

行业关注度与股票横截面收益率

——基于百度行业搜索指数的研究

张学勇 唐国梅*

摘要 本文以行业名称为关键词爬取百度搜索指数作为行业关注度的代理变量,研究投资者行业关注度与股票横截面收益率的关系。结果表明,行业搜索量能正向预测下一天的股票收益率,基于这一指标构建的多空策略每年可获得 4.89% (4.28%) 的等权 (市值加权) 收益率。与投资者关注度假说一致,行业关注度导致的价格压力效应在卖空约束强的股票中更突出,且长期来看价格压力会消退。进一步的研究表明,行业关注度对股价的预测能力强于个股关注度。

关键词 百度指数, 行业, 有限注意力

DOI: 10.13821/j.cnki.ceq.2022.03.03

一、引言

传统资产定价模型通常假设资产价格能够迅速并充分反映市场上的新信息。然而,这一假设的前提是投资者能够密切关注所有信息,并对信息进行及时有效的处理。近年来,众多研究表明,投资者的注意力和信息处理能力都是有限的 (Hirshleifer and Teoh, 2003; 冯旭南, 2017)。随着互联网技术的发展,我们正处于一个信息爆炸的时代,投资者如何分配有限的注意力,有限关注又如何影响其决策行为和资产价格更成为一个重要的研究议题。根植于行为金融学,国内外学者提出了各种关注度指标,为关注度与资产定价关系的研究提供了理论和实证基础 (如: Barber and Odean, 2008; Da *et al.*, 2011; 俞庆进和张兵, 2012; 杨晓兰等, 2016; 杨涛和郭萌萌, 2019)。然而,这些研究大多是从投资者对个股层面的关注度展开,尚未有学者探究投资者的行业关注度的经济影响。

* 张学勇、唐国梅, 中央财经大学金融学院。通信作者及地址: 张学勇, 北京市海淀区学院南路 39 号中央财经大学, 100081; 电话: 13701249262; E-mail: zhangxueyong@cufe.edu.cn。本文感谢国家自然科学基金重大项目 (19ZDA098)、国家自然科学基金项目 (71773152, 71673318)、北京高精尖学科项目 (GJJ2019163), 以及金融科技、金融创新与经济高质量发展研究基地的资助。感谢两位匿名评审专家的建议。

行业作为自上而下的资产配置策略中的关键一环,其本身就存在着重要的研究价值。行业不仅解释了很多个股现象¹,还是一个重要的投资分类。风格投资(style investing)的文献表明,投资者倾向于将股票归为不同的类别(如价值股、成长股、科技股),并在类别而非个股层面上配置资金(Barberis and Shleifer, 2003; Froot and Teo, 2008)。而行业就是一个很直观的投资类别,因为同一个行业的股票在业务模式、资本结构等方面存在很多相似性(Jame and Tong, 2014; Drake *et al.*, 2017)。

此外,从中国实际来看,行业也一直受到政策制定者、金融从业者以及股票投资者的广泛关注。政策层面,国家发展与改革委员会经常以行业为单位制定产业政策,如2009年密集发布了关于多个行业的调整和振兴规划。国务院在2012、2015和2016年也分别发布了《“十二五”国家战略性新兴产业发展规划》《中国制造2025》和《“十三五”国家战略性新兴产业发展规划》等行业发展规划文件。投资层面,行业早已成为市场参与者的重要分析维度。东方财富Choice金融终端研究报告板块数据显示,仅2011—2020年十年间,市场中的行业研报数量就高达233 101份。此外,基于行业的投资策略也占据了基金市场的一定份额,过去十年间平均每年有10只左右新发行的基金是基于行业策略的。²

既然行业在资本市场中扮演着如此重要的角色,一个很自然的问题就是投资者的行业关注度会如何影响资产价格?更进一步地,行业关注度与个股关注度的作用孰强孰弱呢?这是本文主要探究的问题。我国股市有两个突出的特征,一是信息环境相对较差且市场由散户主导(Jones *et al.*, 2020),而散户由于注意力更为有限,因此更容易将股票按照行业来分类分析(Jame and Tong, 2014);二是鲜明的题材转换和行业轮动特征(胡聪慧等, 2015)。因此,探究行业关注度对股票收益率的影响,对于理解我国股市的运作逻辑及定价机理具有重要意义。

