

# 网络搜索能预测股票市场吗?

张谊浩 李元 苏中锋 张泽林

(南京大学商学院, 江苏南京 210093; 中国人民大学商学院, 北京 100872)

**摘 要:**网络搜索与股票市场间相互作用, 投资者网络搜索行为会对资产定价形成影响。本文实证研究发现: (1) 投资者网络搜索强度对股票短期收益率、短期交易量及累积收益率均有影响。 (2) 股票市场能影响网络搜索, 但网络搜索可以在更大程度上影响并预测股票市场的表现, 网络搜索与股票收益间存在的内生性问题对预测效果的影响很小。模拟分析发现, 根据网络搜索构建投资组合可以获取超额收益。 (3) 投资者网络搜索强度指标与传统的投资者情绪和投资者关注指标间存在较强的相关关系; 相较于投资者情绪和投资者关注, 投资者网络搜索对股票市场的解释力及预测效力更强。

**关键词:**网络搜索; 投资者情绪; 投资者关注; 股票市场

**JEL 分类号:** C82, G02, G17 **文献标识码:** A **文章编号:** 1002-7246(2014)02-0193-14

## 一、引言

相比报纸、广播、电视等传统媒介, 互联网具有信息量大、传播速度快、定制化程度高、获取成本低等优势。在“股民”和“网民”高度耦合的数字化时代, 网络搜索对股票投资者有重要影响。股票投资者的主要互联网应用方式有搜索、浏览、论坛、博客和交易等。从网络行为逻辑看, 投资者在搜索引擎中搜索相关股票的关键词之后, 才会有相继的浏览、讨论、发布及交易行为。网络搜索是用户基于个人偏好和特定目标需求找寻信息的过程,

收稿日期: 2013-05-12

作者简介: 张谊浩, 经济学博士, 南京大学商学院教授, Email: zhangyh@nju.edu.cn.

李元, 金融学硕士研究生, 南京大学商学院, Email: 568186953@qq.com.

苏中锋, 管理学博士, 南京大学商学院副教授, Email: zhongfengsu@163.com.

张泽林 (通讯作者), 市场营销学博士, 中国人民大学商学院讲师, Email: zhangzelin@rbs.org.cn.

\* 本文系国家自然科学基金(71202107)及江苏高校优势学科建设工程资助项目(PAPD)的阶段性研究成果, 得到教育部“新世纪优秀人才支持计划”(NCET-13-0286)、中国人民大学新教师启动金项目(13XNF053)及南京大学人文社会科学高级研究院的资助。感谢南京大学经济学院刘志彪教授、安同良教授和林树副教授等的宝贵建议, 感谢匿名审稿专家和编辑部老师的建设性意见, 文责自负。

而用户目标需求又受外部环境的影响,因此对相关关键词的网络搜索与外部环境变化存在必然联系。因此,大量学者利用网络搜索引擎数据对社会经济问题进行预测。

基于网络搜索与经济行为的高度相关性,以及受到行为金融学强调的利用微观个体心理特征及决策行为来研究资产定价并解释、研究和预测金融市场发展的启发,本文借鉴和发展 Joseph, Wintoki & Zhang(2011)的方法(该文献第三作者是本文通讯作者),针对沪深 300 指数的成分股,使用百度搜索统计数据研究股票的网络搜索行为强度(以下简称搜索强度)与市场短期收益率、短期交易量及累积收益率等的关系,并对比网络搜索强度和现有投资者情绪及投资者关注指标对市场收益解释力的差异,进而解释网络搜索能否预测股票市场这一问题。本文的主要贡献是:(1)构造出视角独特且思路新颖的网络搜索影响股票市场的理论分析框架;(2)运用百度搜索统计数据从微观视角考察股票的搜索强度与收益率等指标间的关系,探讨网络搜索能否预测具有起步晚、不成熟、噪音交易者多等特征的中国股票市场这个理论问题;(3)基于网络搜索强度的变动构造出搜索强度正向变动率这一新的反映投资者网络搜索行为动态的指标;且实证发现相较于传统的投资者情绪和投资者关注度指标,该指标预测中国股票市场的效果更好。

## 二、网络搜索对股票市场影响的理论分析

### (一)网络搜索预测股票市场的可行性

股票收益和价格波动能否预测及如何预测,一直是金融学研究焦点问题,国内外学者在此方面积累了大批文献。随着信息技术在金融市场中正扮演越来越重要的角色,通过投资者网络行为来预测股票市场逐步成为一个新课题,一些学者对此也有所尝试。早期研究主要分析网络新闻及论坛对股票市场的影响。如 Antweiler & Frank(2004)对雅虎财经论坛中的股票股评帖子进行分析,发现网络讨论能预测股票市场的波动性,网络讨论意见的分歧与交易量相关。Tumarkin & Whitelaw(2001)发现论坛异常发帖量、内容倾向性变化与股票市场的异常交易量、异常收益相关。董大勇、肖作平(2011)选用川辽两省上市公司数据和东方财富网股吧发帖数据,研究了信息交流中“家乡偏误”问题。

随着网络搜索引擎数据的积累和公开,部分学者开始基于网络搜索来预测股票市场。如 Da 等(2009)利用谷歌的网络搜索量指标,以 Russell 3000 指数成分股为样本,证实网络搜索量可作为投资者关注度的直接度量指标。Joseph 等(2011)结合标准普尔 500 指数(S&P 500)数据,发现网络搜索行为包含投资者情绪,并构造出一个能够代表市场风险因子的情绪指标。贾春新等(2010)以谷歌搜索的历史资讯数量作为投资者关注程度的衡量标准,发现投资者关注会引起股票正回报。俞庆进、张兵(2012)发现百度指数和创业板股票市场表现之间存在相关性,百度指数可以衡量中小投资者的有限关注。上述文献表明,基于网络搜索来预测股票市场是可行的。

