其作为参数做Twitter mood与股票市场之间的Granger因果关系检验,结果表明Twitter中的calm与股票市场之间存在较明显且持久的因果关系,还发现加入了情绪数据的自组织模糊神经网络在股票市场的预测上表现良好。COREA et al.^[17]尝试使用Twitter数据预测NASDAQ的未来走势,结果显示加入Twitter数据的预测模型具有更高的预测精度。WEI et al.^[18]不再具体分析Twitter中的情绪信息,转而使用与股票相关的Twitter的数量分析股票市场,发现Twitter量的波峰与股市波动率、隐含波动率之间具有密切的关联性,并根据研究结论提出了新的股票交易策略。

此领域研究中,也不乏着眼于中国股票市场的研究工作,此类研究多以新浪微博为网络数据来源。赖凯声等^[19]使用协整检验和误差修正模型发现,微博情绪综合指数与同期的上证指数以及下一个交易日的上证指数之间存在显著的长期均衡关系,证明微博与股票市场之间存在相关性。黄润鹏等^[20]使用ROST Content Mining软件对新浪微博文本进行处理,提取出根据积极和消极程度划分的7个情绪倾向时间序列,使用Granger因果分析发现高度积极情绪时间序列的变化能够预测滞后4天的上证指数收盘价变化,引入机器学习的方法,发现加入情绪倾向数据的支持向量机在股票预测方面比对照组提升了13.62%的准确率。朱南丽等^[21]引入博客和微博发布量作为对投资者关注度的代理指标,通过对沪深300指数股的实证研究发现,博客和微博量能够作为投资者关注的有效度量,特别对个人或中小投资者的度量效果最佳。此研究成果一定程度上证明了网络用户与股民之间存在较大重叠或存在较强相互影响,为使用网络情绪做投资者情绪的代理变量提供了实证支持。孟雪井等^[22]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,结合百度关键词推荐系统获得投资者的网络搜索关键词,进而采用时差相关系数法、随机森林算法、因子分析法等将对上述关键词的百度指数时间序列综合成沪市投资者情绪指数。

上述国内外研究,从实证角度证明了网络平台

数据在股票市场预测中的有效性,但尚有不足。首先,多数研究未能细致区分不同市场状态(牛市、熊市以及正常波动情况)下网络情绪与股市收益的相关性;其次,已有研究主要在均值回归框架下开展研究工作,能够揭示投资者情绪对股市收益平均水平的影响,却无法刻画其在不同分位点处的异质影响模式。为此,本研究综合使用均值 Granger 和分位数 Granger 因果检验,探讨网络情绪波动对股市收益的影响,以期有新的发现。

2 网络情绪分析和测量

2.1 网络情绪界定

(1) 实质为情绪氛围

网络情绪并非网络世界中部分人的情绪,而是整个网络环境中的情绪氛围。网络中每一个参与者的个人情绪共同构成了网络中的情绪氛围,这种情绪氛围并非个人情绪的简单相加,而是由所有参与者之间相互影响、共同作用形成。一方面,所有个人参与者的情绪共同构成网络中的情绪氛围,情绪氛围中包含了每个人的情绪;另一方面,情绪氛围会反向影响每一个参与者的情绪,参与者会受同一网络中其他人情绪的感染,改变甚至逆转其原有情绪。曾小梦^[23]研究表明中国网络环境中的个体易受整体情绪左右,参与者之间具有较强的相互作用。这种情绪间的相互影响和共同作用,受参与者交流的畅通程度影响,当参与者之间的交流畅通时,交流频次和质量都有所提升,从而互相间情绪的影响作用也越强。

(2) 范围为网络社交平台

伴随着互联网络的发展,网络社交及其平台得到广泛普及并被接受。尽管网络用户的情绪信息会散布在互联网络世界的各个角落,但主要集中在诸如博客、微博、论坛、QQ等网络社交平台中。新浪微博是目前中国最大的微型博客服务类社交网站,其用户可以通过网页、WAP页面、手机客户端、手机短信、彩信等形式参与微博互动。根据新浪微博2016年5月12日发布的第一季度财务报表显示,截至2016年1季度末,新浪微博月活跃用户达到2.61亿人,日活跃用户1.2亿人。考虑到新浪微博庞大的用户基数及其即时性、开放性、贴近用户等特点,本研究选取新浪微博作为网络社交平台代表开展研究工作,以下网络情绪数据如非特别声明均来自新浪微博。

(3) 类型为多种情绪类型

人的情绪有喜、怒、哀、乐等多种类型,每个人在同一时间其情绪状态可能同时包含两种或多种不同类型,不同类型的情绪对人的行为影响的强度、方向和作用时间可能都有所不同。本研究在探讨网络情绪与股市收益之间关系时,考虑多个不同情绪维度,探讨不同情绪类型对股市收益的异质影响。

2.2 网络情绪分析

网络社交平台中,用户与用户之间的交流方式主要是文字和表情符号,用户在发布微博时,其实时

情绪蕴藏于微博文本之中。因而,微博文本中包含着用户的实时情绪信息,本研究使用中文情感分析方法实现微博文本的情绪信息提取。

情感分析是自然语言处理下的一个分支学科,目前已有的文本情感分析以基于词典的匹配方法和基于机器学习的分类方法为主,其中机器学习分类方法常见的有朴素贝叶斯^[24]、支持向量机方法^[25-26]、神经网络方法^[27]、K近邻方法^[28]以及其他方法^[29]。由于此类方法需要大量训练样本且参数设置复杂,因此在中文情感分析领域目前比较成熟且常用的方法是基于情绪词典的情绪分析方法,如马秉楠等^[30]使用的表情符号词典和蒋盛益等^[31]使用的面向微博语料的语义词典。词典类方法的主要步骤有:①对经过初步数据清洗后的微博文本进行分词,即将完整的语句根据中文语言习惯及语义将其分割为若干词汇或短语;②对每一条微博文本中的情绪因子(词、短语、常见搭配、表情符号等)与情绪词典进行匹配并求和计算,获得每一条微博的情绪向量。提取流程见图1。

中文文本情感分析效果的关键在于情绪词典的质量,本研究综合大连理工大学建立的中文情感词汇本体库和台湾大学建立的中文情感极性词典构建本研究所用情感词典。中文情感词汇本体库是大连理工大学信息检索研究室整理和标注的一个中文本体资源,该情感分类体系在国外比较有影响的Ekman的6大类情感分类体系的基础上,将词汇本体加入情感类别“好”,对褒义情感进行了更细致的划分,最终词汇本体中的情感共分为7大类21小类。本研究使用其中的7大类划分方式。台湾大学建立的中文情感极性词典数据是基于文本情感二元划分方法的一个中文词语数据库,它将11 086个词语分为2 810个积极属性词语和8 276个消极属性词语。在此基础上,本研究将Boson NLP提供的网络情感词库添加进已有词库。Boson NLP情感词典是玻森中文语义开放平台(<http://bosonnlp.com/>)从微博、新闻、论坛等数据来源的上百万篇情感标注数据中自动构建的情感极性词典。因为标注包括微博数据,该词典囊括了很多网络用语及非正式简称,对非规范文本有较高的覆盖率。

表1给出文本情感分析工作使用的情绪词典示例,(a)和(b)分别为前文中所述用于提取7类情感维度的中文情感词汇本体库和用于提取情感极性的中文情感极性词典。此外,本研究还整理并设置了(c)反义词库和(d)程度词库(分5个等级)。当文本中出现“程度词+情绪词”时,该情绪词对应情绪值会加上程度词权重;当出现“反义词+情绪词”时,该情绪会逆转(仅适用于部分情感,如情感极性、开心和悲伤、好和怒)。本研究还对表情符号表达的情绪进行手动设定,表情符号在文本形式的微博中会以“[表情名]”的形式展现,如[开心]。根据所使用情绪词典,本研究得到情感极性、乐、好、怒、悲、惧、厌恶、惊讶共8个维度的日情绪时间序列。