借助个股关注度的研究框架,本文以110个中信证券行业分类标准下的三级行业的行业名称为关键词,爬取各个行业的搜索量(SVI^{IND})作为行业关注度的代理变量。参考Da *et al.* (2015)的研究,我们构建了异常搜索量指标(ASVI^{IND}),发现ASVI^{IND}能够正向预测下一天的股票收益率。后续分析表明,我们的结果与Barber and Odean (2008)的关注度假说是一致的,行业关注度导致的价格压力效应在卖空约束强的股票中更突出,并且会在接下来几天反转。此外,我们发现由于行业关注度领先于个股关注度,因此相比个股关注度而言具有更强的股价预测能力。

¹ 如动量效应(Moskowitz and Grinblatt, 1999)、股票价格(Hou and Robinson, 2006)、公司财务结构(Mackay and Phillips, 2005),以及很多的资产定价异象(Chou *et al.*, 2012)等。

² 国泰安数据库基金主体信息表中收录了每年新发行的基金的基本情况,本文将名称中包含“行业”的基金视作基于行业策略的基金。

本文主要有两方面的贡献。首先，本文拓展了投资者关注度与资产价格关系的研究范围。以前的研究主要是针对个股关注度。³然而，事实上个股获得的关注度中有很很大一部分是由行业和市场因素驱动的 (Drake *et al.*, 2017)。另一方面，注意力有限的投资者本身就会表现出类别思考行为，即分配更多的注意力到市场或者行业层面的信息上 (Peng and Xiong, 2006)。这些研究证据都表明行业关注度的经济含义值得更深入的研究，因而我们的研究具有重要的理论意义。本文不仅探究了行业关注度对资产定价的影响，还考察了行业关注度与个股关注度在中国市场的相对重要性，既是对现有文献的一个补充，也是一个突破。

其次，本文也补充了行业特征对资产定价的影响的文献。Moskowitz and Grinblatt (1999) 研究了行业对动量效应的解释能力，Chou *et al.* (2012) 探究了行业能否解释股票收益率和资产定价异象，Hou and Robinson (2006) 则研究了行业集中度对股票横截面收益率的影响。我们从行业关注度与股票横截面收益率这一独特视角对该领域的文献进行了补充。此外，众多文献都表明投资者存在行业追逐行为 (industry herding)，即投资者倾向于追逐“热门”行业 (Choi and Sias, 2009)。可是如何判断一个行业是热门还是冷门呢？本文构建的行业关注度指标是一个重要参考。

本文剩余部分安排如下：第二部分为理论分析与研究假说；第三部分介绍行业关注度指标及其他变量的构建方法；第四部分实证检验行业关注度与股票横截面收益率的关系，并分析结果是否与 Barber and Odean (2008) 的关注度假说一致；第五部分为稳健性检验；第六部分总结全文。

二、理论分析与研究假说

有限关注的一个直接影响是投资者会表现出注意力驱动购买行为，进而推动股价的短期上涨，这种现象在散户主导的市场尤为突出。Barber and Odean (2008) 提出的关注度假说从理论和实证上对此进行了证实。由于市场中股票数量巨大，投资者在购买股票时面临艰巨的选择问题。解决选择困难的方法之一就是只在那些吸引他们注意力的股票中选股。另一方面，虽然理论上投资者在进行卖出决策时面临相同的选择问题，但是现实中大多数投资者的投资组合里面都只包含少量股票，即使市场不存在卖空约束，投资者通常也只会选择卖出他们已经持有的股票。因此，综合起来，投资者倾向于购买吸引注意力的股票，而注意力对卖出决策的影响则较小，这导致投资者成为吸引注意力的股票的净买入者，这种买入行为会对股价造成短暂的上涨压力。利用投资者的交易数据

³ 少部分学者研究了市场关注度，如 Gilbert *et al.* (2012)、Li and Yu (2012)、Vozlyublennai (2014)、Yuan (2015)。

