



# 社交媒体的投资者 涨跌情绪与证券市场指数

程琬芸, 林 杰

同济大学 经济与管理学院, 上海 200092

**摘要:** 以新浪微博为数据来源, 利用中文文本分析技术, 从5个与证券相关的认证机构用户的微博和评论中提取出大多数投资者对证券市场未来走势涨跌的情绪倾向信息, 构建社交媒体的投资者涨跌情绪指数, 作为社交媒体的度量指标, 运用 Granger 因果关系检验和脉冲响应函数方法, 对社交媒体与证券市场之间的动态影响以及社交媒体对证券市场的预测能力进行分析。研究结果表明, 社交媒体的投资者涨跌情绪指数与证券市场指数收益、成交量之间均存在正相关关系, 证券市场指数的收益、成交量对社交媒体的投资者涨跌情绪指数的影响持续时间超过40个交易日; 社交媒体的投资者涨跌情绪指数对证券市场指数收益仅短期影响显著; 虽然无法帮助预测证券市场指数收益, 但是能够成为预测证券市场指数的交易成交量的有效指标。社交媒体中富含了有助于投资决策的有效信息, 实证研究结果有助于帮助投资者及时判断社交媒体中信息的有效性以及如何更好地利用社交媒体辅助他们的投资决策行为。

**关键词:** 社交媒体; 证券市场指数; 投资者涨跌情绪指数; 文本分析

**中图分类号:** F830.9

**文献标识码:** A

**doi:** 10.3969/j.issn.1672-0334.2013.05.011

**文章编号:** 1672-0334(2013)05-0111-09

## 1 引言

随着 Web 2.0 技术的广泛应用, 以互动社交为特征的社交媒体(如论坛、博客和微博等)被应用到金融领域。投资者通过社交媒体可以实时获取证券市场的相关资讯, 与他人交流对证券市场和个股的看法和感受, 并且参与到社区的各类讨论和交流中<sup>[1-3]</sup>。与此同时, 不同渠道的信息通过社交媒体对信息的分享、聚集和放大等功能在社区中迅速传播, 为投资者的投资决策行为提供了丰富的决策参考信息<sup>[4]</sup>。关于社交媒体与证券市场之间相互关系的研究逐渐成为当今热点研究课题。

目前中国证券市场还只是新兴加转型的市场, 个人投资者在市场参与者中所占比重仍然较大, 个人投资者的严重投机心理和不成熟的投资理念使投

资者很难正确地判断信息有效性并及时地做出投资决策。深入分析社交媒体与证券市场之间的相互影响, 总结社交媒体对证券市场的预测作用, 能够帮助投资者去掉投机心理、树立投资理念, 并有利于深入开展投资者教育工作。

## 2 相关研究评述

在社交媒体被广泛应用的时代, 投资者不再是被动地接收信息, 而是主动地使用网络搜索和发布信息。社交媒体逐渐成为投资者重要的信息来源, 开始蕴涵能够辅助投资者进行投资决策的有效信息。Mao 等<sup>[5]</sup>对传统的投资者情绪指标 Investor Intelligence 和 Daily Sentiment Index 与多个社交媒体(如在线新闻、搜索引擎和微博)的预测证券市场能力进行

收稿日期: 2013-02-22 修返日期: 2013-09-24

基金项目: 国家自然科学基金(71071114); 教育部人文社会科学研究项目(11YJC630216); 上海市重点学科建设项目(B310); 广东省自然科学基金(S2012010010794)

作者简介: 程琬芸(1981-), 女, 广东南海人, 同济大学经济与管理学院博士研究生, 研究方向: 情感分析和信息管理系统等。E-mail: wany.cheng@gmail.com

比较分析,认为从社交媒体中提取出来的度量指标具有更好的预测能力。学者们尝试使用各种不同的方法抽取社交媒体中的有效信息,研究其与证券市场的相互影响。常见的社交媒体度量指标包括数量指标、内容质量指标和结构特征指标,数量指标如发布信息数量、评论数量、转发数量和搜索强度等用户活跃度;内容质量指标如使用褒/贬义词出现的词数、包含褒/贬义词的信息数量和评论数量以及使用文本分析技术提取的情绪指数等方法计算获得的用户情绪和意见分歧度等;结构特征指标如节点位置和连通数量等。

由于数量指标收集和计算都相对简单,是最为常见的社交媒体度量指标。大量研究结果均表明,不同社交媒体的用户活跃度(包括搜索引擎的搜索量<sup>[6]</sup>、博客的发帖量和评论量<sup>[7]</sup>以及微博的发帖量和转发量<sup>[8-9]</sup>等)均与证券市场的交易活动有显著的相关关系。社交媒体的数量指标虽然能表示用户对所讨论话题的关注程度,但在表示用户意愿和用户关联等深层次信息时则有所不足。Sabherwal 等<sup>[10]</sup>利用文本分类器从 TheLion.com 论坛中抽取出投资者股票买卖倾向指标,发现社交媒体的用户情绪指标可以预测证券市场指数价格走势和交易活动,而用户活跃度并不具备相似的预测能力;Ruiz 等<sup>[11]</sup>使用 Twitter 作为数据来源,对 Twitter 中用户活跃度、结构特征等证券市场事件进行相关性分析,不仅发现 Twitter 与交易成交量的相关性强于其与收益的相关性,而且相对于用户活跃度而言, Twitter 的结构特征与股票市场之间的相关性更强。

为了进一步研究社交媒体与证券市场的关系,一些学者利用先进、有效的文本分析技术对社交媒体进行分析,抽取出投资者对证券市场涨跌的情绪倾向和意见分歧程度等深层次信息<sup>[11-18]</sup>。Antweiler 等<sup>[12]</sup>和 Das 等<sup>[13]</sup>分别通过构建文本分类器对论坛帖子进行文本挖掘,研究结果均表明,从社交媒体内容中提取的情绪指标与证券市场指数的多个主要指标(如股价、成交量和波动率等)都具有显著的相关性;Gilbert 等<sup>[14]</sup>和 Zhang 等<sup>[15]</sup>的研究表明,社交媒体中蕴含的用户悲观情绪会导致股价在短期内下跌,用户悲观情绪与证券市场指数呈负相关关系;Bollen 等<sup>[16]</sup>从微博中提取出 6 种情绪,发现政治、经济甚至日常生活中发生的事件都能够影响公众情绪,他们推测,采用情感分析技术对社交媒体内容进行深入的分析和建模可以预测社会和经济的发展趋势。