### (二)网络搜索对股票市场的影响机制

网络搜索是连接信息资源与用户需求的纽带,它在满足投资者信息需求的同时也记

录了投资者的动态,可以指示互联网上投资者整体的行为趋势。股票投资者对特定关键词(如上市公司名称、代码、影响因素等对应名词)的网络搜索趋势在一定程度上能够反映投资者预期,相应决策会影响股票市场的交易结果。因此,投资者使用网络搜索获取信息的过程及结果,会从质和量两方面影响投资者判断和决策,进而表现为股市波动。

股票市场上的个体及机构投资者都会用到网络搜索,但两者的网络搜索行为具有明显差异。由于规模、人才和资金实力更强,机构投资者在信息渠道、获取技术和处理效率上具有比较优势,一般较少依赖网络搜索进行股票投资。而个体投资者,无论是股龄长、买卖经验丰富的“老股民”,还是新进入或投资经验不足的“新股民”,在专业投资软件对网络搜索软件无法完全替代的条件下,出于各种动机都倾向于使用网络搜索来获取股票投资信息。股票市场价格走势由噪音交易者(noise traders)和套利交易者(arbitrageurs)的博弈所决定(Shleifer & Summers, 1990),当市场上有大量的依据错误主观信念或与公司基本面无关的信息做出非理性决策的噪音交易者时,噪音交易者会对股票价格造成持续性影响,导致资产价格并非如有效市场理论所预测的那样。因此,在个体投资者至今仍是中國股票市场最大资金供给者,同时也是市场中最主要交易者的背景下(谭松涛、陈玉宇, 2012),更接近于噪音交易者的个体投资者主导的网络搜索对股票市场的影响可能更加显著,这进一步凸显本文的研究价值。

网络搜索对股票市场存在的实质性影响,引起学者们注意网络搜索影响股票市场的机理。但遗憾的是目前还没有公认的观点。结合已有研究,本文尝试利用图1来阐释网络搜索影响股票市场的具体机理。

按照通常定义,投资者情绪(investor sentiment)是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,不能完全反映当前已有的事实(Baker & Wurgler, 2006)。投资者关注(investor attention)则体现外界输入信息被理解(记忆)和反应选择这类感觉层次上的需求,投资者关注分为关注和有限关注(Engelberg等, 2012)。图1中,作为体现投资者在理性和非理性思维交叉驱动下的信息获取行为,个体投资者主导的网络搜索与投资者情绪、投资者关注以及其它认知、意志和从众等因素之间存在作用与反作用关系,这种互动关系类似于一个只有部分可观测的“灰箱(Gray box)”(以下简称灰箱A)。在灰箱A内部,一方面投资者情绪和投资者关注会驱使投资者进行特定的网络搜索(传导机制①)。以个体投资者对股票名称(或代码)的网络搜索为例,当某个时点上投资者从众多股票中搜索一只或多只股票时,就体现出投资者关注变化对投资者网络搜索的驱动作用;当某一时段内投资者群体对同一只股票或多只股票的搜索频度出现周期性起伏时,则反映出个体投资者之间基于信息交流及心理传染下的投资者情绪起伏对投资者网络搜索的推动作用。另一方面,投资者网络搜索又会引起投资者情绪和投资者关注的进一步改变(传导机制②)。当个体投资者对股票名称(或代码)的网络搜索由初次(浅层)向多次(深层)发展时,信息的搜索框架调整及获取容量放大会进一步左右投资者关注或投资者情绪的嬗变。