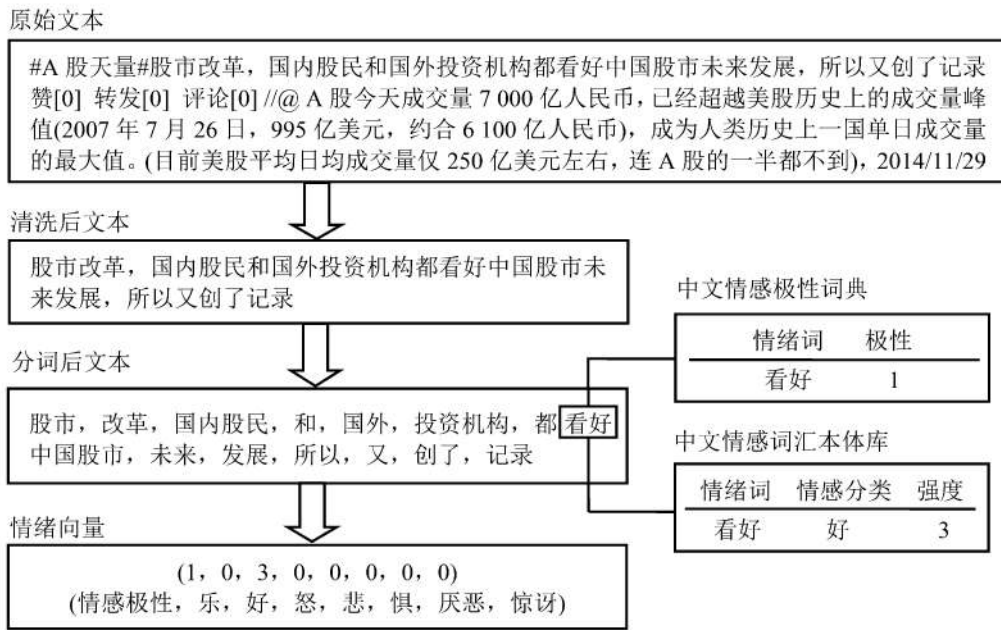


图1 情绪向量提取流程
Figure 1 Procedure of Sentiment Vector Extraction

表1 情绪词典示例
Table 1 Example of Sentiment Word Dictionary

(a) 中文情感词汇本体库			(b) 中文情感极性词典		(c) 反义词库	(d) 程度词库	
情绪词	情感分类	强度	情绪词	极性	词	程度词	程度
脏乱	恶	7	强悍	1	不	一点儿	微弱
贼眼	恶	5	强势	1	非	有点	微弱
晴朗	好	5	巧夺天工	1	无	更	更
倍儿棒	好	7	巧妙	1	未	越来越	更
少见多怪	恶	3	俏丽	1	不曾	实在	强
哀莫大于心死	哀	9	俏皮	1	没	很	强
标兵	好	5	煽动	-1	没有	极度	极强
倒霉	哀	3	琐碎	-1	请勿	奇	极强
败家子	恶	7	疑心	-1	不用	微	不足
阴沉	怒	5	疑虑	-1	无须	相对	不足

2.3 网络情绪测量

为获得原始微博文本,本研究使用网络爬虫技术抓取2014年1月1日至2016年4月1日34 466名微博用户共计4 934 549条新浪微博条目。根据上文中网络社交平台情绪的界定,本研究没有选择特定目标用户群(如经常参与股票市场相关话题讨论的用户群体)作为抓取范围,而是进行随机抓取。考虑到新

浪微博庞大的活跃用户基数,要实施全样本数据抓取几乎无法实现,本研究以微博用户之间相互“关注”关系为线索,使用深度遍历的检索模式实现对样本的随机抽样。韩运荣等^[32]研究发现,微博舆论传递呈漩涡蔓延型、星系扩散型和节外生枝型,表明用户之间存在紧密联系;尹书华^[33]的研究表明,新浪微博用户关系网络具有小世界特性。所谓小世界特

性是指在新浪微博用户关系网络中,任意两个节点(用户)之间的平均路径长度较短,具体为3.314,即平均约每3个节点就可以将任意两个用户关联起来。基于此,本研究采取上述数据抓取规则并认为此规则实现了随机抽样。

在进行数据抓取时,为保证情绪数据可靠性和有用性,只抓取符合以下筛选条件的微博用户的微博数据:①非认证用户,即非名人、官方、自媒体等带有“V”字标记的微博用户。此类用户关注者数量巨大,本身多偏娱乐或官方,通常不含有效情绪信息。②粉丝数大于20且小于1500的用户,用于剔除广告营销账号和非活跃用户。广告营销类微博账户通常拥有1500以上的关注者(粉丝),个人用户关注者数目一般不超过1500;非活跃用户包含两类,一类为虽注册微博账号但不经常使用的用户,此类用户通常不会使用微博表达其个人情绪故剔除,另一类为所谓“僵尸粉”。

在对抓取到的微博文本内容进行数据清洗时,主要剔除以下内容:①转发及评论内容,为保证情绪提取的准确性,此处仅保留代表微博发布者主观观点的微博内容。②用户昵称,包括本人用户昵称及好友用户昵称。③话题内容,部分微博围绕某一热门话题时会加上“#话题名称#”,此处剔除。④定位信息,客户端信息等不包含情绪的无关信息。完成上述数据清洗工作以后,用于下一步情绪提取的数据仅含微博发布者的原创微博,此类微博内容由发布者编辑并发布,其中包含更多且更主观的用户情绪信息,无论是情绪提取的难易程度或是情绪提取的准确性都有显著提升。

在计算网络情绪得分时,主要分为3个步骤。①使用网络情绪分析中所述方法,对经过清洗后的微

博文本进行情绪提取,得到每条微博的情绪向量;②对同一天内所有微博的情绪向量各维度数值求均值,得到当天的情绪向量;③将某一时间段内所有情绪向量按日期排列获得该时间段内的情绪时间序列。图2给出本研究提取到的2014年1月1日至2016年4月1日微博情绪时间序列。

3 计量模型和方法

为研究网络情绪是否影响股市收益,本研究使用Granger因果检验方法,探讨其对股市收益的预测能力。

3.1 均值Granger因果检验

均值Granger因果检验主要考察解释变量X是否有利于预测Y的行为。一般地,如果增加X及其滞后值有助于改善对Y条件均值的预测能力,则认为X是Y的均值Granger原因,记为 $X \xrightarrow{E} Y$;反之,则认为X不是Y的均值Granger原因,记为 $X \not\xrightarrow{E} Y$ 。将 $X \not\xrightarrow{E} Y$ 作为原假设 H_0 ,检验(1)式是否成立,即

$$E[Y_t | \Omega_{t-1}(Y, X)] = E[Y_t | \Omega_{t-1}(Y)] \tag{1}$$

其中, $E(\cdot)$ 为条件期望, t 为时间, $\Omega_{t-1}(\cdot)$ 为直到 $(t-1)$ 时刻的信息集。

为检验网络情绪波动S是否为股市收益R的均值Granger原因,考虑原假设 $H_0: S \not\xrightarrow{E} R$ 或 $H_0: \beta_1 = \beta_2 = \dots = \beta_q = 0$,需要建立有约束($\beta_1 = \beta_2 = \dots = \beta_q = 0$)和无约束两个均值回归模型,即

$$E(R_t) = a + \sum_{i=1}^p \alpha_i R_{t-i} \tag{2}$$

$$E(R_t) = a + \sum_{i=1}^p \alpha_i R_{t-i} + \sum_{j=1}^q \beta_j S_{t-j} \tag{3}$$