针对中文社交媒体与证券市场之间关系的研究正在初步探索中,林振兴<sup>[3]</sup>发现虽然论坛中发帖数量与 IPO 抑价、首日交易量存在显著正相关,但是采用人工判断的方法整理的基于 IPO 的投资者乐观/悲观情绪指标对中国证券市场 IPO 溢价和首日交易量并不存在显著相关性;张永杰等<sup>[17]</sup>利用百度搜索关键词所返回的网页数量作为社交媒体信息含量的度量指标,对社交媒体信息和资产定价进行分析,发现社交媒体中富含影响资产定价的有效信息;施

荣盛<sup>[18]</sup>发现论坛中超额发帖量与股票交易成交量显著正相关。可见,不断发展的中文社交媒体中富含影响中国证券市场交易活动的有效信息。

上述研究对社交媒体与证券市场的关系做出了有益尝试,但大多研究成果来自于英文社交媒体与欧美等发达国家的证券市场的相互关系研究,中国作为新兴加转型的资本市场,无论是相关制度和市场的信息效率,还是投资群体的构成等方面都具有独特性,因此国外现有成果是否适用于中国证券市场仍有待考察。此外,针对中文社交媒体与证券市场的关系研究主要是针对数量指标展开的,而利用文本分析技术对中文社交媒体做深层次挖掘有效信息的研究基本空白。因此,本研究使用知网等中文文本分析工具对新浪微博的内容进行语义层面的分析,尝试对中文社交媒体做进一步的研究,并考察社交媒体与证券市场之间的相互影响以及社交媒体对证券市场的预测作用。

### 3 研究设计

#### 3.1 构建社交媒体的投资者涨跌情绪指数

##### (1) 计算词语的涨跌情绪值

自然语言是人们表达和交流想法、感情的工具,而词语是自然语言处理的基本单位。表示情感倾向的词语通常称为极性词,本研究按照人们对证券市场未来走势的涨跌判断,将表示用户情绪的词语分为“看涨”和“看跌”两类涨跌极性词;另外,还有程度词和否定词两类词语,程度词是表示加重或减弱涨跌极性的词,否定词是表示否定涨跌极性的词。

在知网知识框架中,每个词语可以包含若干个概念,每个概念可以表述成若干个义原。参考 Li 等<sup>[19]</sup>的方法,使用知网可以判断任意词语的类别并计算其语义倾向值。首先建立一个种子义原库,包括“看涨”义原(bull)、“看跌”义原(bear)、程度义原(degree)和否定义原(deny)共 4 类种子义原。任意一个词语  $w$ ,分别计算该词语作为“看涨”词、“看跌”词、程度词和否定词的语义倾向值,使用  $P_{bull}$ 、 $P_{bear}$ 、 $P_{degree}$  和  $P_{deny}$  表示,所有语义倾向值均在  $[0, 1]$  之间。本研究按照以下方法识别词语  $w$  的类别并完成相应计算。

① 判断词语  $w$  是否为否定词,若  $P_{deny} \geq \max(P_{bull}, P_{bear}, P_{degree})$  且  $P_{deny} \geq 0.750$ ,则词语  $w \in deny$ ,累计否定词出现的次数;

② 若词语  $w \notin deny$ ,则再判断词语  $w$  是否为程度词,若  $P_{degree} \geq \max(P_{bull}, P_{bear}, P_{deny})$  且  $P_{degree} \geq 0.750$ ,则词语  $w \in degree$ ,其程度值  $Wdeg_w = P_{degree}$ ;

③ 若词语  $w \notin deny \wedge w \notin degree$ ,则词语  $w$  为涨跌极性词,其涨跌情绪值  $Wten_w = P_{bull} - P_{bear}$ 。

##### (2) 计算句子的涨跌情绪值

在中文文本分析之前需要对文本进行分词处理,本研究使用中国科学院计算技术研究所研制的 ICTCLAS 3.0 系统进行中文分词。

给定一个句子  $s$ ,句中否定词、程度词和涨跌极性

词的个数分别用  $N_{deny}$ 、 $N_{degree}$  和  $N_{moods}$  表示, 则句子  $s$  的涨跌情绪值  $Sten_s$  的计算公式为

$$Sten_s = \frac{(-1)^{N_{deny}} \prod_{i=1}^{N_{degree}} (1 + Wdeg_{w_i}) \sum_{j=1}^{N_{moods}} Wten_{w_j}}{N_{deny} + N_{degree} + N_{moods}} \quad (1)$$

其中,  $Wdeg_{w_i}$  为第  $i$  个程度词  $w_i$  的程度值,  $i = 0, 1, \dots, N_{degree}$ ;  $Wten_{w_j}$  为第  $j$  个涨跌极性词  $w_j$  的涨跌情绪值,  $j = 0, 1, \dots, N_{moods}$ 。