灰箱A内部,投资者情绪、投资者关注和其他认知、意志和从众等因素之间也具有高

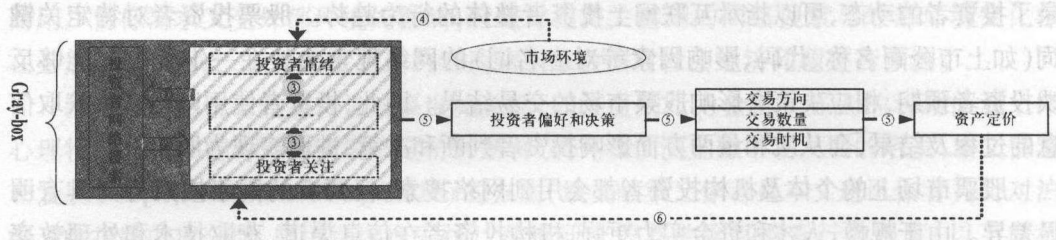


图1 网络搜索影响股票市场的机理

度交叉性(传导机制③),理论上很难厘清其相互关联,实践中也不易将其严格区分(Da等,2009)。以投资者对股票名称(或代码)的搜索为例,如果投资者在同一时点对一只股票或多只股票的关注趋同时,会导致异质信念的个体投资者预期同质化,引发投资者群体情绪的涌现及波动,即投资者关注引发投资者情绪。反之,投资者情绪的唤醒水平又影响投资者关注程度的分配,进而影响投资者的认知操作效率(权小锋、吴世农,2012)。投资者持乐观情绪时会扩大投资者注意范围,投资者关注度提高;投资者持悲观情绪时会集中注意范围,投资者关注度降低,即投资者情绪引起投资者关注(Rowe等,2007)。由于现实的个体投资者并非独立决策,如果考虑个体投资者相互间的信息交流、情绪传染及群体压力对投资者(包括群体和个体)关注水平的影响(Banerjee,1992),投资者情绪和投资者关注之间的这种互动关联可能更复杂。从灰箱A外部来讲,它还受法律制度、政策安排、牛熊状态等市场环境的持续性及非对称性影响(传导机制④)。以灰箱A中的投资者情绪、投资者关注受市场环境制约为例:一方面,投资者情绪和市场行情密切相关。当股市处于牛市态势下,持有积极情绪的投资者会提高未来的预期回报、低估风险,发生更多的交易行为;相反,当股市处于熊市态势时,持有消极情绪的投资者则降低预期回报、高估风险,产生更多的卖出行为(Loewenstein等,2001)。另一方面,投资者关注也受市场状态的影响(权小锋等,2012)。投资者在市场低迷时期查看投资组合的频率远低于市场繁荣时期,表明投资者分配注意力时会表现出“鸵鸟效应”;与市场处于低迷状态时相比,投资者在市场上升阶段对股市表现出更程度的关注(Karlsson等,2009)。此外,市场环境还直接影响灰箱A内投资者网络信息搜索的广度、深度和强度。

对于灰箱A特别是其中投资者关注、投资者情绪及其他因素如何决定投资者偏好与决策,进而影响股票交易方向、数量及时机,最终作用于资产定价的整个过程(传导机制⑤),行为金融学文献中可找到大量支持(参见De Long等,1990;Thaler,2005;Engelberg等,2012等)。需要指出的是,灰箱A对投资者股票交易方向、数量及行为的影响方式是动态复杂的,如个体投资者的网络搜索反映更多的是个体投资者的买盘压力。当某个体投资者对某上市公司股票名称(或代码)进行网络检索时,这更可能预示着对陌生公司股票买入行为,而非对在手的熟知公司股票的卖出行为(Barber & Odean,2008)。当个体投资者购买股票时,希望通过网络搜索了解到该股票历史交易信息、人们对该股票(上市公司)的网络评论及相关传言。而这些信息特别是评论和传言并不能在股票交易平台上

完全体现。当个体投资者卖出股票时,他们依据更多的是以购买价格作为参照点的预期收益,此时网络搜索信息的需求远小于购买时。

事实上,资产定价走势反过来也会影响灰箱 A(传导机制⑥)。如股票市场定价走势就会引发投资者情绪或投资者关注的变化,由此会驱动投资者网络搜索行为的改变。网络搜索和股票市场之间可能存在循环影响与反馈机制。总之,基于相较于投资者情绪及投资者关注更综合、更直观、更易观测的投资者网络搜索,来预测股票市场更符合逻辑。

### 三、实证设计

#### (一)样本选择和数据来源

本文选择 2006 年 6 月到 2012 年 5 月时间段的沪深 300 指数 A 股公司作为研究样本。有三点考虑:(1)沪深 300 指数覆盖了沪深 A 股市场近六成市值,具有市场代表性,能反映中国证券市场股票价格变动的整体走势和运行状况。(2)中小盘股票的波动性大于沪深 300 指数的大盘股票,基于网络搜索的套利风险因此会变大,从而增加预测的不确定性。(3)作为相对成熟的股票,沪深 300 指数公司上市时间大都对应于 2006 年 6 月起始的百度指数。国内仅有的关于股票网络搜索文献基本上都是以创业板公司为研究对象,但考虑到创业板的推出时间较短、市值偏低,其代表性与沪深 300 指数相比不够充分,其较强的波动性还会带来预测困难等问题。

沪深 300 指数成分股的网络搜索数据来自于百度指数(<http://index.baidu.com/>)。作为中国大陆主流中文网络搜索引擎,百度有着最高的市场覆盖率,国内大多数投资者都使用它搜索相关信息。百度指数提供了从 2006 年 6 月至今主要关键字的搜索信息。本文使用用户关注度的统计指标,时间跨度从 2006 年 6 月 9 日至 2012 年 5 月 20 日。中国股市在这段时间里,恰好经历了相对完整的牛熊市周期。

本文对样本进行筛选:剔除 2006 年 6 月后上市的公司 85 个加数据存在缺失公司 1 个,共计 86 个,原有 300 个上市公司最后余 214 个(沪深 300 指数成分股每半年调整一次,本文使用 2012 年 1 月调整的样本股)。过长的时间窗口不利于考查投资者网络搜索强度对股票市场的真实影响,为提高研究准确性并与百度指数数据相匹配,本文选择周作为研究周期。在时间选择上,也删减掉因春节、国庆等主要假期所导致的网络搜索量统计与股票市场数据统计之间数据无法匹配的 19 周,保留对应的 292 周数据。最后获得 214 家上市公司 5 年 292 周共计 62488 个样本数据。本文实证所需的股市交易数据和财务数据,不作特别说明,均来自于 Wind 数据库,包括沪深 300 指数 A 股上市公司样本的周收益变动率、周交易量、周市值、季度所有者权益、月平均市盈率、月平均市净率等。