构建净买入指标, Barber and Odean (2008) 验证了该假说。

我们认为, 同样的逻辑对行业关注度也是成立的。无论是个股还是行业选择, 投资者都同样面临选择困难, 将其有限的注意力聚焦到部分引人注目的行业上, 依然是简化投资决策的一个方式。因此, 当投资者对某个行业的关注度增加时, 投资者会对该行业的股票形成购买压力, 进而推动该行业股价的上涨。据此我们提出本文的第一个研究假说:

假说 1 行业关注度能正向预测该行业的股票收益率。

在 Barber and Odean (2008) 的关注度假说框架下, 除了关注度与股票收益率的正向关系外, 还有两个重要的实证内涵。第一个是价格压力在卖空约束强的股票中会更为突出。理论上, 关注度不仅能影响投资者的买入行为, 也会影响卖出决策, 买卖双方力量的相对强弱决定了价格的变化方向及大小。然而, 现实中关注度更多影响的是买方的交易行为, 这是因为在全球大部分市场, 股票的买入和卖出条件并非对等。对于买方而言, 只需在市场中买入关注度高的标的即可。而对于卖方而言, 卖出关注度高的股票的前提是持有这些股票, 满足该条件的卖方数量远远小于买方数量。即使悲观投资者可以做空高关注度的股票, 卖空也存在诸多约束。以我国为例, 2010 年以前 A 股市场禁止融券操作, 这在机制上就阻碍了卖方力量。虽然目前已逐步放开融券限制, 但由于卖空操作涉及融券费用等直接成本, 以及更高的投资风险等间接成本 (Jones and Lamont, 2002), 在我国卖空约束依然普遍存在 (包锋和徐建国, 2015)。当投资者关注度上升时, 卖空约束强的股票由于卖方力量更弱, 会有更强的价格压力。据此我们提出假说 2:

假说 2 行业关注度导致的价格压力效应在卖空约束强的股票中更为突出。

关注度假说的第二个实证内涵是如果股价的上涨是由于对吸引注意力的股票的净买入压力导致的, 那么这种价格上涨会出现反转; 相反, 如果是由基本面驱动的, 则不会出现反转效应 (Barber and Odean, 2008; Da *et al.*, 2011)。关注度领域的众多文献都对此进行了检验 (如: Seasholes and Wu, 2007; Vozlyublenniaia, 2014)。特别的, 作为首次采用搜索指数来度量个人投资者关注度的研究, Da *et al.* (2011) 发现谷歌搜索指数导致的价格上涨压力会在一年内反转。俞庆进和张兵 (2012) 基于中国市场的研究也发现, 百度指数代表的个股关注度所导致的正向价格压力很快会发生反转, 这是因为之前的高搜索量带来的溢价缺乏基本面的支撑, 一旦投资者的关注度转移到其他地方, 这种上涨将无法持续。因而, 我们预期行业搜索量导致的价格上涨也会在短期内反转, 由此提出假说 3:

假说 3 行业关注度导致的价格压力效应会反转。

过去大量的文献表明个股关注度对股票收益率有显著影响, 自然就会引发究竟是个股关注度还是行业关注度对股价的影响更大的疑问。Drake *et al.* (2017) 基于美国市场的研究表明, 行业和市场关注度解释了大约 1/5 的个股关

关注度，表明个股获得的关注度中有很大大一部分是与行业和市场获得的关注度相关的。然而，我们认为行业关注度在中国市场中的地位可能会更为突出。首先，从市场环境来看，相比于发达经济体而言，中国股市的信息不对称问题更严重(Chan *et al.*, 2008)，这导致信息搜寻的成本更高，降低了投资者挖掘和处理信息的积极性(Chan and Hameed, 2006)，从而更加依赖于行业信息来决策。在极端情况下，投资者甚至会完全忽略个股信息，而只关注市场或者行业层面的信息(Peng and Xiong, 2006)。其次，从投资者结构来看，中国股市是散户主导的市场，沪市约85%的交易量是由散户贡献的(Jones *et al.*, 2020)。与专业投资者相比，个人投资者在认知和注意力资源上都处于劣势，因此更容易将股票按照行业来分类分析(Jame and Tong, 2014)。最后，从信息传递的角度，国内很多政策都是针对行业颁布的，随后个股才对这些政策做出反应，这意味着行业层面的信息会在个股信息到来之前捕获投资者的注意力。