### (3) 计算帖子的涨跌情绪值

社交媒体中一条帖子通常只表达一个主题, 因此每条帖子的涨跌情绪值可以使用帖子中所有句子的平均涨跌情绪值表示。使用标点符号将帖子  $m$  分为  $N_{sen}$  个句子, 帖子  $m$  的涨跌情绪值  $Mten_m$  的计算公式为

$$Mten_m = \frac{\sum_{k=1}^{N_{sen}} Sten_{s_k}}{N_{sen}} \quad (2)$$

其中,  $Sten_{s_k}$  为第  $k$  个分句  $s_k$  的涨跌情绪值,  $k = 0, 1, \dots, N_{sen}$ 。

### (4) 构造基于社交媒体的投资者涨跌情绪指数

本研究使用一天内社交媒体中所有与证券相关的帖子的平均涨跌情绪值表示当天社交媒体的投资者涨跌情绪。令  $m$  为在  $(t-1)$  日下午 15:00 到  $t$  日下午 15:00 之间发布的第  $m$  条帖子, 且该时间段内共发布的帖子数为  $N_{msg}$  条, 则  $t$  日该社交媒体的投资者涨跌情绪指数  $OSMten_t$  的计算公式为

$$OSMten_t = \frac{\sum_{m=1}^{N_{msg}} Mten_m}{N_{msg}} \quad (3)$$

其中,  $Mten_m$  为第  $m$  条帖子的涨跌情绪值。若投资者涨跌情绪指数  $OSMten_t > 0$ , 表示  $t$  日内社交媒体中投资者心态乐观, 认为证券市场行情转好, 并且大盘指数和个股股价会上涨; 相反,  $OSMten_t < 0$ , 表示  $t$  日内社交媒体中投资者心态转为悲观, 认为证券市场行情不乐观, 而且大盘指数和个股股价下跌的可能性增加。 $OSMten_t$  绝对值越大, 表示投资者认为证券市场上涨(下跌)的可能性越大, 程度也越大。

### (5) 投资者涨跌情绪指数的有效性检验

从新浪微博上收集与股票相关的帖子, 通过人工标注进行分类整理后作为测试语料。分别由 3 人对帖子进行独立判断, 当帖子内容包含看涨或看跌含义时, 归类为“看涨”或“看跌”, 若未表达明确的涨跌倾向, 则归类为“中性”。本研究使用的测试语料包括“看涨”和“看跌”的帖子各 140 条。

计算测试语料帖子的投资者涨跌情绪指数, 若  $OSMten_t > 0$  为看涨帖子,  $OSMten_t < 0$  为看跌帖子, 否则为中性帖子。测试结果见表 1。

采用分类正确率评价算法效果, 即

$$\text{分类正确率} = \frac{\text{分类正确的帖子数}}{\text{测试语料帖子总数}} \times 100\% \quad (4)$$

表 1 算法有效性测试结果  
Table 1 Test Results of Effectiveness of Proposed Method

测试集	算法计算结果		
	看涨	中性	看跌
看涨帖子(140 条)	97	1	42
看跌帖子(140 条)	46	3	91

本研究的投资者涨跌情绪指数计算方法的平均分类正确率为 67.143%, 虽然在自动识别投资者涨跌情绪倾向上表现一般, 但考虑到社交媒体的帖子具有用语简洁、特征稀疏和指代不明确等特点, 而本研究使用的方法无须中文情感词典和中文情感语料库, 只需利用知网及少量的种子义原即可实现, 因此本研究使用的计数方法仍是有效且切实可行的。

### 3.2 研究模型

向量自回归 (VAR) 模型把需要分析的变量纳入一个系统进行分析, 在系统中待分析变量均设定为内生变量, 模型的每个方程中每一个内生变量利用自身的滞后变量和系统内其他内生变量的滞后变量进行回归。VAR 模型的优点在于可以通过将变量之间的动态性特征纳入考虑范围提高预测精度, 然而 VAR 模型无法直接观测到系统内变量之间的关系。Sims<sup>[20]</sup> 认为, 研究变量之间的动态关系比单独考察变量系数更有意义, 可以通过脉冲响应函数观察系统内各个经济变量之间的结构关系。脉冲响应函数是通过分析系统内变量对其他内生变量带来的冲击的响应, 分析系统内变量之间的动态影响和时滞关系。

社交媒体中投资者在受到证券市场交易活动影响的同时, 他们针对证券市场走势所发表的言论也影响着社区中其他投资者的投资决策行为。由于投资者阅读信息和进行投资决策行为的时间并不相同, 因此社交媒体与证券市场的交易活动可能在当期和未来几期都存在相互影响。本研究使用三元向量自回归模型检验社交媒体的投资者涨跌情绪指数与证券市场指数的收益、成交量之间的动态关系以及三者之间的时滞关系。本研究使用的  $p$  阶 VAR 模型可以表示为

$$y_t = c_0 + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \varepsilon_t \quad (5)$$

其中,  $y_t$  为内生变量向量;  $y_{t-u}$  为内生变量的  $u$  期滞后向量,  $u = 1, 2, \dots, p$ ;  $A_u$  为内生变量滞后项的系数矩阵,  $u = 1, 2, \dots, p$ ;  $c_0$  为常量向量;  $\varepsilon_t$  为协方差矩阵为单位矩阵的白噪声向量;  $p$  为滞后阶数;  $t = 1, 2, \dots, T$ ,  $T$  为时间长度。

### 3.3 数据来源

#### (1) 股票数据

本研究选取的证券市场指数样本来自上证综合指数和深圳成分指数, 包括上证综合指数的每日收盘指数  $PSH$ 、日成交量  $VLSH$  和深圳成分指数的每日

收盘指数  $PSZ$ 、日成交量  $VLMSZ$ 。股票数据均使用深圳市财富趋势科技股份有限公司提供的通达信软件下载获得。 $t$  日上证综合指数的股指收益率  $RSH_t$  和深圳成分指数的股指收益率  $RSZ_t$  的计算公式为

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (6)$$

其中,  $R_t = (RSH_t, RSZ_t)$ ,  $P_t$  为  $t$  日股指的收盘指数,  $P_{t-1}$  为  $(t-1)$  日股指的收盘指数。

样本区间为 2011 年 1 月 1 日至 2012 年 2 月 29 日, 每个股票数据样本序列包含 280 个有效数据。

#### (2) 社交媒体的投资者涨跌情绪指数数据

本研究采用新浪微博作为社交媒体的数据来源。新浪微博从 2009 年开通后便成为中国最重要的微博服务提供网站之一。截至 2011 年底, 新浪微博的注册用户已突破 3 亿, 用户每日发博量超过 1 亿条。但与 Twitter 用户主要关注时事新闻的特点不同, 新浪微博用户大多关注休闲娱乐方面的话题<sup>[21]</sup>。为了尽可能地获取投资者对证券市场信息的看法和情绪, 首先把数据来源的范围限定在与财经、证券相关的用户群体中。由于新浪微博信息源的可信赖程度、专业程度、吸引程度和使用多媒体信息的数量都会影响信息的转发数量<sup>[22]</sup>, 因此, 与普通用户相比, 认证媒体机构用户不仅更具可信赖性和专业性, 还能实时报道和分析盘中走势, 并且分享来自不同渠道的证券市场相关信息, 从而受到更多投资者的关注和评论。