#### (二)变量说明

##### 1. 网络搜索

现有文献大多直接使用股票关键词搜索量或取对数的数据作为相应指标。这种绝对数处理方法存在着明显不足:(1)不同规模上市公司的网络搜索量可能有很大差异,而这

种差异与股票投资无关。(2)相对于搜索量绝对水平的高低,搜索量变动更值得重视。以绝对数处理方法进行股票间搜索量的比较,有时会产生误导。为解决上述问题,本文使用相对数处理方法,即将样本中每只股票所记录的每周搜索量除以样本期内每周平均搜索量,得到一个均值为 1 的序列:

$$SI_{i,t} = SV_{i,t} / \frac{1}{n} \sum_{i=1}^n SV_{i,t} \quad (1)$$

其中,  $SV_{i,t}$  (Search Volume) 表示百度指数提供的第  $i$  只股票第  $t$  周的网络搜索量。标准化后的数据  $SI_{i,t}$  (Search Intensity) 称之为网络搜索强度,若大于 1,说明网络搜索强度大于样本期的平均水平,反之亦然。本文重点观察网络搜索量变动所产生的影响,该方法并没有丢失这方面的信息,同时可以解决过去利用绝对数处理方法存在的两大缺陷。上述网络搜索量  $SV$  是以样本股票名称为关键词的搜索量。选择股票名称而不是股票代码作为关键词的理由在于,对于股龄短或经验少的新股民,他们大多数人可能无法记住有关上市公司的数字股票代码;对于股龄长或经验多的老股民,同样既无必要也不可能记住众多上市公司的股票代码。从实际操作来看,中国散户也更倾向于使用股票名称来进行搜索。在网络搜索强度  $SI$  的基础上,本文进一步构造出一个新指标:网络搜索强度正向变动率  $SIPC$  (Search Intensity Positive Change),其计算方法如下:

$$SIPC_t = \frac{1}{n} \sum_{i=1}^n I(SI_{i,t} - SI_{i,t-1}) \quad (2)$$

其中,  $n$  为组合中的股票数目;  $I(*)$  为示性函数,当  $*$   $> 0$  时  $I(*) = 1$ ,其余  $I(*) = 0$ 。 $SIPC_t$  表示第  $t$  周网络搜索强度相对于前一期上升的股票数占样本总数的比率。 $SIPC$  基本能够反映整个市场投资者网络搜索行为的变动趋势。

## 2. 投资者情绪

常用的股票市场投资者情绪指标主要包括封闭式基金折价、IPO 数量、IPO 上市首日收益、交易所每月新开户数等(易志高、茅宁,2009)。由于 IPO 在 2008 年 9 月 25 日至 2009 年 6 月 18 日有 9 个月的暂停, IPO 数量及 IPO 上市首日收益都不适于作为投资者情绪指标。本文选择封闭式基金折价及交易所每周新开户数作为投资者情绪的代理指标。封闭式基金折价 ( $DCEF$ ) 计算公式如下:

$$DCEF_t = \sum_{i=1}^n [(P_{i,t} - NAV_{i,t}) * N_i] / \sum_{i=1}^n (N_i * NAV_{i,t}) \quad (3)$$

其中,  $n$  为沪深市场在第  $t$  周公开发行的封闭式基金数量,  $P_{i,t}$  是  $i$  基金每周最后一个交易日的收盘价,  $NAV_{i,t}$  是每周最后一个交易日的  $i$  基金的单位净值,  $N_i$  是基金  $i$  的份额。市场上的封闭式基金大多以折价方式交易,折价小说明投资者情绪高涨,折价大意味着投资者情绪低落,因而  $DCEF$  通常为负且与投资者情绪正相关。

股票市场每周新开户数 ( $NA$ ),采用中登公司发布的《一周股票账户情况统计表》报告,选取每周新增 A 股开户数。出于统计考虑,取自然对数:

$$\ln NA_t = \ln(NA_t) \quad (4)$$

其中,  $NA_t$  为第  $t$  周新增 A 股开户数。  $LnNA$  也与投资者情绪正相关。

### 3. 投资者关注

传统的投资者关注度量指标主要包括超额收益、交易量、媒体报道和广告支出等(Da 等, 2011)。本文涉及沪深 300 指数, 考虑到媒体报道范围不容易界定, 广告支出主要针对具体公司, 故选择超额收益及交易量作为度量投资者关注的指标。超额收益率( $ARet$ )采用市场调整模型计算:

$$ARet_t = Ret_t - MRet_t \quad (5)$$

其中,  $Ret_t$  分别是第  $t$  周沪深 300 指数周收益率;  $MRet_t$  是第  $t$  周的市场收益率;  $ARet_t$  是超额收益率。参照伍燕然、韩立岩(2007)的研究, 本文使用上证综指收益率作为市场收益率。交易量( $TV$ )指标使用沪深 300 指数的交易量取自然对数:

$$LnTV_t = Ln(TV_t) \quad (6)$$

其中,  $TV_t$  为第  $t$  周沪深 300 指数的交易量。

### 4. 股票市场交易指标

本文使用股票市场交易指标数据主要是沪深 300 指数各成分股的周收盘价及交易量。这里定义累积收益率指标为:

$$r_{i,j,t} = (p_{i,t+j-1} - p_{i,t-1}) / p_{i,t-1} \quad (7)$$

其中,  $r_{i,j,t}$  表示第  $i$  只股票相对第  $t-1$  周累积  $j$  周的收益率,  $p_{i,t-1}$  表示第  $i$  只股票在  $t-1$  周的收盘价。此外, 定义类似  $SIPC$  的正收益比率  $PR$  (Positive Return), 计算方法为:

$$PR_t = \frac{1}{n} \sum_{i=1}^n I(ARet_{i,t}) \quad (8)$$

其中,  $n$  为组合中的股票数目;  $ARet_{i,t}$  为第  $i$  只股票在第  $t$  周的收益率;  $I(*)$  为示性函数,  $* > 0$  时  $I(*) = 1$ , 其余  $I(*) = 0$ 。  $PR$  越大说明样本中收益率为正的股票数目越多。基于沪深 300 成分股的代表性,  $PR$  基本可反映整个股票市场的变动方向。