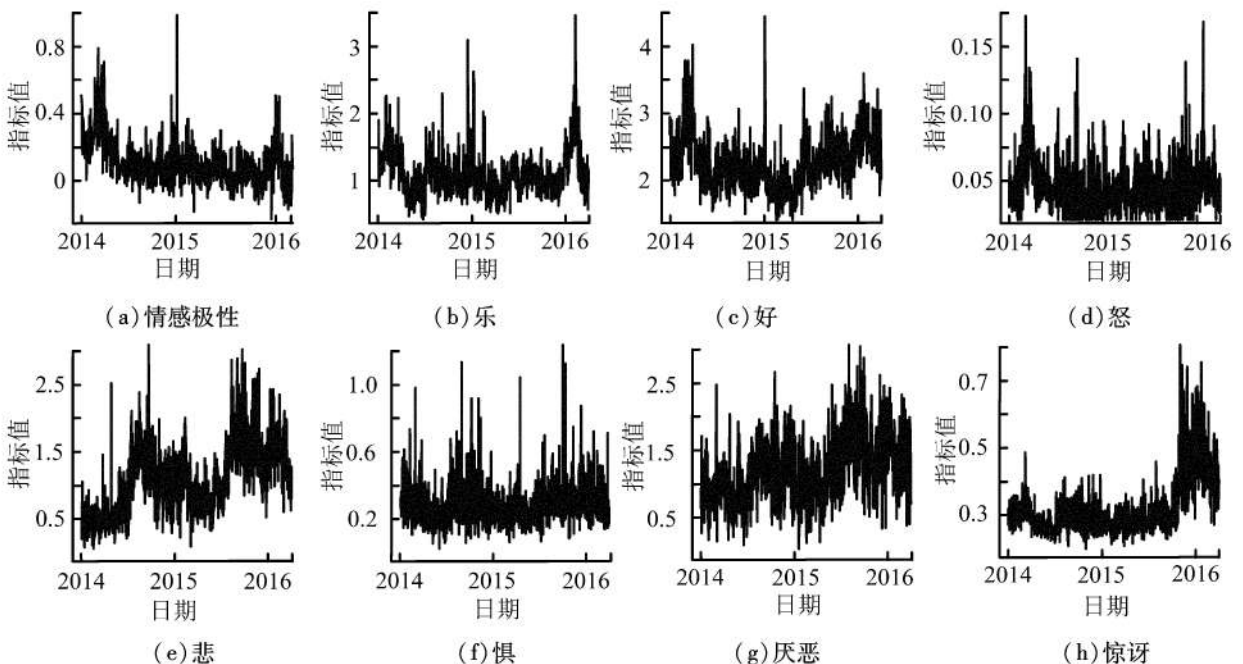


图2 微博情绪时间序列
Figure 2 Time Series of Weibo Sentiment

其中, R_t 为 t 时刻股票收益率, i 为收益率的滞后阶, p 为收益率的最大滞后阶数, j 为网络情绪指标的滞后阶, q 为网络情绪指标的最大滞后阶数, S_{t-j} 为 $(t-j)$ 时刻网络情绪波动指标值, a 为截距项。(2) 式为有约束均值回归模型, 表明 S 对 R 无显著影响; (3) 式为无约束均值回归模型, 表明 S 对 R 有显著影响。通过两个均值回归模型, 得到无约束模型残差平方和 RSS_U 和有约束模型残差平方和 RSS_R , 进而构建 F 检验统计量, 即

$$F = \frac{RSS_R - RSS_U}{\frac{RSS_U}{n-k}} \sim F(m, n-k) \quad (4)$$

其中, m 为滞后项的个数, n 为观测值个数, k 为无约束回归中待估计参数的个数。如果 F 统计值的尾概率小于 0.05, 则在 5% 显著性水平下拒绝原假设 H_0 , 认为网络情绪是股市收益的均值 Granger 原因, 意味着可以使用网络情绪预测股市收益均值的变动规律。

3.2 分位数 Granger 因果检验

均值 Granger 因果检验只能探测解释变量对响应变量条件均值的预测能力, 无法细致刻画其对响应变量在其他位置 (如上尾或下尾) 的预测能力。分位数 Granger 因果检验可以弥补这一缺憾, 它将分位数回归与均值 Granger 因果检验相结合, 能够找到变量间因果关系成立的具体位置 (分位数区间), 从而可能捕获均值 Granger 因果检验无法发现的因果关系。关于分位数回归的基本原理可以参见 KOENKER et al.^[34-35] 的研究, 关于分位数 Granger 因果检验的具体原理和方法参见 CHUANG et al.^[36]、GEBKA et al.^[37] 和 LEE et al.^[38] 的研究。

分位数 Granger 因果检验旨在检验原假设 $H_0: X \not\rightarrow Y$, 检验 (5) 式是否成立, 即

$$Q_{Y_t}[\tau | (Y, X)_{t-1}] = Q_{Y_t}(\tau | Y_{t-1}), \quad \forall \tau \in (b, c) \quad (5)$$

其中, $Q_{Y_t}(\tau | \cdot)$ 为条件 τ 分位数; τ 为分位数; (b, c) 为特定的分位数区间, 且 $0 < b < c < 1$ 。

为检验变量 X 是否为变量 Y 的分位数 Granger 原因, 建立原假设 $H_0: X \not\rightarrow Y$, 建立分位数回归模型, 即

$$Q_{Y_t}(\tau | Z_{t-1}) = a(\tau) + Y'_{t-1,p} \alpha(\tau) + X'_{t-1,q} \beta(\tau) = Z'_{t-1} \theta(\tau) \quad (6)$$

其中, $a(\tau)$ 为截距项, $\alpha(\tau)$ 和 $\beta(\tau)$ 为回归系数列向量; $\theta(\tau)$ 为回归系数向量, $\theta(\tau) = [a(\tau), \alpha(\tau)', \beta(\tau)']'$; $Y'_{t-1,p} = (Y_{t-1}, \dots, Y_{t-p})$; $X'_{t-1,q} = (X_{t-1}, \dots, X_{t-q})$; $Z'_{t-1} = (Y'_{t-1,p}, X'_{t-1,q})$ 。

由 (6) 式的分位数回归可以推导构建确定分位数 τ 处的 Wald 检验统计量为

$$W_{T(\tau)} = \frac{T \hat{\beta}'_{T(\tau)} (\Psi \hat{M}_{\tau}^{-1} \Psi')^{-1} \hat{\beta}_{T(\tau)} \hat{f}^2}{\tau(1-\tau)} \quad (7)$$

其中, T 为观测数据数; $\hat{\beta}'_{T(\tau)}$ 为 $\hat{\beta}_{T(\tau)}$ 的转置; $\hat{\beta}_{T(\tau)}$ 为 $\beta(\tau)$ 的估计值; Ψ 为选择矩阵, $\Psi \theta(\tau) = \beta(\tau)$; $\hat{M}_{\tau} =$

$\frac{1}{T} \sum_{i=1}^T Z_{i-1} Z'_{i-1}; (\cdot)^{-1}$ 为矩阵的逆; \hat{f} 为概率密度函数。