基于以上分析，本文认为在我国特殊的信息环境和散户主导的投资者结构下，行业在投资决策中扮演着更为重要的角色，且行业信息往往又领先于个股信息，因而行业关注度对股价的影响会强于个股关注度。据此我们提出假说4：

假说4 行业关注度相比个股关注度对股票收益率的影响更强。

三、数据及变量定义

(一) 搜索量数据及样本构建

本文的搜索量数据来自百度指数⁴。百度提供了来自电脑端和移动端的搜索数据，前者最早追溯到2006年6月5日，后者则为2011年1月1日。考虑到智能手机的推广，我们采用电脑端与移动端的搜索量之和作为百度指数，从而样本区间为2011年1月至2019年8月。

为了度量行业关注度，我们以中信证券行业分类标准下的三级行业的行业名称为关键词，搜索并爬取了各个行业的日度搜索量。中信三级行业包含189个行业，在此基础上，我们剔除了以下行业：(1) 7个金融行业，这是考虑到金融股的高杠杆(Fama and French, 1992)；(2) 综合行业，因为行业名称太宽泛；(3) 71个行业名称或者其很直接的同义词没有被百度指数收录的行业。⁵最终，我们的样本为2011年1月至2019年8月来自110个中信三级行业的2208只A股股票。

⁴ <http://index.baidu.com/v2/index.html>，访问时间：2019年8月17日。

⁵ 在我们的110个样本行业中，有7个行业的搜索关键词与行业名称不完全一致。具体的，我们将“电影动画”行业的搜索关键词替换为“电影动漫”，“日用化学品”替换为“日化用品”，“石油开采”替换为“采油”，“电信运营”替换为“通信运营”，“系统集成及IT咨询”拆分为“系统集成”+“IT咨询”。另外，媒体行业下有一个“互联网”子行业，考虑到互联网更像一个IT行业的子行业，我们将该行业的搜索关键词调整为“互联网媒体”。限于篇幅，具体行业及其搜索关键词未放入正文，可来函索取。

(二) 异常关注度指标

表1的Panel A展示了各个行业的日度搜索指数 SVI^{IND} 的描述性统计。表格中汇报的数值为对应时间段内各变量的时间序列均值。可以看到,行业搜索量数据是高度右偏的:在样本期内, SVI^{IND} 的均值为904.40,为中位数的两倍多。此外,横截面标准差的均值高达1506.24,表明不同行业间的搜索量存在巨大差别。

为了解决这一问题,我们参考Da *et al.* (2015)的方法,对搜索量数据进行如下处理:首先,我们计算 t 天的搜索量相对于其过去7天的搜索量均值的对数差值为

$$\Delta SVI_{i,t}^{IND} = \log(1 + SVI_{i,t}^{IND}) - \log[1 + \text{Mean}(SVI_{i,t-1}^{IND}, \dots, SVI_{i,t-7}^{IND})]. \quad (1)$$

然后,我们将 $\Delta SVI_{i,t}^{IND}$ 进行1% (首尾各0.5%)水平的缩尾处理。接着,为了消除行业效应,我们将 $\Delta SVI_{i,t}^{IND}$ 对行业虚拟变量回归,并保留残差。最后,将残差序列标准化,得到最终的异常搜索量 ($ASVI^{IND}$) 指标。

表1的Panel B展示了 $ASVI^{IND}$ 的描述性统计结果。可以看到,调整后的指标偏度显著降低。更为重要的是,与原始的搜索数据相比, $ASVI^{IND}$ 捕捉到了投资者对某个行业的关注度的增量,并且该指标在不同行业间是可比的。