社交媒体用户在阅读帖子时, 当认为帖子内容对朋友或他人具有一定价值时, 他有可能会原帖转发; 当认为帖子内容很重要, 尤其是涉及到自身利益时, 会倾向于撰写评论表达自己的看法, 抒发自己的情感, 并期望得到他人的认可。可见, 社交媒体的评论表达了用户对所讨论话题的真实意见和想法。

本研究选择 5 个与时事新闻或财经证券市场相关的认证媒体机构用户所发表的微博以及相应的评论作为计算投资者涨跌情绪指数的数据来源, 分别

为新浪证券、新浪财经、头条新闻、证券市场红周刊和证券市场周刊, 他们均与时事新闻或财经证券市场相关, 而且截至 2011 年 11 月, 他们都已发布超过 1 万条的微博和拥有 10 万人以上的粉丝, 具有一定的影响力和代表性。本研究共收集微博 66 317 条和评论 1 207 693 条作为数据样本, 时间跨度从 2011 年 1 月 1 日到 2012 年 2 月 29 日。按照前文的方法从数据样本中抽取并合成新浪微博的投资者涨跌情绪指数  $OSMten$ , 为了与证券市场指数数据做对比分析, 投资者涨跌情绪指数  $OSMten$  仅使用证券市场交易日的数据作为分析序列, 包含 280 个有效数据。

#### 4 实证分析和结果

实验分析使用 Eviews 6.0 软件完成, 数据均保留小数点后 3 位。

##### 4.1 变量的基本统计描述

表 2 给出所有变量的基本统计描述信息。在样本区间内, 证券市场指数的收益和投资者涨跌情绪指数的均值均为负值, 可以看出在这一年多时间内, 中国证券市场总体走势低迷, 社交媒体中的投资者对证券市场的未来表现感到悲观。此外, 证券市场指数收益的波动远大于社交媒体的投资者涨跌情绪指数的波动。

由于 VAR 模型的建立和 Granger 因果检验都要求使用的时间序列是平稳时间序列, 采用 Phillips Perron 单位根检验方法对各变量进行平稳性检验, 检验方法使用带有漂移项和趋势项的一般形式, 结果见表 2。所有变量的 PP 检验值都小于 1% 显著性水平的统计量, 显示全部时间序列都是平稳的。

表 3 给出所有变量间的相关系数。投资者涨跌情绪指数与证券市场指数的收益、成交量均呈现正向相关关系, 且与交易成交量的正相关关系在 1% 显著性水平下显著。显然, 社交媒体的投资者涨跌情绪指数与证券市场指数存在正相关关系, 并且投资者涨跌情绪指数与交易成交量之间的正相关关系强于其与收益之间的正相关关系。

表 2 描述性统计  
Table 2 Descriptive Statistics

变量	平均值	最大值	最小值	标准差	偏度	峰度	PP 检验值
$OSMten$	-0.116	0.185	-0.713	0.090	-0.862	9.185	-8.979***
$RSH$	-0.045	4.179	-3.792	1.188	-0.004	3.760	-17.651***
$RSZ$	-0.066	4.946	-4.551	1.444	0.124	3.649	-17.597***
$VLMSH$	8.610E+05	1.685E+06	3.963E+05	2.804E+05	0.587	2.785	-5.922***
$VLMSZ$	7.136E+04	1.762E+05	3.261E+04	2.814E+04	1.079	3.916	-6.583***

注: \*\*\* 为显著性水平为 1%, 下同; PP 检验标准为 1% 显著性水平统计量为 -3.991。

表 3 变量间的相关系数  
Table 3 Correlation Coefficients of Variables

	<i>OSMten</i>	<i>RSH</i>	<i>RSZ</i>	<i>VLMSH</i>	<i>VLMSZ</i>
<i>RSH</i>	0.053	1.000			
<i>RSZ</i>	0.050	0.958 ***	1.000		
<i>VLMSH</i>	0.334 ***	0.144 **	0.140 **	1.000	
<i>VLMSZ</i>	0.368 ***	0.211 ***	0.213 ***	0.902 ***	1.000

注: \*\* 为显著性水平为 5%, 下同。

4.2 Granger 因果关系检验

Granger 因果关系检验可以从统计意义上分析两个平稳时间序列  $X$ 、 $Y$  之间的因果关系,判断在对  $X$ 、 $Y$  变量进行回归时,加入  $X$  的滞后项能否显著地改进对  $Y$  的预测,如果能够显著改进对  $Y$  的预测,则认为  $X$  是  $Y$  的 Granger 原因,或认为  $X$  Granger 引起  $Y$ 。如果  $Y$  同时也是  $X$  的 Granger 原因,则认为  $X$  与  $Y$  之间具有双向 Granger 因果关系。虽然 Granger 因果关系不等于因果关系,但若  $X$  是  $Y$  的 Granger 原因,则  $X$  可作为预测  $Y$  的有效预测变量<sup>[23]</sup>。因此,本研究采用

Granger 因果关系检验分析社交媒体的投资者涨跌情绪指数对证券市场指数收益和成交量的预测能力。

本研究分别将社交媒体的投资者涨跌情绪指数与证券市场指数的收益、成交量做两两检验,表 4 列出相应的 Granger 因果关系检验结果。在滞后期为 1 阶~5 阶时,可以发现,①证券市场指数收益不是投资者涨跌情绪指数的 Granger 原因均被拒绝,且显著性水平都达到 5%;投资者涨跌情绪指数不是证券市场指数的收益的 Granger 原因均被接收,说明在中国证券市场中,证券市场指数的收益是社交媒体的投资者涨跌情绪指数的 Granger 原因;②证券市场指数的成交量与社交媒体的投资者涨跌情绪指数存在双向 Granger 关系,且在显著性水平 10% 下有效,表明证券市场指数的成交量与社交媒体的投资者涨跌情绪指数互为 Granger 原因。