更进一步, 可以得到用于检验特定分位数区间内分位数 Granger 因果检验结果显著性的 Sup-Wald 检验统计量, 它为分位数区间 (b, c) 内若干分位数对应 Wald 检验统计量的上确界, 即

$$\sup W_T = \sup_{i=1, \dots, n} W_{T(\tau_i)} \quad (8)$$

其中, τ_i 为第 i 个分位数。根据 $\sup W_T$ 检验统计量值及相应临界值可以判断对应分位数区间内分位数 Granger 因果是否成立, 进而判断因果关系成立的具体分位数区间, 如若检验统计量表明在分位数区间 $[0.05, 0.95]$ 内非 Granger 原因不成立, 但 $[b, c]$ ($b > 0.05$ 且 $c < 0.95$) 内非 Granger 原因成立, 则意味着 Granger 因果成立的分位数区间为 $[0.05, b] \cup [c, 0.95]$ 。进一步, 可以证明

$$W_{T(\tau)} \Rightarrow \left\| \frac{B_s(\tau)}{\sqrt{\tau(1-\tau)}} \right\|^2, \quad \tau \in (b, c) \quad (9)$$

故有

$$\sup_{\tau \in (a,b)} W_{T(\tau)} \xrightarrow{D} \sup_{\tau \in (a,b)} \left\| \frac{B_s(\tau)}{\sqrt{\tau(1-\tau)}} \right\|^2 \quad (10)$$

其中, $B_s(\tau)$ 为 s 维布朗桥向量, $\|\cdot\|$ 为欧几里得范数, \Rightarrow 为弱收敛, \xrightarrow{D} 为依分布收敛。由此分布, 可以通过模拟计算的方式得到 Sup-Wald 检验统计量的临界值。

实证中, 首先, 在 $[0.05, 0.95]$ 区间内等距取 1 000 个分位数, 进行如下的分位数回归, 即

$$Q_{Y_t}[\tau | (R'_{t-1,p}, S'_{t-1,q})] = a(\tau) + R'_{t-1,p} \alpha(\tau) + S'_{t-1,q} \beta(\tau) \quad \tau \in (b, c) \quad (11)$$

其次, 根据分位数回归结果, 由 (7) 式和 (10) 式得到 Sup-Wald 检验统计量的值, 进而筛选出在 $[0.05, 0.95]$ 分位数区间内与股市收益有显著因果关系的情绪时间序列, 并确定因果关系成立的分位数区间。Sup-Wald 统计量在区间 (b, c) 上的临界值可以通过 Monte Carlo 模拟方式得到, 部分临界值已由 ANDREWS^[39] 和 KOENKER et al.^[40] 给出, 表 2 给出分位数区间 $[0.05, 0.95]$ 的 Sup-Wald 检验临界值。

表 2 分位数区间 [0.05, 0.95] 的 Sup-Wald 检验临界值
Table 2 Critical Values of the Sup-Wald Test within Quantile Interval of [0.05, 0.95]

置信水平	待检验参数向量维度				
	1	2	3	4	5
10%	8.19	11.20	13.47	15.62	17.69
5%	9.84	12.93	15.15	17.56	19.61
1%	13.01	16.44	19.06	21.54	24.18

注: 待检验参数向量维度即本研究实证中的滞后天数; 数据来源: ANDREWS D W K. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 1993, 61(4): 821-856.

4 实证研究

分别使用均值 Granger 因果检验和分位数 Granger 因果检验,探讨网络情绪对股市收益的影响,并将两种方法取得的结果进行对比。

4.1 数据选取

为探讨网络情绪与股市收益之间的因果关系,除网络情绪日数据外,还需要股市收益数据。考虑到沪深300指数的代表性,本研究使用其作为整个沪深股市收益的代表,样本选取2014年1月1日至2016年4月1日期间的沪深300指数日收盘价,数据来源于Wind数据库。表3给出样本期间沪深300指数对数收益率和网络情绪描述性统计结果,情感极性可以取负值(表示消极情绪),其他情绪值大于或等于0,偏度和峰度的标准值均为0。根据沪深300指数对数收益率的偏度和峰度可以看出其具有明显的尖峰厚尾特性,表明中国证券市场不服从正态分布,适合使用分位数回归方法加以分析。

为探讨不同股市发展阶段下网络情绪对股票市场影响的差异,本研究将股票市场划分为3个阶段。①股市波动阶段,即在此期间股指有正常的上下波动但整体上基本持平;②股市上升阶段,即在此期间股指总体走势呈明显上升趋势;③股市下降阶段,即在此期间股指总体走势呈明显下降趋势。

根据各走势阶段的定义,结合图3的沪深300指数价格波动情况,对股票市场进行阶段划分,详见表4。整个样本期间内,股市波动阶段共有3个,考虑到2015年1月5日至2015年3月1日和2015年8月25日至2015年12月21日两区间为股市大幅波动前后的盘整阶段,此时股市虽整体表现为上下波动但不能视为“正常”,此处不适用;股市下降阶段共有3个,但2015年8月17日至2015年8月25日股市下降时间区间过短(去除休息日及滞后项数据后仅4条数据),此处不适用。剔除上述3个时间区间后,本研究选取

2014年1月1日至2014年10月26日作为股市波动时间段(S1),选取2014年10月31日至2015年1月5日和2015年3月5日至2015年6月2日两个时间段分别作为股市上升阶段1(S2)和股市上升阶段2(S3),选取2015年6月7日至2015年7月8日和2015年12月21日至2016年1月27日两个时间段分别作为股市下降阶段1(S4)和股市下降阶段2(S5)。

在使用Granger因果检验探讨网络情绪与股票市场之间关系时,要求数据对象必须是平稳的时间序列。为此,在股指价格和网络情绪指数基础上,通过一阶差分,分别得到股指收益和网络情绪波动为

$$R_t = \ln(P_t) - \ln(P_{t-1}) \tag{12}$$

$$S_t = M_t - M_{t-1} \tag{13}$$

其中, P_t 为 t 时刻沪深300收盘价, M_t 为 t 时刻网络情绪指数。

4.2 结果分析

分别使用均值 Granger 因果检验和分位数 Granger 因果检验,探讨网络情绪波动对股市收益的影响。考虑到股市每周有5个交易日,本研究分别使用滞后1阶到滞后5阶,进行Granger因果检验。

在进行Granger因果检验之前,为避免出现伪回归,首先对所有8个情感时间序列和一个股市收益率时间序列进行单位根检验。为此,选取PP检验和KPSS检验两个非参数检验方法,前者以“时间序列非平稳”作为原假设,容易犯“存伪错误”;后者则把“时间序列平稳”作为原假设,可以视为PP检验的补充。同时使用两种检验方法,可以降低“存伪错误”的可能,使检验更全面、更具说服力。表5给出PP检验和KPSS检验的尾概率,两种检验方法一致表明,在1%显著性水平下都拒绝原假设,认为不存在单位根,意味着9个时间序列都平稳,可以进行Granger因果检验。

表3 各变量描述性统计结果
Table 3 Results of Description Statistics of Variables

变量名称	符号	最小值	最大值	均值	中位数	方差	标准差	偏度	峰度
沪深300指数对数收益率	CSI300	-9.15	6.50	0.03	0.04	3.02	1.74	-0.79	4.43
情感极性	POL	-0.21	1.20	0.19	0.16	0.03	0.16	1.69	5.19
乐	HAPPY	0.63	3.47	1.33	1.29	0.11	0.34	1.50	5.27
好	GOOD	1.30	4.87	2.34	2.27	0.25	0.50	1.01	2.06
怒	ANGER	0.003	0.21	0.04	0.03	0.004	0.03	1.93	6.11
悲	SAD	0.27	2.01	0.86	0.83	0.10	0.31	0.65	0.29
惧	FEAR	0.04	0.94	0.26	0.24	0.01	0.11	1.68	5.89
厌恶	HATE	0.49	2.06	1.09	1.07	0.07	0.01	0.52	2.00
惊讶	AMAZE	0.003	0.60	0.12	0.10	0.26	0.09	0.21	4.78