由于各股票规模不同, 股票交易量之间可能存在巨大差异, 因而将交易量也进行消除量纲处理。本文采用如下公式计算交易量的变动, 称为异常交易量( $AV$ ):

$$AV_{i,t} = (V_{i,t} - V_{i,avg}) / V_{i,avg} \quad (9)$$

其中,  $V_{i,t}$  表示  $i$  股票在  $t$  周的交易量,  $V_{i,avg}$  表示  $i$  股票在整个样本期的平均交易量。这样处理所得数据可以反映交易量相对平均水平的变动, 使交易量规模差异明显的股票之间也具有可比性。

### (三) 检验模型

模型 1: 在研究  $SI$  对股票市场的影响时, 每周日对样本中各股票按照网络搜索强度  $SI$  进行递减排序, 将排序后股票分为 5 组, 前四组每组包含 43 个样本, 最后一组包含剩余 42 个样本。按搜索强度由高到低分组为 Q1 至 Q5。计算各组合股票下一周及以后的交易数据( $TD$ ), 取均值作为组合的交易数据, 并计算其在样本期的平均值。

$$TD_{Qi,t} = \frac{1}{n_{Qi,j \in Qi}} \sum TD_{j,t}, TD_{Qi} = \frac{1}{n} \sum_{t=1}^n TD_{Qi,t}, i = 1, \dots, 5 \quad (10)$$

其中,  $TD_{Qi,t}$  表示组合  $Q_i$  在第  $t$  周的交易数据,  $TD_{Qi}$  为组合  $Q_i$  在整个样本期交易数据的均值。在考察网络搜索强度对股票市场的影响时, 将相应的数据, 如短期累积收益率, 交易量等作为  $TD_{j,t}$  代入公式(10)算出  $TD_{Qi}$  即可。另外, 按照上述方法所构造的各组合股票在每周日根据网络搜索强度  $SI$  变动进行重新调整。

模型 2: 本文使用 Granger 因果检验及 AR 模型考察网络搜索与股票市场的相互作用, 并探讨内生性问题可能产生的影响。AR 模型的具体形式如下:

$$MRet_t = \beta_0 + \sum_{i=1}^n \beta_i MRet_{t-i} + \varepsilon_t$$

(11)

$$MRet_t = \beta_0 + \sum_{i=1}^n \beta_i MRet_{t-i} + \alpha_0 SIPC_{t-1} + \varepsilon_t$$

(12)

其中,  $MRet$  代表市场收益。公式(11)是利用历史收益信息解释当期市场收益的普通 AR 模型, 公式(12)增加滞后一期网络搜索指标考察网络搜索对市场收益的解释效果。

模型 3: 为比较网络搜索强度正向变动率  $SIPC$ 、投资者情绪与投资者关注对股票市场的解释及预测效力, 采用如下回归模型:

$$MRet_t = \beta_0 + \beta_1 SIPC_{t-1} + \beta_2 DCEF_t + \beta_3 LnNA_t + \beta_4 ARet_t + \beta_5 LnTV_t + \varepsilon_t$$

(13)

首先分别考察各指标单独作为自变量对因变量  $MRet$  的解释力; 其次分别将  $SIPC$  与投资者情绪指标,  $SIPC$  与投资者关注指标共同作为自变量对  $MRet$  进行回归; 最后包含所有自变量对  $MRet$  进行回归。在实证分析中, 根据需要对系数  $\beta_i$  进行选择, 在对应的变量不进入方程时, 取其为 0 即可。

四、实证结果与分析

(一) 网络搜索对股票市场的影响

1. 网络搜索强度对短期收益率

按照模型 1 思路, 将累积一周的收益率  $r_{i,1,t}$  代入公式(10)中计算, 结果见表 1。

表 1 网络搜索强度与短期收益率

组合	r	$\alpha$	$R_m - R_f$	SMB	HML	R 方
Q1	1.314	1.074*** (7.422)	1.270*** (37.770)	0.501*** (8.248)	-0.282*** (-5.388)	0.831
Q2	0.733	0.625*** (5.311)	1.135*** (41.496)	0.455*** (9.210)	-0.155*** (-3.623)	0.857
Q3	0.327	0.269** (2.399)	1.114*** (42.698)	0.555*** (11.764)	-0.14*** (-3.443)	0.866
Q4	0.289	0.195 (1.641)	1.040*** (37.654)	0.478*** (9.58)	-0.157*** (-3.657)	0.832
Q5	0.490	0.442*** (3.781)	1.021*** (38.189)	0.683*** (14.210)	-0.147*** (-3.311)	0.851
Q1 - Q5	0.824	0.621*** (5.084)	0.267*** (9.407)	-0.148*** (-2.879)	-0.126*** (-2.836)	0.290

注: 收益率单位为%; \*\*、\* 分别表示在 1%、5%、10% 的水平显著; 下同。括号内数据为 T 检验量。



表1中左边第二列 $r$ 表示未经调整的各股票组合在样本期的平均收益,第三列 $\alpha$ 为经过Fama-French三因素模型调整过的股票组合在样本期的平均收益。从Q1开始至Q5,无论是 $r$ 还是 $\alpha$ ,数据基本是呈现递减趋势,这意味着随着组合网络搜索强度的降低,下一周组合的周平均收益也随之降低。这种下降趋势在Q5组合发生逆转,可能源于投资者非持续的网络搜索行为与股票收益所具有的动量效应共同作用所致。