表1 描述性统计

年份	行业数量	均值	标准差	最大值	最小值	25%分位数	50%分位数	75%分位数
Panel A: 行业搜索量 (SVI^{IND})								
2011	109	736.78	1 079.81	8 129.19	5.33	213.53	397.89	870.53
2012	109	912.77	1 468.83	9 745.97	8.22	221.48	415.14	979.27
2013	109	952.52	1 977.06	15 848.1	18.94	210.40	396.59	903.28
2014	109	964.19	1 798.86	13 513.9	14.53	212.90	420.70	1 034.93
2015	109	1 051.87	1 743.03	13 799.3	5.86	284.62	527.94	1 231.80
2016	109	937.94	1 490.40	12 187.9	8.70	263.98	466.00	1 172.74
2017	110	901.40	1 367.74	10 762.4	4.63	244.04	465.56	1 208.49
2018	110	872.38	1 400.20	11 178.8	4.09	238.32	463.87	1 166.85
2019	110	762.09	1 091.87	8 338.30	10.88	228.34	412.41	1 056.60
所有年	109	904.40	1 506.24	11 622.0	8.95	235.56	441.77	1 069.88
Panel B: 异常行业搜索量 ($ASVI^{IND}$)								
2011	109	0.01	0.78	2.70	-2.90	-0.37	0.02	0.43
2012	109	-0.01	0.80	2.85	-2.95	-0.39	-0.00	0.41
2013	109	0.00	0.77	2.62	-2.81	-0.38	0.01	0.42

(续表)

年份	行业数量	均值	标准差	最大值	最小值	25%分位数	50%分位数	75%分位数
Panel B: 异常行业搜索量 (ASVI ^{IND})								
2014	109	0.01	0.81	2.71	-3.38	-0.37	0.02	0.42
2015	109	0.02	0.72	2.31	-2.64	-0.33	0.03	0.40
2016	109	0.00	0.72	2.38	-2.66	-0.37	0.01	0.39
2017	110	-0.01	0.73	2.36	-3.02	-0.36	0.00	0.38
2018	110	-0.01	0.80	2.52	-3.52	-0.38	0.02	0.41
2019	110	-0.01	0.84	2.59	-3.59	-0.41	0.01	0.44
所有年	109	0.00	0.77	2.56	-3.03	-0.37	0.01	0.41

注：样本区间为2011年1月至2019年8月，行业由中信三级行业分类标准定义。

(三) 其他变量

除了ASVI^{IND}以外，本文还涉及很多变量。财务数据和股票收益率数据来源于国泰安数据库，所有财务变量都在1%水平上缩尾，以避免极端值的影响。

首先，我们控制影响股票横截面收益率的公司特征变量。参考Jiang *et al.* (2021)，市值 (*Size*) 定义为收盘价与总股数的乘积的自然对数，数值每日更新。账面市值比 (*B/M*) 为账面价值与市值的比值，其中市值每日更新，*t* 年7月到 *t*+1 年6月的账面价值则是使用 *t*-1 年年末的账面价值，账面价值等于总权益减去优先股权益。动量 (*Mom*) 为过去 *t*-252 到 *t*-21 个交易日的累计收益率。价格 (*Price*) 为上一个交易日的收盘价。

另外，参考Barber and Odean (2008) 和 Da *et al.* (2011)，本文也引入了几个个股关注度指标。*Ret* 为上一个交易日的收益率，单位为%。*Abn_volume* 为分组日的交易量与过去一年（即，252个交易日）的平均交易量之比。类似的，*Abn_turnover* 为分组日的换手率与过去一年的平均换手率之比。*Analyst* 为1加上一年度跟踪某公司的分析师（团队）数量的自然对数。ASVI^{STK} 为以股票代码为搜索关键词的个股异常搜索量指标，计算方法与ASVI^{IND} 类似，即对于股票 *j*，先计算 *t* 天的搜索量相对于其过去7天的搜索量均值的对数差值为

$$\Delta SVI_{j,t}^{STK} = \log(1 + SVI_{j,t}^{STK}) - \log[1 + \text{Mean}(SVI_{j,t-1}^{STK}, \dots, SVI_{j,t-7}^{STK})], \quad (2)$$

然后对 $\Delta SVI_{j,t}^{STK}$ 进行首尾各0.5%水平的缩尾，再将 $\Delta SVI_{j,t}^{STK}$ 对股票虚拟变量回归并保留残差。最后将残差序列标准化，得到最终的异常搜索量 (ASVI^{STK}) 指标。