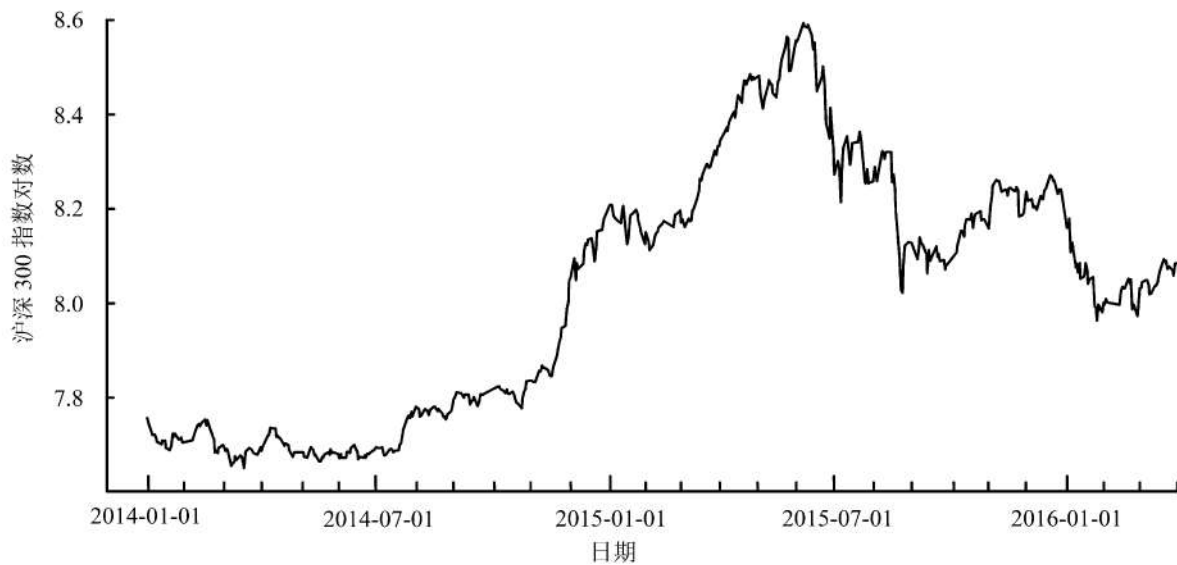


图3 2014年1月1日至2016年4月1日沪深300指数价格
Figure 3 Price of CSI 300 from 2014-01-01 to 2016-04-01

表4 股市走势阶段划分
Table 4 Partitioned Stages of Stock Market Trend

股市阶段	时间区间	是否选择	划分结果
股市波动	2014-01-01 ~ 2014-10-26	是	股市波动阶段:S1
	2015-01-05 ~ 2015-03-01	否	
	2015-08-25 ~ 2015-12-21	否	
股市上升	2014-10-31 ~ 2015-01-05	是	股市上升阶段 1:S2
	2015-03-05 ~ 2015-06-02	是	股市上升阶段 2:S3
股市下降	2015-06-07 ~ 2015-07-08	是	股市下降阶段 1:S4
	2015-08-17 ~ 2015-08-25	否	
	2015-12-21 ~ 2016-01-27	是	股市下降阶段 2:S5

表5 单位根检验结果
Table 5 Results of Unit Root Tests

检验方法	情感极性	乐	好	怒	悲	惧	厌恶	惊讶
PP 检验	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
KPSS 检验	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

4.2.1 均值Granger因果检验

在股市不同阶段,对各网络情绪波动与股市收益之间关系进行均值Granger因果检验,结果见表6。表6给出的滞后阶数为最优滞后阶数,由SIC准则从滞后1天~滞后5天中确定。由表6可得到如下结论。

(1)在股市上升阶段网络情绪波动与股市收益之间存在更多因果关系

各网络情绪波动与股市收益之间的因果关系在股市波动阶段(S1)和股市下降阶段(S4和S5)整体不显著。不过,在股指两个上升阶段(S2和S3),网络情绪波动与股市收益之间出现了比股市波动阶段和股

表6 不同股市走势阶段下均值Granger因果检验结果
Table 6 Results of Mean Granger Causality Tests under Different Stages of Stock Market Trend

阶段	模型信息	情感极性	乐	好	怒	悲	惧	厌恶	惊讶
S1	滞后阶/天	4	5	3	4	2	2	2	5
	SIC	546.44	548.38	545.56	545.95	546.21	547.85	545.84	545.00
	p 值	0.27	0.51	0.19	0.22	0.21	0.48	0.18	0.18
S2	滞后阶/天	5	5	5	5	5	5	4	4
	SIC	180.54	173.74	177.37	174.42	176.38	178.44	169.07	174.41
	p 值	1.00	0.29	0.69	0.32	0.52	0.78	0.02*	0.54
S3	滞后阶/天	5	5	4	4	5	5	4	5
	SIC	239.74	248.44	236.85	239.42	243.38	248.82	243.66	248.26
	p 值	0.06	0.76	0.04*	0.02*	0.21	0.81	0.23	0.74
S4	滞后阶/天	4	5	4	5	4	4	5	5
	SIC	104.40	105.12	109.85	111.63	107.31	108.40	101.18	109.32
	p 值	0.28	0.12	0.93	0.96	0.76	0.86	0.39	0.95
S5	滞后阶/天	5	5	5	5	5	5	4	5
	SIC	179.08	179.48	176.76	180.99	181.39	174.06	179.99	177.36
	p 值	0.30	0.31	0.29	0.46	0.67	0.09	0.97	0.82

注:p 值为均值Granger因果检验中的Wald检验p值;*为在5%水平下显著,下同。

市下降阶段更为广泛的因果关系,意味着当股票市场处于上升时期时更容易受到网络情绪的影响。究其原因,可能由于在股票市场整体向上的阶段,投资者心理压力减轻,对股市投资的谨慎程度降低,从而导致其决策的随意性、非理性程度比平时要强,其投资决策容易受各类情绪的影响。

(2)不同股市走势阶段股市收益受网络情绪波动影响差异明显

这种差异主要体现在与股市收益有显著因果关系的情绪种类上。例如,同为股市上升阶段,但上升阶段1和上升阶段2中能够作为股市收益Granger原因的网络情绪种类并不一致,在股市上升阶段1中主要为厌恶,在股市上升阶段2中主要为好和怒。这可能是由不同波动阶段的具体特性造成,经过盘整之后继续上涨的股市上升阶段2相较于股市上升阶段1投资者心理发生变化,能够对其决策产生显著影响的情绪维度有所不同,这一结果反映了两个阶段内市场中的投资者情绪状态有所不同。

4.2.2 分位数Granger因果检验

在表6的40个均值Granger因果检验中,存在3个显著的Granger因果关系,占比仅为7.50%,似乎说明

网络情绪并不能显著影响且准确预测股市收益。这一结果是否可靠,本研究使用分位数Granger因果检验,探讨网络情绪波动与股市收益之间的关系,结果见表7。表7给出的滞后阶数为最优滞后阶数,也是根据SIC准则从滞后1天~滞后5天中确定。在表7的40个分位数Granger因果检验中,发现存在23个显著的Granger因果关系,占比上升到57.50%。由表7可以进一步得到如下主要结论。

(1)使用分位数Granger因果检验能够揭示均值Granger因果检验难以发现的因果关系

无论股市处于哪个阶段,在0.05~0.95分位点之间表现出与股市收益之间存在显著因果关系的网络情绪波动数量都多于均值Granger因果检验。这一结果表明,一些在特定分位点区间真实存在的网络情绪与股市收益之间的因果关系,使用均值Granger因果检验方法往往难以发现,而分位数Granger因果检验则可以弥补这一缺憾,能够准确地捕捉到在不同分位点区间的因果关系。例如,在波动阶段(S1),均值Granger因果检验认为情感极性波动不是股市收益的Granger原因,而分位数Granger因果检验却发现,在[0.92,0.95]分位区间上,情感极性波动是股市收益