可见,社交媒体的投资者涨跌情绪指数同时受证券市场指数的收益和成交量变动的影响。同时,社交媒体的投资者涨跌情绪指数能影响投资者的投资决策行为,从而引起证券市场指数的交易成交量变动。本研究认为,社交媒体的投资者涨跌情绪指数虽然无法帮助预测证券市场指数收益,但能够帮助预测证券市场指数的交易成交量,可以成为预测交易成交量的有效预测指标。

表 4 Granger 因果检验结果  
Table 4 Granger Causality Test Results

原假设	F 统计量					结论
	滞后 1 阶	滞后 2 阶	滞后 3 阶	滞后 4 阶	滞后 5 阶	
<i>RSH</i> 不是 <i>OSMten</i> 的 Granger 原因	9.186 *** (0.003)	7.134 *** (0.001)	4.966 *** (0.002)	4.317 *** (0.002)	4.078 *** (0.001)	拒绝
<i>OSMten</i> 不是 <i>RSH</i> 的 Granger 原因	1.001 (0.318)	0.553 (0.576)	0.518 (0.670)	0.781 (0.538)	0.553 (0.736)	接受
<i>RSZ</i> 不是 <i>OSMten</i> 的 Granger 原因	5.267 ** (0.023)	4.187 ** (0.016)	2.790 ** (0.041)	2.718 ** (0.030)	2.681 ** (0.022)	拒绝
<i>OSMten</i> 不是 <i>RSZ</i> 的 Granger 原因	1.146 (0.285)	0.644 (0.526)	0.564 (0.639)	0.766 (0.548)	0.595 (0.704)	接受
<i>VLMSH</i> 不是 <i>OSMten</i> 的 Granger 原因	7.968 *** (0.005)	3.375 ** (0.036)	4.198 *** (0.006)	2.796 ** (0.027)	1.919 * (0.091)	拒绝
<i>OSMten</i> 不是 <i>VLMSH</i> 的 Granger 原因	6.197 ** (0.013)	4.287 ** (0.015)	3.306 ** (0.021)	1.987 * (0.097)	1.925 * (0.090)	拒绝
<i>VLMSZ</i> 不是 <i>OSMten</i> 的 Granger 原因	8.560 *** (0.004)	3.761 ** (0.025)	4.293 *** (0.006)	3.575 *** (0.007)	2.172 * (0.058)	拒绝
<i>OSMten</i> 不是 <i>VLMSZ</i> 的 Granger 原因	9.674 *** (0.002)	6.372 *** (0.002)	4.346 *** (0.005)	3.246 ** (0.013)	2.881 ** (0.015)	拒绝

注: 括号内数据为伴随概率, \* 为显著性水平为 10%。

### 4.3 脉冲响应函数分析

使用新浪微博的投资者涨跌情绪指数分别与两个证券市场指数的收益和成交量构造两个三元向量,即  $(OSMten, RSH, VLMSH)'$  和  $(OSMten, RSZ, VLMSZ)'$ ,建立如3.2节所描述的三元 VAR 模型。根据 AIC 和 SC 准则为两个 VAR 模型选择最优滞后阶数, AIC 准则建议选择滞后 4 阶, SC 准则建议选择滞后 1 阶,考虑到低估滞后阶数会丢失变量间的动态关系,因此本研究选择滞后 4 阶作为两个 VAR 模型的最优滞后阶数。使用 AR 特征多项式根的倒数检验两个 VAR 模型的稳定性,两个 VAR 模型的根模的倒数都小于 1,即落在单位圆内,说明两个 VAR 模型都是稳定的,可以进行脉冲响应函数分析。

脉冲响应函数是通过在一个扰动项上施加一个一次性的冲击,用以研究对系统内生变量的当前值和未来值所带来的影响。因此,本研究在  $(t+0)$  期期末给样本序列施加一个正向单位标准差冲击,两个 VAR 模型中各变量相应的动态响应如图 1 和图 2 所示。图中实线为脉冲响应函数,虚线为正负 2 倍标准差大小的置信区间。

图 1 描述了证券市场指数的收益和成交量对社交媒体的投资者涨跌情绪指数冲击的响应。图 1(a)

和图 1(b) 中,证券市场指数收益受投资者涨跌情绪指数的一个正向单位标准差冲击时,立即产生正向响应,正向响应持续 5 期左右后立刻变小并趋近于 0,说明社交媒体的投资者涨跌情绪指数对证券市场指数的收益只存在短期同向影响,而且响应集中在前 5 个交易日内。

图 1(c) 和图 1(d) 中,证券市场指数的成交量受投资者涨跌情绪指数的一个正向单位标准差冲击时,第 1 期没有明显的响应,但第 2 期出现的正向响应快速变大,在第 4 期左右达到最大值后逐渐减弱,响应持续时间超过 40 期。可见,社交媒体的投资者涨跌情绪指数的变动会对证券市场指数的交易成交量造成正向影响,且持续时间超过 40 个交易日。

图 2 描述了社交媒体的投资者涨跌情绪指数对证券市场指数的收益和成交量冲击的响应。从图 2(a) 和图 2(b) 可以看出,受证券市场指数的收益的一个正向单位标准差冲击时,社交媒体的投资者涨跌情绪指数表现出同向响应,响应在前 5 期内达到最大值,在前 10 期内显著存在,随后正向响应逐渐减弱,在 40 期左右趋近于 0。当证券市场指数上涨,大多数投资者变得乐观,社交媒体的投资者涨跌情绪指数也上升;当证券市场指数下跌,大多数投资者转