最后,为了检验关注度效应是否在卖空约束强的股票中更突出,我们引入两个卖空约束代理变量。*Ins_ownership*为上一季度的机构投资者持股比例,数值越小则卖空约束越强。*Short_available*为融券标的股标识,能够融券则记为“是”,否则为“否”,非融券标的股有更强的卖空约束。

四、实证结果

(一) 行业关注度与股票收益率

1. 单变量组合分析

本文首先采用单变量排序方法检验 $ASVI^{IND}$ 与股票横截面收益率的关系。每天我们将样本中的行业按照 $ASVI^{IND}$ 的 30% 和 70% 分位数分为低、中、高三组,然后将个股按照其所处的行业放到三个组合中,再计算下一天各个组合的收益率。组合收益率的计算方法参考 Hou and Robinson (2006),即先将同一个行业的股票收益率平均得到行业收益率,再将同一个组合内不同行业的收益率平均得到组合的收益率(单位为%)。

表2的Panel A和Panel B分别汇报了以等权重和总市值加权方式计算行业收益率时各组合的收益情况。第(1)列展示了三个组合的原始收益率。可见, $ASVI^{IND}$ 与股票收益率显著正相关:从低 $ASVI^{IND}$ 组到高 $ASVI^{IND}$ 组,组合的等权重收益率从每年 12.57% ($0.0499\% \times 252$) 单调递增至 17.46%,市值加权收益率从 7.01% 单调递增至 11.29%。做多高 $ASVI^{IND}$ 组股票同时做空低 $ASVI^{IND}$ 组股票这样一个策略能够获得每年 4.89% (4.28%) 的等权重(市值加权)收益率。

表2 单变量组合分析

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel A: 等权重					
低	0.0499 (1.13)	0.0298** (2.31)	0.0002 (0.03)	0.0009 (0.16)	0.0047 (0.80)
中	0.0566 (1.28)	0.0364*** (2.85)	0.0065 (1.16)	0.0074 (1.39)	0.0104* (1.89)
高	0.0693 (1.58)	0.0493*** (3.85)	0.0200*** (3.17)	0.0208*** (3.47)	0.0244*** (3.92)
高一低	0.0194*** (2.98)	0.0195*** (2.98)	0.0198*** (2.97)	0.0198*** (2.98)	0.0197*** (2.95)

(续表)

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel B: 市值加权					
低	0.0278 (0.68)	0.0085 (0.94)	-0.0087 (-1.45)	-0.0085 (-1.41)	-0.0050 (-0.83)
中	0.0320 (0.78)	0.0126 (1.49)	-0.0033 (-0.58)	-0.0029 (-0.51)	-0.0002 (-0.04)
高	0.0448 (1.10)	0.0256*** (2.90)	0.0097 (1.58)	0.0101* (1.65)	0.0134** (2.19)
高一低	0.0170** (2.35)	0.0171** (2.35)	0.0184** (2.48)	0.0185** (2.50)	0.0184** (2.47)

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

表 2 的余下几列检验了各组合的收益率能否被经典的风险因子解释。具体的，我们用组合收益率对 CAPM、Fama-French (1993) 三因子模型、Carhart (1997) 四因子模型和 Fama-French (2015) 五因子模型对应的因子回归⁶，检验截距项 (α) 是否显著。可以看到，在所有回归模型中，多空组合的 α 都是显著的，进一步表明基于 $ASVI^{IND}$ 的策略能获取剔除市场共同因子影响的超额收益率。并且多空组合的收益是源自多头收益率显著更高，这与 Barber and Odean (2008) 的关注度假说是一致的，即能够吸引注意力行业的股票有更高的收益率。由此，假说 1 得到初步验证。

为了更直观地分析多空组合的收益率，同时增强本文结论在实际股票市场中的应用性，图 1 绘制了基于行业关注度的多空组合以及沪深 300 指数的净值曲线。可以看到，L/S 策略在市场剧烈波动的情况下仍然保持了净值的持续增长，且回撤率远远小于市场指数策略。当然，在不同的市场热度情况下，行业关注度的效果也存在差异。具体表现为股灾前策略的净值增长更快，这从侧面表明行业搜索量代表了一种关注度以及情绪效应，这种效应在市场热度较高的时期更强。另外，整体来看等权重策略要优于市值加权策略，说明行业关注度策略在小市值股票中会更为有效，这与关注度假说是一致的 (Da *et al.*, 2011)。