表7 不同股市走势阶段下分位数Granger因果检验

Table 7 Results of Quantile Granger Causality Tests under Different Stages of Stock Market Trend

阶段	模型信息	情感极性	乐	好	怒	悲	惧	厌恶	惊讶
S1	滞后阶/天	3	4	4	3	5	3	4	3
	SIC	725.03	576.67	712.29	631.38	638.80	709.20	710.43	725.03
	Sup-Wald	28.40**	14.79	7.56	9.64	9.42	33.83**	24.47**	16.64*
	分位区间	[0.92,0.95]					[0.91,0.95]	[0.92,0.95]	[0.94,0.95]
S2	滞后阶/天	4	3	4	5	2	3	4	4
	SIC	196.93	204.13	211.62	185.59	199.71	192.23	182.48	192.01
	Sup-Wald	20.31*	13.36	26.60**	25.40**	21.29*	4.59	47.88**	9.69
	分位区间	[0.05,0.16]		[0.05,0.10]	[0.05,0.19]	[0.05,0.10]		[0.05,0.13]	
S3	滞后阶/天	5	4	5	4	4	3	4	4
	SIC	282.25	295.37	279.91	286.58	304.35	308.48	290.16	313.61
	Sup-Wald	33.00**	12.79	21.69*	28.43**	19.14*	24.98**	25.20**	23.45**
	分位区间	[0.05,0.09]		[0.05,0.07]	[0.05,0.12]	[0.05,0.08]	[0.05,0.09]	[0.05,0.13]	[0.05,0.08]
S4	滞后阶/天	4	5	4	4	5	3	4	3
	SIC	106.67	106.41	116.08	104.07	94.72	113.27	110.80	109.57
	Sup-Wald	19.58*	12.64	2.84	5.73	4.69	3.78	8.21	16.49*
	分位区间	[0.05,0.07]							[0.05,0.01]
S5	滞后阶/天	2	5	4	3	4	4	3	4
	SIC	210.33	194.39	192.19	188.41	185.95	179.06	217.98	193.68
	Sup-Wald	13.23*	20.26*	5.74	17.50*	9.39	24.11**	19.60**	18.14
	分位区间	[0.93,0.95]	[0.92,0.95]		[0.92,0.95]		[0.93,0.95]	[0.92,0.95]	

注: Sup-Wald为分位数Granger因果检验中的Sup-Wald检验统计值;**为在1%水平下显著;*为在5%水平下显著;分位区间为分位数Granger因果关系显著成立的分位点所在区间,没有数据表示不存在分位区间使分位数Granger因果成立。

的Granger原因。此外,在S1阶段,均值Granger因果检验认为不存在对股市收益存在显著因果影响的情绪,而分位数Granger因果检验却发现情感极性、惧、厌恶、惊讶等4个情绪波动对股市收益存在显著的因果关系。在5个不同阶段,均值Granger因果检验认为情感极性都不是股市收益的Granger原因,而分位数Granger因果检验却发现情感极性一直都是股市收益的Granger原因。

这一结果可能存在两个方面的原因。①就方法论而言,基于分位数回归的分位数Granger因果检验能够揭示解释变量对响应变量在不同分位点处的异质影响,网络情绪的影响可能表现为部分分位点区间为正向且显著影响、部分分位点区间为负且显著影响、部分分位点区间无显著影响等情形,从而出现一些特定分位点区间存在网络情绪与股市收益之间

的因果关系。这一特定分位点区间的Granger因果关系很难被基于均值回归的均值Granger因果检验发现,因为部分分位点区间(如上尾部)的正向且显著影响与部分分位点区间(如下尾部)的负向且显著影响可能互相抵消,导致均值回归中网络情绪对股市收益的均值影响不显著,得出不存在Granger因果关系的结论。②就实际问题而言,行为金融理论认为,市场上存在众多非理性投资行为,投资者情绪反映了市场上的投机倾向,在一些极端收益目标追求下,投资者情绪往往更容易与极端收益建立起关联,从而表现为在一些极端分位点区间上形成对股市收益的Granger因果影响。

(2)不同市场走势阶段下股市收益受网络情绪影响具有异质性

从具有显著因果关系的具体分位点区间看,当

市场处于波动阶段(S1)期间,所有因果关系均在高分位点(靠近0.95)处成立;当市场处于股市上升阶段(S2和S3)期间,所有因果关系均在低分位点(靠近0.05)处成立。不过,同为股市下降阶段S4和S5,其具有因果关系的分位点区间却截然不同。在S4阶段,情感极性在分位区间 $[0.05, 0.07]$ 上构成股市收益的Granger原因;在S5阶段,其构成Granger原因的分位区间为 $[0.93, 0.95]$ 。这可能源于两个下降阶段各自的特殊性造成心理状态的差异性。股市下降阶段1为本次股指大幅涨落期间的首次大幅下跌,且下跌速度快,下跌程度剧烈,下跌之前为股指疯涨阶段,在下跌局面发生之前多数投资者对此次下跌没有预期和心理准备;而股市下降阶段2为本期期间股指的第3次大跌,下跌相对缓和且大跌之前市场处于盘整阶段,下跌发生之前投资者已有一定程度的心理准备。以上结果表明,当市场处于不同走势阶段时情绪对股市收益的影响具有异质性,其异质性不仅表现在影响收益的情绪种类上,还表现在所影响收益率的分位区间上。

更进一步,本研究认为造成这种异质性的根本原因在于,不同心理状态下投资者情绪受网络情绪影响效果的异质性。当市场处于波动状态时,市场上的意外情况较少,投资者处于放松的心理状态,此时投资者受网络情绪影响较多且情绪对市场收益率的影响发生在高分位区间;当市场处于上升阶段时,投资者处于乐观轻松的心理状态,此时市场受情绪影响较多,且此时情绪的影响发生在低分位区间;当市场处于下降阶段尤其是突发、剧烈的下降阶段时,市场整体不乐观且意外情况频发,投资者处于紧张的心理状态,此时投资者受网络情绪的影响极少且均发生在低分位区间。

(3)分位数 Granger 因果检验结果包含了均值 Granger 因果检验结果

首先,比较不同股市走势阶段下的分位数 Granger 因果检验结果,发现在股市上升阶段与股市收益之间存在显著因果关系的情绪总体上要多于股市下降阶段,此结果与均值 Granger 因果检验一致,表明在股市上升阶段股市收益受情绪的影响更大。其次,不同股市走势阶段下,股市收益受网络情绪影响差异明显,其中股市上升阶段存在显著因果关系的情绪种类较多,如在上升阶段2(S3)有情感极性、好、怒、悲、惧、厌恶、惊讶等7种情绪波动与股市收益之间存在显著的因果关系,比上升阶段1(S2)多了惧和惊讶。本研究认为,这是因为在经过盘整之后继续上涨的股指上升阶段2(S3),投资者的心理状态比股指上升阶段1(S2)期间担忧的成分有所增加,更容易受到诸如惧和惊讶等情绪的影响。而在处于股市下降阶段时,网络情绪与股市收益之间的因果关系总体偏少,如在股市下降阶段1(S4)时,仅有情感极性和惊讶具有显著的因果关系。因此,从情绪种类角度看,股票市场在不同市场走势阶段下受情绪影响是十分敏感的。