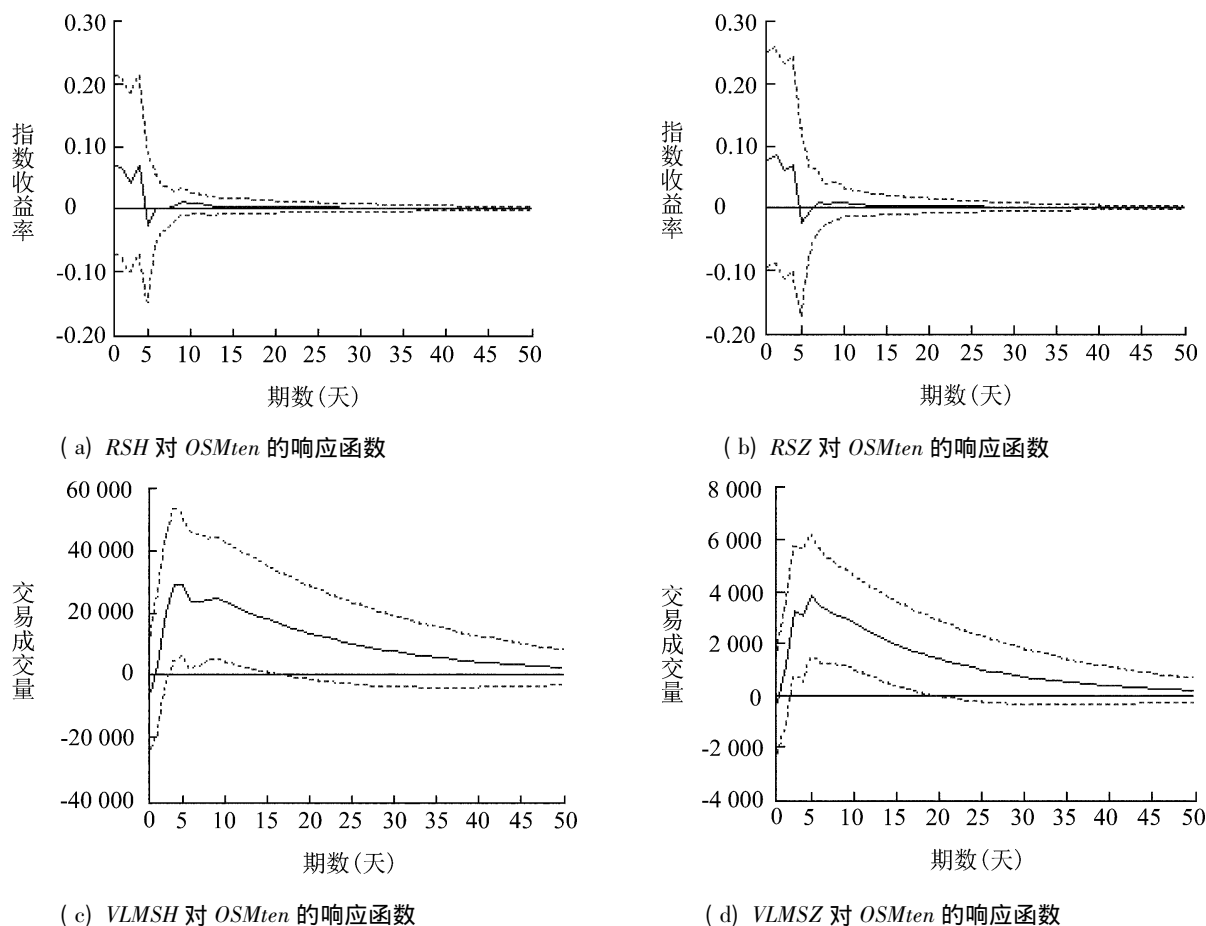


图 1 证券市场指数的收益、成交量对社交媒体的投资者涨跌情绪指数冲击的响应函数  
Figure 1 Impulse Response Function of Stock Market Indices Returns and Trading Volumes to Investors' Bullish Sentiment Indices of Social Media