⁶ 因子数据来源于中国资产管理研究中心，指标的具体计算方法见中心网站：<http://sf.cufe.edu.cn/info/1198/7881.htm>，访问时间：2019 年 9 月 20 日。

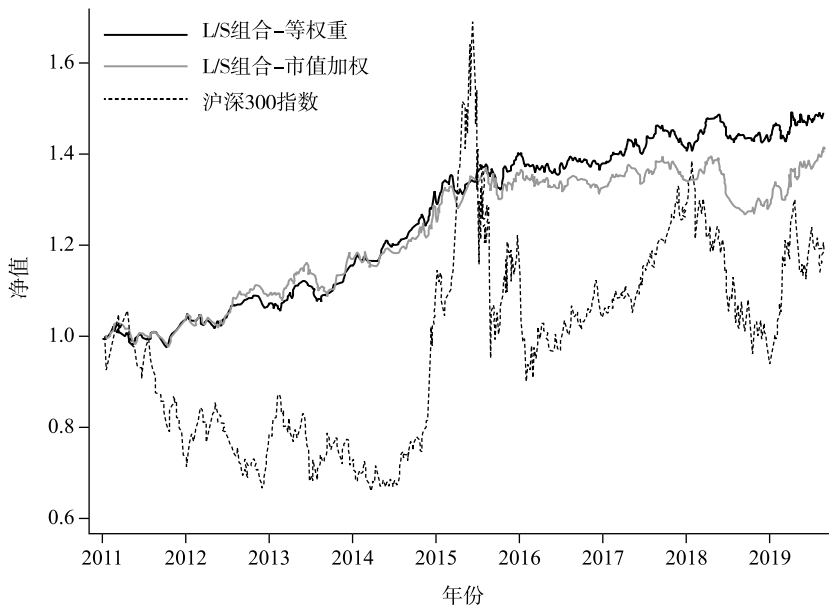


图1 基于行业关注度的策略的净值曲线

2. 双变量组合分析

接下来我们采用双变量排序分组方法, 检验 $ASVI^{IND}$ 与股票收益率的正相关关系在控制了个股特征和个股关注度之后是否依然显著。

(1) 控制个股特征。很多股票特征都会影响横截面收益率, 如规模和账面市值比 (Fama and French, 1992)、动量 (Jegadeesh and Titman, 1993)、股票价格 (Fritzemeier, 1936) 等。基于以上分析, 本文控制规模 (*Size*)、账面市值比 (*B/M*)、动量 (*Mom*)、收盘价 (*Price*) 四个股票特征变量。

每天我们按照给定个股特征变量的 30% 和 70% 分位数将样本中的股票分为三组, 然后在每个组内识别出唯一的行业, 并将行业按照 $ASVI^{IND}$ 的 30% 和 70% 分位数分为三组, 从而得到 9 个 (3×3) 个股-行业双变量分组的组合。同样的, 参考 Hou and Robinson (2006), 我们先将下一天同一个行业的股票收益率平均得到行业收益率⁷, 再将同一个组合内的行业收益率平均得到组合收益率 (单位为%)。

表 3 的结果表明, 在所有个股特征子样本中, 高 $ASVI^{IND}$ 组与低 $ASVI^{IND}$ 组的收益率之差均显著为正, 且无法被标准风险因子解释。值得注意的是, 小规模股票中多空组合年收益率为 5.52%, 而大规模股票中则只有 4.66%, 这与 Da *et al.* (2011) 的结果一致——注意力驱动的价格压力在小规模股票中更显著, 因为这些股票的价格更容易被交易量影响。