5 结论

本研究旨在揭示网络情绪对股市收益的影响,一方面给出网络情绪的分析方法和测量结果,另一方面综合使用均值 Granger 因果和分位数 Granger 因果检验方法进行实证检验。选取新浪微博数据和沪深300指数为研究对象,实证检验网络情绪波动对股市收益的影响。研究表明,在一些极端分位点区间两者存在显著且广泛的 Granger 因果关系,这在特定条件下股市收益的可预测性提供了佐证。①使用分位数 Granger 因果检验,能够揭示均值 Granger 因果检验难以发现的因果关系,前者的发现率为57.50%,远高于后者的7.50%。②在一些特定分位点区间存在网络情绪波动与股市收益之间的因果关系,表明蕴藏于互联网络中的情绪信息有助于增加对股市收益的理解,可以将其纳入经济计量模型之中,与宏观或(和)微观经济变量一起研究股市收益行为,能够提高模型的解释能力和预测效果。③网络情绪波动对股市收益的影响表现出随着分位点变化而变化的异质效应,需要深入考虑并揭示在不同分位点处或不同分位点区间上网络情绪的作用和贡献,如好和悲只在低分位点区间上成为股市收益的 Granger 原因。④不同市场走势阶段下网络情绪波动与股市收益之间的 Granger 因果关系有着不同的表现,主要体现在 Granger 因果关系成立的分位点区间和情绪种类的差异上,当市场处于波动阶段(S1),所有因果关系均在高分位点(靠近0.95)处成立;当市场处于股市上升(S2和S3)期间,所有因果关系均在低分位点(靠近0.05)处成立;而同为股市下降阶段的S4和S5,其具有因果关系的分位点区间却截然不同。

本研究意义在于两个方面。①为如何有效挖掘并使用微博信息提供一种新的可行思路。鉴于中国网络用户群体基数大、网络普及范围广泛的现状,在海量的互联网络信息中必然蕴藏着大量能够揭示中国经济、政治等方面当前状态及民情、民意、民生等关乎国计民生的有用信息,因而有效挖掘并使用这些信息能够有助于企业制定合理的发展规划和竞争战略,有助于政府进行科学决策和制度设计。②为在特定条件下股市收益的可预测性提供了佐证。实证发现,在一些特定分位点区间网络情绪与股市收益之间存在显著的因果关系,意味着网络情绪对股市收益的条件分位数具有一定的预测能力,可以利用网络情绪变化预测股市收益的尾部行为特征,为金融风险防范提供决策参考。需要注意的是,本研究的分位数 Granger 因果关系大多发生在极端分位点区间,如上尾的 $[0.90, 0.95]$ 和下尾的 $[0.05, 0.10]$,表明网络情绪对股市收益的预测能力有较强的限制条件。

本研究使用新的计量方法得到一些有意义的发现,但受客观条件限制,研究工作尚存在不足之处。①网络情绪对股票市场的影响主要通过影响投资者情绪来完成,其背后的作用机理需要进一步明确,可

以开展进一步的理论研究工作来探索网络情绪对股市波动的影响渠道、影响方式和影响程度等。②目前中文自然语言处理准确率尚有所不足,从文本中提取情绪信息还存在一定程度的误差,进一步研究工作可以通过人工标注大量语料,配合机器学习方法(如神经网络方法等)提高情绪提取准确率。③本研究主要针对中国股市开展,得到的研究结果是否适合发达国家或者成熟市场国家的股市值得进一步探讨。

参考文献:

- [1] NOFSINGER J R. Social mood and financial economics. *Journal of Behavioral Finance*, 2005, 6(3):144-160.
- [2] PRECHTER, Jr R R. *The wave principle of human social behavior and the new science of socioeconomics*. Gainesville, Ga: New Classics Library, 1999:241-254.
- [3] PRECHTER, Jr R R, PARKER W D. The financial/economic dichotomy in social behavioral dynamics: the socioeconomic perspective. *Journal of Behavioral Finance*, 2007, 8(2):84-108.
- [4] YAO J, MA C, HE W P. Investor herding behaviour of Chinese stock market. *International Review of Economics & Finance*, 2014, 29:12-29.
- [5] HUANG D, JIANG F, TU J, et al. Investor sentiment aligned: a powerful predictor of stock returns. *The Review of Financial Studies*, 2015, 28(3):791-837.
- [6] KIM S H, KIM D. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, 2014, 107(2):708-729.
- [7] JOSEPH K, WINTOKI M B, ZHANG Z. Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *International Journal of Forecasting*, 2011, 27(4):1116-1127.
- [8] 张谊浩,李元,苏中锋,等.网络搜索能预测股票市场吗?.*金融研究*,2014(2):193-206.
ZHANG Yihao, LI Yuan, SU Zhongfeng, et al. Can internet search predict the stock market?. *Journal of Financial Research*, 2014(2):193-206. (in Chinese)
- [9] 洪涛,厉伟.基于网络搜索数据的住房价格预期与实际价格波动分析.*统计与信息论坛*,2015,30(11):49-53.
HONG Tao, LI Wei. A study on the relationship between expectation and real housing prices based on query index. *Statistics & Information Forum*, 2015, 30(11):49-53. (in Chinese)
- [10] GILBERT E, KARAHALIOS K. Widespread worry and the stock market // *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Washington, DC: AAAI Press, 2010:58-65.
- [11] GRUHL D, GUHA R, KUMAR R, et al. The predictive power of online chatter // *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*. Chicago, Illinois: ACM Press, 2005:78-87.
- [12] MISHNE G, DE RIJKE M. Capturing global mood levels using blog posts // *Proceedings of the AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*. California: AAAI Press, 2006:145-152.
- [13] LIU Y, HUANG X, AN A, et al. ARSA: a sentiment-aware model for predicting sales performance using blogs // *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Amsterdam, Netherlands: ACM Press, 2007:607-614.
- [14] PENG C L, LAI K L, CHEN M L, et al. Investor sentiment, customer satisfaction and stock returns. *European Journal of Marketing*, 2015, 49(5/6):827-850.
- [15] ZHANG X, FUEHRES H, GLOOR P A. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences*, 2011, 26:55-62.
- [16] BOLLEN J, MAO H, ZENG X. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011, 2(1):1-8.
- [17] COREA F, CERVELLATI E M. The power of micro-blogging: how to use twitter for predicting the stock market. *Eurasian Journal of Economics and Finance*, 2015, 3(4):1-7.
- [18] WEI W, MAO Y, WANG B. Twitter volume spikes and stock options pricing. *Computer Communications*, 2016, 73(2):271-281.
- [19] 赖凯声,陈浩,钱卫宁,等.微博情绪与中国股市:基于协整分析.*系统科学与数学*,2014,34(5):565-575.
LAI Kaisheng, CHEN Hao, QIAN Weiming, et al. Weibo emotion and Chinese stock market: a study based on co-integration analysis. *Journal of Systems Science and Mathematical Sciences*, 2014, 34(5):565-575. (in Chinese)
- [20] 黄润鹏,左文明,毕凌燕.基于微博情绪信息的股票市场预测.*管理工程学报*,2015,29(1):47-52,215.
HUANG Runpeng, ZUO Wenming, BI Lingyan. Predicting the stock market based on microblog mood. *Journal of Industrial Engineering and Engineering Management*, 2015, 29(1):47-52,215. (in Chinese)
- [21] 朱南丽,邹平,张永平等.基于博客/微博信息量的投资者关注度测量研究:来自中国股票市场的经验数据.*经济问题探索*,2015(2):159-166.
ZHU Nanli, ZOU Ping, ZHANG Yongping, et al. Investor attention measurement based on the number of blogs and microblogs. *Inquiry into Economic Issues*, 2015(2):159-166. (in Chinese)
- [22] 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.*宏观经济研究*,2016(1):144-153.
MENG Xuejing, MENG Xianglan, HU Yangyang. Research on the investor sentiment index based on text mining and baidu index. *Macroeconomics*, 2016(1):144-153. (in Chinese)
- [23] 曾小梦.基于网民心理的网络舆情引导研究.长沙:湖南大学,2013:20-28.
ZENG Xiaomeng. *Research on how to guide network public opinion based on the internet psychology*. Changsha: Hunan University, 2013:20-28. (in Chinese)