2. 网络搜索强度对短期交易量

按照模型1中的方法,将 $AV_{i,t}$ 代入公式(10)计算,具体结果见表2。表2第二行数据反映了从组合Q1到Q5,随着网络搜索强度降低,异常交易量 $AV$ 也对应降低。直接使用原始数据计算,如表2第三行及第四行所示,从Q1到Q5,随着搜索强度降低,各组合交易量的均值、中位数相应减少。这说明网络搜索强度与短期交易量存在正向关系。

表2 网络搜索强度和短期交易量

	Q1	Q2	Q3	Q4	Q5	Q1-Q5
AV	0.441	0.076	-0.100	-0.274	-0.284	0.726
均值	1.30E+08	99671487	85567161	73895896	61958147	68048874
中位数	1.22E+08	91404025	75723005	65937769	56375364	65580248

注:AV的单位为%。

3. 网络搜索强度对累积收益率

在网络搜索强度对短期收益的分析中,基本存在搜索强度高的组合的短期收益也较高的规律,那么这种规律在多长时间段内能够得到保持?使用模型1方法,分别计算各组合股票累积1,2,3,5,8周的收益率,即 $j=1,2,3,5,8$ 的 $r_{i,j,t}$ 。将模型1中的 $TD$ 分别用上述周数的累积收益率 $r_{i,j,t}$ 替换,计算各个股票组合在样本期的平均收益,结果见表3。

表3显示,第2周Q1对Q5还存在着0.335%的超额收益,尽管相对于第1周来说已经下降很多。Q1组累积3周的收益率已经小于Q5组,这种趋势随着时间的延长变得更加显著,Q1组累计8周的收益率已经比Q5组少了2.909%。表3还显示,网络搜索强度较大的Q1和Q2组合中,累积周平均收益率是递减的,而在网络搜索强度较小的Q3至Q5组合中,这些收益是递增的。随着时间的延长,网络搜索强度对股票短期累积收益的影响将由正转为负。

表3 网络搜索强度和累积收益率

组合	Q1		Q2		Q3		Q4		Q5		Q1-Q5	
	累积	平均	累积	平均	累积	平均	累积	平均	累积	平均	累积	平均
1周	1.314	1.314	0.733	0.733	0.327	0.327	0.289	0.289	0.49	0.49	0.824	0.824
2周	1.579	0.79	1.315	0.658	0.849	0.425	0.973	0.487	1.244	0.622	0.335	0.168
3周	1.929	0.643	1.809	0.603	1.499	0.5	1.787	0.596	2.143	0.715	-0.214	-0.071

续表

组合	Q1		Q2		Q3		Q4		Q5		Q1 - Q5	
	累积	平均	累积	平均	累积	平均	累积	平均	累积	平均	累积	平均
5 周	2.661	0.532	2.835	0.567	2.799	0.56	3.348	0.67	4.126	0.825	-1.465	-0.293
8 周	4.05	0.506	4.72	0.59	4.878	0.61	5.746	0.718	6.959	0.87	-2.909	-0.364

(二)网络搜索与股票市场关系的进一步分析

1. 网络搜索与股票市场的内生性

使用 Granger 因果检验以及 AR 模型讨论网络搜索与股票市场之间可能存在的内生性问题。Granger 因果检验需要确定时间序列的平稳性,经检验变量 *SIPC* 与 *MRet* 均为平稳,根据 AIC 等准则确定的滞后期为 3,使用软件 Eviews6.0 进行分析,所得实证结果见下表 4。表 4 中,在 5% 的显著性水平下,可以拒绝零假设,即两者互为 Granger 因果。在 1% 的显著性水平下,可以拒绝 *SIPC* 不是 *MRet* 的 Granger 因,不能拒绝 *MRet* 不是 *SIPC* 的 Granger 因。由此可知,*SIPC* 对 *MRet* 的影响要大于 *MRet* 对 *SIPC* 的影响。模型 2 中 AR 项滞后阶数仍通过 AIC 等准则确定,经检验滞后阶数应选 3,具体结果见表 5。

表 4 Granger 因果检验

Null Hypothesis:	F - Statistic	Prob.
SIPC does not Granger Cause MRet	55.417	0.000
MRet does not Granger Cause SIPC	4.240	0.040

表 5 AR 模型检验

MRet	C	MRet(-1)	MRet(-2)	MRet(-3)	SIPC(-1)	调整 R 方
回归(1)	0.119(0.623)	0.012(0.839)	0.162(0.006)	0.071(0.226)		0.022
回归(2)	-4.023(0.000)	-0.047(0.393)	0.108(0.047)	0.075(0.168)	7.114(0.000)	0.167

注:括号中的数字为该变量在回归方程中的显著性水平,下同。

回归(1)(公式 11)和回归(2)中(公式 12)结果显示,在包含了市场收益的历史信息之后,网络搜索指标 *SIPC*(-1)在 1% 的水平下显著,并且在相当大程度上提高了回归方程对市场收益的解释力,使得调整 R 方从 0.022 提高到 0.167。该实证结果充分说明:股票市场某种程度上可以影响网络搜索,但是网络搜索包含了更为丰富的信息,其本身可以在更大程度上影响并据此预测股票市场的表现。同时,网络搜索与股票收益之间存在的内生性问题对预测效果的影响很小。

2. 网络搜索预测股票市场的模拟分析

网络搜索强度最高的股票组合 Q1 与市场(上证综指)组合的每周收益率差额,在整个样本期 292 周中有 203 周为正;且在整个样本期中,Q1 组合对市场组合的超额收益达到平均每周 1.165%。这说明按照网络搜索强度构造股票投资组合,可获得显著高于市