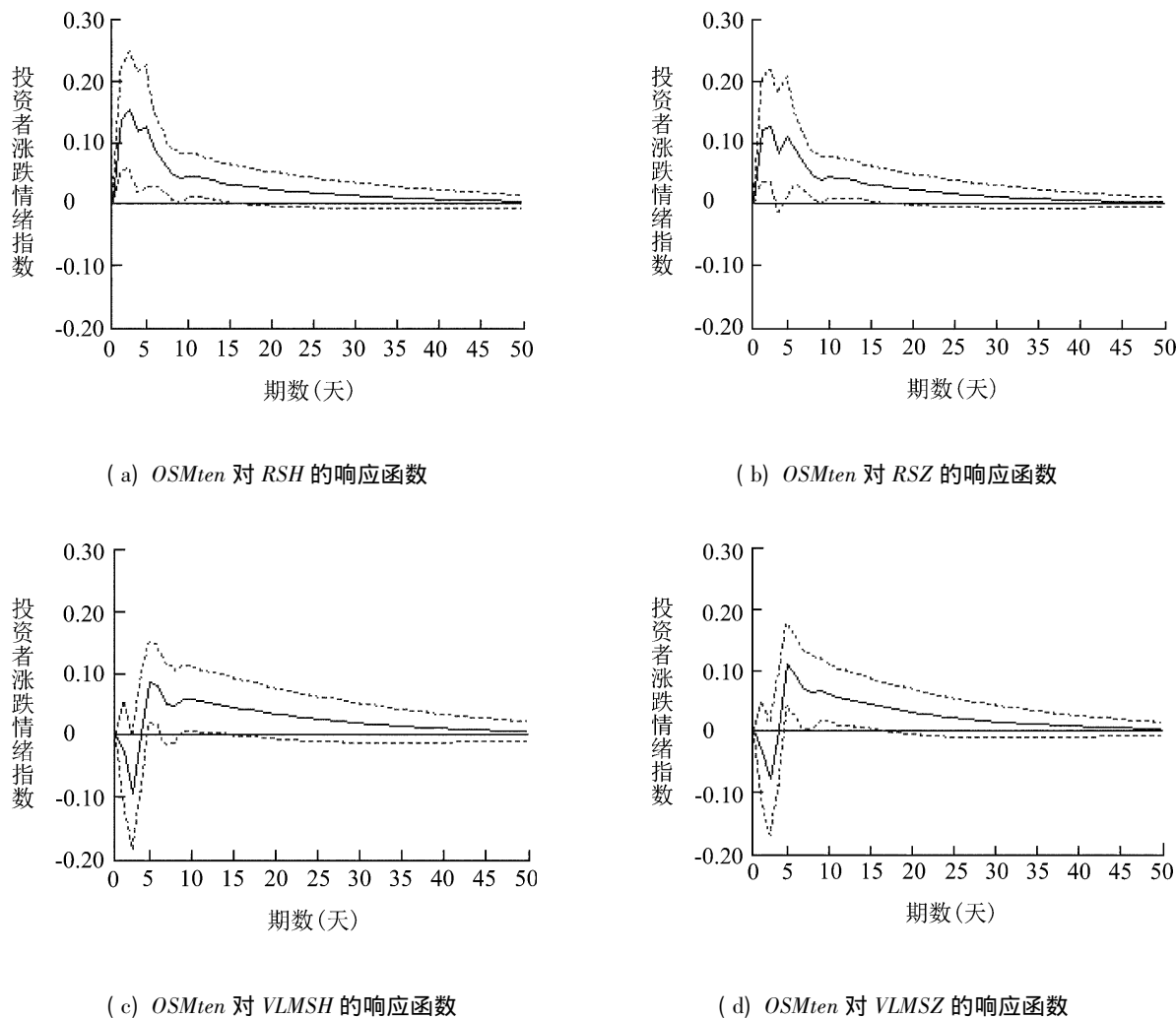


图2 社交媒体的投资者涨跌情绪指数对证券市场指数的收益及成交量冲击的响应函数

Figure 2 Impulse Response Function of Investors' Bullish Sentiment of Social Media to Stock Market Indices Returns and Trading Volumes

为悲观,社交媒体的投资者涨跌情绪指数下降。

在图2(c)和图2(d)中,受证券市场指数的交易成交量的一个正向单位标准差冲击时,社交媒体的投资者涨跌情绪指数第1期没有响应,在之后2期内响应达到负向最大值,第4期响应减弱并趋近于0,但第5期出现最大正向响应,随后正向响应逐渐变小,持续时间超过40期。可以看出,成交量的变动对投资者会造成一定的困扰,社交媒体的投资者涨跌情绪指数除前3期出现反向响应外,总体上表现出同向响应,且同向响应持续时间在40个交易日左右。

综合对图1和图2的分析,社交媒体的投资者涨跌情绪指数与证券市场指数的收益、成交量之间均存在正向影响,但影响持续的时间不同。证券市场指数的收益的变动会在较长时间内影响社交媒体中投资者的涨跌情绪,而社交媒体中投资者的涨跌情绪仅对证券市场指数的收益带来短期影响;此外,社交媒体中投资者的涨跌情绪与证券市场的交易成交量之间的正向影响会持续超过40个交易日。

## 5 结论

本研究以新浪微博为样本,采用中文文本分析技术从中提取出大多数投资者对证券市场未来走势的涨跌情绪倾向,合成社交媒体的投资者涨跌情绪指数,将该指数用以研究社交媒体与证券市场指数的收益和成交量的相互影响及预测能力。研究结果表明,社交媒体的投资者涨跌情绪指数与证券市场指数的收益、成交量之间都存在正相关关系。社交媒体的投资者涨跌情绪指数表现乐观(悲观)时,证券市场指数的收益会在短期内出现上涨(下跌),而且市场反应积极(消极),成交量会持续较长时间放大(缩小)。当证券市场指数的收益和成交量发生变动时,投资者对两者的反应虽然稍有不同,但投资者对两者的关注时间都较长,在社交媒体中投资者涨跌情绪在40个交易日左右都受到两者变化的影响。

证券市场指数的收益和成交量的变动都能够引起投资者情绪的变动,此结论与现实相符。而社交媒体的投资者涨跌情绪指数无法帮助预测证券市场

指数的收益,却有助于预测短期内证券市场指数的交易成交量。原因可能是社交媒体的信息融入证券市场指数价格的时间稍短,使用每日数据仍很难观测到社交媒体与证券市场指数的收益的因果关系;同时,由于投资者的专业知识和心理素质有所差异,对信息的理解时间和做出交易决策的时间也有长短,因此社交媒体的投资者涨跌情绪指数可以成为预测交易成交量的有效指标。

以投资者涨跌情绪指数为代表,社交媒体中蕴含着大量能辅助投资者进行有效投资决策的有效信息。随着中国证券市场的快速发展,社交媒体在证券市场的影响能力日益增加,本研究的实证结论对帮助投资者正确、有效地理解和使用社交媒体并辅助投资决策具有重要现实意义,并且可以帮助投资者去掉投机心理,树立投资理念,增强风险意识和投资决策能力。