- [24] ZHOU G, ZHAO J, ZENG D. Sentiment classification with graph co-regularization // *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, 2014;1331-1340.
- [25] POURSEPANJ H, WEISSBOCK J, INKPEN D. uOttawa; system description for SemEval 2013 task 2 sentiment analysis inTwitter // *Proceedings of the Second Joint Conference on Lexical and Computational Semantics: Seventh International Workshop on Semantic Evaluation*. Atlanta, Georgia, 2013;380-383.
- [26] 李婷婷, 姬东鸿. 基于SVM和CRF多特征组合的微博情感分析. *计算机应用研究*, 2015, 32(4): 978-981.
LI Tingting, JI Donghong. Sentiment analysis of micro-blog based on the featured combinations of SVM and CRF. *Application Research of Computers*, 2015, 32(4): 978-981. (in Chinese)
- [27] 苏小英, 孟环建. 基于神经网络的微博情感分析. *计算机技术与发展*, 2015, 25(12): 161-164, 168.
SU Xiaoying, MENG Huanjian. Sentiment analysis of micro-blog based on neural networks. *Computer Technology and Development*, 2015, 25(12): 161-164, 168. (in Chinese)
- [28] 刘晓菲, 丁香乾, 石硕, 等. 基于改进KNN的消费者评价信息情感分类研究. *微型机与应用*, 2014, 33(24): 81-83, 86.
LIU Xiaofei, DING Xiangqian, SHI Shuo, et al. Research on classifying sentiment of consumer evaluation based on modified KNN algorithm. *Microcomputer & Its Applications*, 2014, 33(24): 81-83, 86. (in Chinese)
- [29] YANG B, CARDIE C. Context-aware learning for sentence-level sentiment analysis with posterior regularization // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore, Maryland, 2014: 325-335.
- [30] 马秉楠, 黄永峰, 邓北星. 基于表情符的社交网络情绪词典构造. *计算机工程与设计*, 2016, 37(5): 1129-1133.
MA Bingnan, HUANG Yongfeng, DENG Beixing. Generating sentiment lexicon of online social network based on emotions. *Computer Engineering and Design*, 2016, 37(5): 1129-1133. (in Chinese)
- [31] 蒋盛益, 黄卫坚, 蔡茂丽, 等. 面向微博的社会情绪词典构建及情绪分析方法研究. *中文信息学报*, 2015, 29(6): 166-171, 202.
JIANG Shengyi, HUANG Weijian, CAI Maoli, et al. Building social emotional lexicons for emotional analysis on microblog. *Journal of Chinese Information Processing*, 2015, 29(6): 166-171, 202. (in Chinese)
- [32] 韩运荣, 高顺杰. 微博舆论传播模式探究. *现代传播: 中国传媒大学学报*, 2012, 34(7): 35-39.
HAN Yunrong, GAO Shunjie. Research on the mode of micro-blog public opinion communication. *Modern Communication: Journal of Communication University of China*, 2012, 34(7): 35-39. (in Chinese)
- [33] 尹书华. 基于复杂网络的微博用户关系网络特性研究. *西南师范大学学报: 自然科学版*, 2011, 36(6): 57-61.
YIN Shuhua. A research of user relations properties based on a complex network of microblog. *Journal of Southwest China Normal University: Natural Science Edition*, 2011, 36(6): 57-61. (in Chinese)
- [34] KOENKER R. *Quantile regression*. New York: Cambridge University Press, 2005;145-271.
- [35] KOENKER R, HALLOCK K F. Quantile regression. *Journal of Economic Perspectives*, 2001, 15(4): 143-156.
- [36] CHUANG C C, KUAN C M, LIN H Y. Causality in quantiles and dynamic stock return-volume relations. *Journal of Banking & Finance*, 2009, 33(7): 1351-1360.
- [37] GEBKA B, WOHAR M E. Causality between trading volume and returns: evidence from quantile regressions. *International Review of Economics & Finance*, 2013, 27: 144-159.
- [38] LEE T H, YANG W. Granger-causality in quantiles between financial markets: using copula approach. *International Review of Financial Analysis*, 2014, 33: 70-78.
- [39] ANDREWS D W K. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 1993, 61(4): 821-856.
- [40] KOENKER R, MACHADO J A F. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 1999, 94(448): 1296-1310.

Exploring the Relationship between Internet Sentiment and Stock Market Returns Based on Quantile Granger Causality Analysis

XU Qifa^{1,2}, BO Zhongpu¹, JIANG Cuixia¹

1 School of Management, Hefei University of Technology, Hefei 230009, China

2 Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education, Hefei University of Technology, Hefei 230009, China

Abstract: According to the behavioral finance theory, besides fundamental macro-economic factors, the change of stock market price is also influenced by individual investors or noise traders to a large extent. There are a large amount of individual investors

in Chinese stock market. Most of them are also Internet users. Due to the high degree of coincidence between the group of individual investors and Internet users, it is possible to explore the fundamental rule of stock returns in China using internet sentiment information.

To investigate the impacts of individual investor behavior, we examine the relationship between investor sentiment and stock returns. In particular, we focus on and evaluate the impacts of internet sentiment information on stock returns from the perspective of individual investor sentiment. First, we extract several Internet Sentiment time series from Sina Weibo texts using Chinese text sentiment analysis approach. Second, we explore the causal relationship between internet sentiment changes and stock market returns using both the mean and quantile Granger causality tests. Third, Chinese stock market is divided into three stages based on its trends and we study the Granger causal relationship between internet sentiment changes and stock returns in different stages.

An empirical study is conducted on the CSI 300 index. The empirical results show that there is a widespread and significant causal relationship between the two variables at extreme quantile intervals via the quantile Granger causality test, instead of the naive mean Granger causality test. Specifically, in 40 causal relationship tests, the quantile Granger causality test confirms that there are 23 significant causal relationships. The discovery ratio is 57.5%, which is obviously larger than that of 7.5% of the naive mean Granger causality test. In addition, the effects of internet sentiment changes on stock returns are heterogeneous over the different market stages.

The empirical findings have great significance both in theory and practice. A significant causal impact of internet sentiment on the stock market returns has proved to be for some specific quantile intervals. It provides evidences for the predictability of stock market returns under certain conditions. Therefore, internet sentiment can be used to predict the tail (for instance upper or lower tail) of stock return, which offers a decision-making scheme to avoid financial risk. These results eventually provide a new idea for stock market research involving asset pricing, return prediction, volatility estimation, and so on. They have also shed some lights on a new direction for the application of internet sentiment information.

Keywords: quantile regression; Granger causality test; stock market returns; internet sentiment; Weibo data

Received Date: August 20th, 2016 **Accepted Date:** March 21st, 2017

Funded Project: Supported by the National Natural Science Foundation of China(71671056), the National Social Science Foundation of China (15BJY008) and the Humanity and Social Sciences Research of Ministry of Education(14YJA790015)

Biography: XU Qifa, doctor in management, is a professor in the School of Management at Hefei University of Technology. His research interests include theoretical and methodology of quantitative economics, and financial econometrics. His representative paper titled "Composite quantile regression neural network with applications" was published in the *Expert Systems with Applications* (Volume 76, 2017). E-mail: xuqifa1975@126.com

BO Zhongpu is a master degree candidate in the School of Management at Hefei University of Technology. His research interest focuses on financial econometrics. E-mail: zp2576302465@163.com

JIANG Cuixia, doctor in management, is an associate professor in the School of Management at Hefei University of Technology. Her research interests include financial time series analysis and financial econometrics. Her representative paper titled "Hedge fund investment strategies based on LASSO quantile regression" was published in the *Journal of Management Sciences in China* (Issue 3, 2016). E-mail: jiangcx1973@163.com

□