场的超额收益,网络搜索强度能够作为预测股票市场的硬指标。

### (三)网络搜索、投资者情绪与投资者关注三种指标对股票市场解释力之比较

#### 1. 网络搜索指标与正收益比率

考察反映搜索强度变动的指标  $SIPC$  与正收益比率  $PR$  的关系。对某只股票来讲,其网络搜索强度上升意味着该股票会面临更大的买盘压力,进而推动其股价上升。可以推断,在整个股票组合中, $SIPC$  和  $PR$  之间应该存在着正向的关系。需强调的是,股票收益率在每周五就可以获得,而对其网络搜索量的统计则在周日进行。由于投资者的网络搜索行为要先于其对股票的交易行为,根据这种先行-滞后关系可以推断:本期的  $SIPC$  与下期的  $PR$  应有更强的相关关系。实证研究结果见表6。

表6 网络搜索指标与正收益比率

	$SIPC$	$SIPC(-1)$	$C$	调整 $R$ 方
$PR$ (相关性)	0.158*** (0.007)	0.362*** (0.000)		
$PR$ (回归)	0.211*** (0.007)		0.419*** (0.000)	0.022
$PR$ (回归)		0.481*** (0.000)	0.317*** (0.000)	0.128

注:括号内的数值为显著性水平,下同。

表6中使用 Pearson 相关性检验发现,同期的  $SIPC$  和  $PR$  之间的相关系数为 0.158,而  $SIPC(-1)$  与  $PR$  之间的相关系数为 0.362。在分别将  $SIPC$  和  $SIPC(-1)$  对  $PR$  做回归时,以  $SIPC$  为自变量回归所得的经调整的  $R$  方为 0.022,远小于以  $SIPC(-1)$  为自变量回归所得的 0.128,这说明  $SIPC(-1)$  对  $PR$  有着更强的解释力。结果佐证了本文推断: $SIPC$  和  $PR$  之间存在着正向相关关系,本期的  $SIPC$  与下期的  $PR$  有着更强的相关关系,这意味着  $SIPC$  对股票市场具有预测效力。

#### 2. 网络搜索、投资者情绪与投资者关注的相关性

接着考察  $SIPC$  与投资者情绪的代理指标封闭式基金折价  $DCEF$  和交易所每周新开户数  $LnNA$  以及投资者关注的代理指标超额收益率  $ARet$  和调整的交易量  $LnTV$  之间的相关性。同样基于投资者的网络搜索行为要先于对股票的交易行为这一逻辑,推断本期的  $SIPC$  与下一期的投资者情绪及投资者关注指标相比同期具有更强的相关性。使用 Pearson 相关性分析的实证结果见表7。

表7 网络搜索、投资者情绪与投资者关注的相关性

相关性	$DCEF$	$LnNA$	$ARet$	$LnTV$
$SIPC$	0.127** (0.03)	-0.113* (0.054)	0.071 (0.228)	-0.176*** (0.003)
$SIPC(-1)$	0.115* (0.050)	0.185*** (0.002)	0.112* (0.057)	0.358*** (0.000)

表7结果说明投资者网络搜索强度变化混合蕴含着投资者情绪和投资者关注,本文构造的投资者网络搜索强度指标领先于传统的投资者情绪指标和投资者关注指标。

3. 网络搜索、投资者情绪与投资者关注的解释力比较

进一步考察 *SIPC* 相对于投资者情绪指标和投资者关注指标对股票市场的解释力。具体统计分析前,可通过简单的作图进行更直观的比照。从图 2 至图 4<sup>①</sup> 容易发现,相对于投资者情绪指标以及投资者关注指标,*SIPC* 的曲线与市场收益曲线的基本保持着一致性变动趋势,特别是 2009 年之后两者的变动几乎重合。

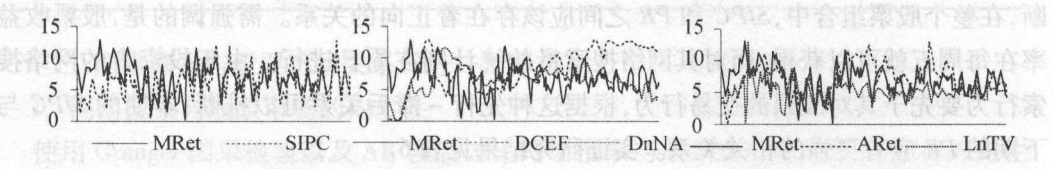


图 2 市场收益与网络搜索 图 3 市场收益与投资者情绪 图 4 市场收益与投资者关注

下面对上述直观发现基于模型 3 来数量化检验:以市场收益 *MRet* 为因变量,以网络搜索指标 *SIPC*、投资者情绪指标、投资者关注指标为因变量分别进行回归。基于上文的分析,直接使用滞后一期的 *SIPC*(-1)进入方程,结果见表 8。回归 1-5 说明,*SIPC*(-1)相比投资者情绪及投资者关注指标,对市场收益有着更高的解释力。回归 6-8 说明,除投资者情绪和关注以外,*SIPC* 包含了更丰富的内容。

表 8 网络搜索、投资者情绪与投资者关注的解释力比较

MRet	回归 1	回归 2	回归 3	回归 4	回归 5	回归 6	回归 7	回归 8
SIPC(-1)	6.977 *** (0.000)					7.369 *** (0.000)	5.534 *** (0.000)	5.631 *** (0.000)
DCEF		-2.175 (0.332)				-3.661 * (0.083)		-4.249 ** (0.037)
LnNA			0.063 (0.836)			-0.254 (0.381)		
ARet				0.9 *** (0.000)			0.718 *** (0.000)	0.710 *** (0.000)
LnTV					2.281 *** (0.000)		1.228 ** (0.011)	1.346 *** (0.005)
C	-2.495 *** (0.000)	-0.28 (0.594)	-0.613 (0.872)	0.066 (0.782)	-54.57 *** (0.000)	-0.246 (0.946)	-31.51 *** (0.006)	-35.24 *** (0.002)
调整 R 方	0.126	0	0	0.057	0.071	.0.132	0.181	0.19