如何更好、更及时地挖掘社交媒体中的有效信息,分析社交媒体的内容和结构等特征与证券市场的深层次关系是今后研究的重要内容。

#### 参考文献:

- [1] Mizrach B ,Weerts S. Experts online: An analysis of trading activity in a public internet chat room [J]. *Journal of Economic Behavior & Organization* ,2009 , 70( 1/2) : 266-281.
- [2] Van Bommel J. Rumors [J]. *The Journal of Finance* , 2003 ,58( 4) : 1499-1520.
- [3] 林振兴. 网络讨论、投资者情绪与 IPO 抑价 [J]. *山西财经大学学报* ,2011 ,33( 2) : 23-29.  
Lin Zhenxing. Internet discussions ,investor sentiment and IPO underpricing [J]. *Journal of Shanxi Finance and Economics University* ,2011 ,33( 2) : 23-29. ( in Chinese)
- [4] Hong H ,Kubik J D ,Stein J C. Social interaction and stock-market participation [J]. *The Journal of Finance* ,2004 ,59( 1) : 137-163.
- [5] Mao H ,Counts S ,Bollen J. Predicting financial markets: Comparing survey ,news ,Twitter and search engine data [J/OL]. Ithaca NY: ArXiv e-prints , 2011. <http://arxiv.org/pdf/1112.1051.pdf>.
- [6] Joseph K ,Wintoki M B ,Zhang Z. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search [J]. *International Journal of Forecasting* ,2011 ,27( 4) : 1116-1127.
- [7] De Choudhury M ,Sundaram H ,John A ,Seligmann D D. Can blog communication dynamics be correlated with stock market activity? [C] // *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*. New York ,NY: ACM ,2008: 55-60.
- [8] Mao Y ,Wei W ,Wang B ,Liu B. Correlating S&P 500 stocks with Twitter data [C] // *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*. New York ,NY: ACM ,2012: 69-72.
- [9] Zhang X ,Fuehres H ,Gloor P A. Predicting asset value through Twitter buzz [C] // Altmann J ,Baumöl U ,Krämer B J. *Advances in Collective Intelligence* 2011. Berlin: Springer ,2012: 23-34.
- [10] Sabherwal S ,Sarkar S K ,Zhang Y. Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news [J]. *Journal of Business Finance & Accounting* , 2011 ,38( 9/10) : 1209-1237.
- [11] Ruiz E J ,Hristidis V ,Castillo C ,Gionis A ,Jaimes A. Correlating financial time series with micro-blogging activity [C] // *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*. New York ,NY: ACM ,2012: 513-522.
- [12] Antweiler W ,Frank M Z. Is all that talk just noise? The information content of internet stock message boards [J]. *The Journal of Finance* ,2004 ,59( 3) : 1259-1294.
- [13] Das S R ,Chen M Y. Yahoo! For Amazon: Sentiment extraction from small talk on the web [J]. *Management Science* ,2007 ,53( 9) : 1375-1388.
- [14] Gilbert E ,Karahalios K. Widespread worry and the stock market [C] // *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* ,2010: 58-65.
- [15] Zhang X ,Fuehres H ,Gloor P A. Predicting stock market indicators through Twitter “I hope it is not as bad as I fear” [J]. *Procedia-Social and Behavioral Sciences* ,2011 ,26: 55-62.
- [16] Bollen J ,Mao H ,Pepe A. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena [C] // *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* , 2011: 450-453.
- [17] 张永杰 ,张维 ,金曦 ,熊熊. 互联网知道的更多么?: 网络开源信息对资产定价的影响 [J]. *系统工程理论与实践* ,2011 ,31( 4) : 577-586.  
Zhang Yongjie ,Zhang Wei ,Jin Xi ,Xiong Xiong. Does the internet know more?: Open source information and asset pricing [J]. *Systems Engineering - Theory & Practice* ,2011 ,31( 4) : 577-586. ( in Chinese)
- [18] 施荣盛. 投资者关注与分析师评级漂移: 来自中国股票市场的证据 [J]. *投资研究* ,2012 ,31( 6) : 133-145.  
Shi Rongsheng. Investor attention and drift of analyst ratings: Evidence from the Chinese stock market [J]. *Review of Investment Studies* ,2012 ,31( 6) : 133-145. ( in Chinese)



- [19] Li D ,Ma Y ,Guo J. Words semantic orientation classification based on HowNet [J]. The Journal of China Universities of Posts and Telecommunications , 2009 ,16( 1) : 106-110.
- [20] Sims C A. Macroeconomics and reality [J]. Econometrica ,1980 ,48( 1) : 1-48.
- [21] Yu L ,Asur S ,Huberman B A. What trends in Chinese social media [C] // Proceedings of the 5th Workshop on Social Network Mining and Analysis. New York ,NY: ACM ,2011: 1-9.
- [22] Liu Z ,Liu L ,Li H. Determinants of information retweeting in microblogging [J]. Internet Research , 2012 ,22( 4) : 443-466.
- [23] 易会文. 格兰杰因果检验用法探讨 [J]. 中南财经政法大学研究生学报 ,2006( 5) : 34-36 23.
- Yi Huiwen. Discussion on how to do Granger causality test [J]. Journal of the Postgraduate of Zhongnan University of Economics and Law ,2006( 5) : 34-36 , 23. ( in Chinese)

## Investors' Bullish Sentiment of Social Media and Stock Market Indices

Cheng Wanyun , Lin Jie

School of Economics and Management , Tongji University , Shanghai 200092 , China

**Abstract:** Based on data from SINA Weibo and the Chinese text analysis techniques , the study extracted information of major investors' sentiment orientation about the stock markets' ups and downs from SINA weibo and comments of five verified institutional users related with Chinese stock markets. Constructing bullish sentiment indices of investors on social media as the measuring indices of social media and based on Granger causality tests and impulse response functions , we analyze the dynamic relationship between social media and stock market , as well as the forecasting power of social media for stock market. The empirical results demonstrate that: ①there are positive relationships among investors' bullish sentiment indices , stock market indices returns and trading volumes. The impacts of stock market indices returns and trading volumes on investors' bullish sentiment indices of social media last more than 40 trading days; and ②contrarily , only the short-term impacts of investors' bullish sentiment indices of social media on stock market indices returns are significant. Moreover , the investors' bullish sentiment indices of social media can help to predict stock market trading volumes , but not stock market indices returns. There is abundant and effective information in social media that is helpful for investment decision. The research conclusions will not only help investors to determine the effective information of social media timely and correctly , but also help them to fully apply social media for making investment decision.

**Keywords:** social media; stock market indices; investors' bullish sentiment indices; text analysis

**Received Date:** February 22<sup>nd</sup> , 2013      **Accepted Date:** September 24<sup>th</sup> , 2013

**Funded Project:** Supported by the National Natural Science Foundation of China( 71071114) , the Program of Humanities and Social Sciences of Education Ministry( 11YJC630216) , the Shanghai Leading Academic Discipline Project( B310) and the Natural Science Foundation of Guangdong Province( S2012010010794)

**Biography:** Cheng Wanyun , a Guangdong Nanhai native ( 1981 - ) , is a Ph. D. candidate in the School of Economics and Management at Tongji University. Her research interests include sentiment analysis and Management information system , etc. E-mail: wany. cheng@ gmail. com      □