① 为消除原变量序列的过度波动,便于观察变量的变动趋势,此处使用移动平均法(期数为 4)进行平滑处理;同时为方便各变量在图形中进行比较,考虑将各变量等比例映射到一个相同区间。

#### (四) 稳健性检验

为保证研究结论的可靠性,本文对实证方法进行多种方式的调整。结果显示,这些敏感性分析不会对本文结论产生实质性的影响,说明本文的结果具有较好的稳健性。首先,在网络搜索强度与股票市场指数的实证部分主要对样本期划分以及分组方式进行了调整:(1)将总的样本期结合牛熊周期划分为5个小样本期;(2)将样本分组方式由原先的5组细化为10组。其次,在三种指标对股票市场的解释力比较部分,主要对市场收益、投资者情绪等指标进行滞后期选择进行了调整:(1)市场收益率改用深证综指计算;(2)将网络搜索、投资者情绪及投资者关注指标均提前1期以检验其预测效力。

### 五、结论与启示

本文实证研究发现:(1)投资者的网络搜索强度对股票的短期收益率、短期交易量以及累积收益率均有影响。(2)股票市场某种程度上可以影响网络搜索,但是网络搜索能在更大程度上影响并据此预测股票市场的表现,网络搜索与股票收益之间存在的内生性问题对预测效果的影响很小。根据网络搜索构建组合可以获取超额收益,利用网络搜索能够预测股价走势。(3)投资者网络搜索强度指标与传统的投资者情绪指标和投资者关注指标之间存在较强的相关关系;且相较于投资者情绪和投资者关注,投资者网络搜索对股票市场的解释力及预测效力更强。这说明,复合了投资者情绪及投资者关注等众多因素的网络搜索指标更适合于作为预测股票市场的—个重要指标。

本文研究有着重要的实践启示。考虑到国内投资者基于网络搜索获取信息的普遍性及利用网络搜索预测股票市场的有效性,中国的股票投资者和证券监管部门都应“与时俱进”地从更新颖的视角考察并重视网络搜索在股票市场中所起到的作用。网络搜索引擎提供的数值精确、信息量大、更新快、易获取,个体投资者使用网络搜索的强大功能可以提升自身的信息获取能力,同时也能够从网络搜索所蕴含的行为金融逻辑中找寻“跑赢市场”的获利机会,通过网络搜索强度构造独特的股票组合攫取超额收益,进而提高股票投资绩效。机构投资者则需将个体投资者网络搜索指标作为其预测散户行为动态、改善资产组合的重要参考指标之一。而对于证券监管部门而言,若能挖掘、捕捉并监测好投资者网络搜索指标的话,则能够改进其股票市场监管的技术手段,优化市场监管绩效。

### 参考文献

- [1]贾春新、赵宇、孙萌、汪博,2010,《投资者有限关注与限售股解禁》,《金融研究》第11期108~122页。
- [2]权小锋、洪涛、吴世农,2012,《选择性关注、鸵鸟效应与市场异象》,《金融研究》第3期109~123页。
- [3]谭松涛、陈玉宇,2012,《投资经验能够改善股民的收益状况吗》,《金融研究》第5期164~178页。
- [4]伍燕然、韩立岩,2007,《不完全理性、投资者情绪与封闭式基金之谜》,《经济研究》第3期117~129页。
- [5]易志高、茅宁,2009,《投资者情绪测量实证研究:CICSI的构建》,《金融研究》第11期174~184页。
- [6]Baker, M., Wurgler, J., 2006, “Investor sentiment and the cross section of stock returns”, *Journal of Finance*, 61;

1645 ~ 1680.

- [7] Barber, B. M., Odean, T., 2008, "All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors", *Review of Financial Studies*, 21: 785 ~ 818.
- [8] Da Z., Engelberg, J., Gao P., 2011, "In Search of Attention", *The Journal of Finance*, 66(5): 1461 ~ 1499.
- [9] Engelberg J., Sasseville C., Williams J., 2012, "Market madness? The case of mad money", *Management Science*, 58(2): 351 ~ 364.
- [10] Kissan Joseph, M. Babajide Wintoki, Zelin Zhang, 2011, "Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search", *International Journal of Forecasting*, 27: 1116 ~ 1127.
- [11] Richard Thaler, 2005, *Advances in Behavioral Finance*, Oxford Press.

## Can Internet Search Predict the Stock Market?

ZHANG Yihao LI Yuan SU Zhongfeng ZHANG Zelin

(Business School, Nanjing University; Business School, Renmin University of China)

**Abstract:** An intrinsic mechanism is suggested to exist among the interactions between the individual trader's internet search behavior and the target stock's market performance. Individual traders' internet search behavior on the stock symbol will substantially influence the intertemporal stock price. We find that: (1) Internet search intensity can predict the stock's short-run return, trading volume and cumulative return, which implies a positive relationship between internet search and stock market. (2) Internet search intensity can affect stock market further and predict stock market effectively. Moreover, a simulation analysis reveals that a stock portfolio constructed according to the internet search will yield excess abnormal return. (3) Internet search intensity index is strongly correlated with traditional investor sentiment index and investor attention index, and is more powerful to explain and predict stock market.

**Key words:** Internet search, Investor sentiment, Investor attention, Stock market

(责任编辑:杨启庸)(校对:YY)