⁷ 这里仅展示了等权重的情况。

表3 控制个股特征

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel A: 控制 <i>Size</i>					
小	0.0219** (2.32)	0.0220** (2.37)	0.0216** (2.31)	0.0214** (2.29)	0.0210** (2.26)
中	0.0189** (2.12)	0.0188** (2.14)	0.0179** (2.02)	0.0180** (2.03)	0.0178** (2.01)
大	0.0185* (1.91)	0.0186* (1.89)	0.0191* (1.95)	0.0193** (1.97)	0.0192* (1.96)
Panel B: 控制 <i>B/M</i>					
低	0.0209* (1.90)	0.0209* (1.89)	0.0217* (1.93)	0.0217* (1.93)	0.0221** (1.97)
中	0.0167** (1.98)	0.0167** (1.99)	0.0162* (1.91)	0.0161* (1.91)	0.0166** (1.96)
高	0.0159* (1.77)	0.0160* (1.76)	0.0177* (1.95)	0.0178** (1.96)	0.0162* (1.78)
Panel C: 控制 <i>Mom</i>					
低	0.0234*** (2.61)	0.0233** (2.58)	0.0226** (2.47)	0.0226** (2.47)	0.0230** (2.51)
中	0.0202** (2.55)	0.0202** (2.55)	0.0209*** (2.63)	0.0210*** (2.64)	0.0203** (2.55)
高	0.0250** (2.23)	0.0251** (2.26)	0.0233** (2.09)	0.0231** (2.08)	0.0225** (2.01)
Panel D: 控制 <i>Price</i>					
低	0.0265*** (3.01)	0.0264*** (3.04)	0.0262*** (2.97)	0.0262*** (2.96)	0.0260*** (2.96)
中	0.0152* (1.83)	0.0151* (1.83)	0.0146* (1.75)	0.0146* (1.75)	0.0144* (1.71)
高	0.0324*** (2.91)	0.0323*** (2.81)	0.0314*** (2.73)	0.0313*** (2.73)	0.0309*** (2.69)

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

(2) 控制个股关注度。众多研究表明个股关注度会影响投资者的交易行为(如: Barber and Odean, 2008; Da *et al.*, 2011; Lou, 2014)。因此, 在探究行业关注度与股票收益率的关系时, 有必要控制个股关注度的影响。参考已有文献, 本文控制前一天收益率 (Ret)、异常交易量 (Abn_volume)、异

常换手率 ($Abn_turnover$)、分析师跟踪数 ($Analyst$) 四个间接关注度指标, 以及个股异常搜索量 ($ASVI^{STK}$) 这一直接关注度指标。

表4的结果表明, 无论是控制了间接关注度 (Panel A 至 Panel D) 还是直接关注度 (Panel E) 指标, 高 $ASVI^{IND}$ 组与低 $ASVI^{IND}$ 组的收益率之差基本都显著为正。并且, 无论是原始收益率还是 alpha 都表现出这样的规律, 表明行业关注度与股票横截面收益率之间的正向关系在剥离了个股关注度的影响之后依然存在。

表4 控制个股关注度

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel A: 控制 Ret					
低	0.0170* (1.86)	0.0169* (1.89)	0.0169* (1.89)	0.0168* (1.88)	0.0158* (1.77)
中	0.0138* (1.76)	0.0139* (1.78)	0.0140* (1.76)	0.0140* (1.76)	0.0130 (1.60)
高	0.0254** (2.16)	0.0255** (2.17)	0.0250** (2.14)	0.0251** (2.15)	0.0251** (2.14)
Panel B: 控制 Abn_volume					
低	0.0314*** (3.79)	0.0316*** (3.81)	0.0320*** (3.80)	0.0322*** (3.83)	0.0305*** (3.60)
中	0.0198** (2.57)	0.0198** (2.53)	0.0191** (2.39)	0.0191** (2.39)	0.0190** (2.37)
高	0.0270** (2.30)	0.0270** (2.36)	0.0278** (2.41)	0.0278** (2.41)	0.0281** (2.42)
Panel C: 控制 $Abn_turnover$					
低	0.0314*** (3.73)	0.0314*** (3.72)	0.0314*** (3.67)	0.0315*** (3.67)	0.0304*** (3.55)
中	0.0159** (2.04)	0.0159** (2.06)	0.0155** (1.97)	0.0156** (1.97)	0.0150* (1.88)
高	0.0271** (2.28)	0.0272** (2.31)	0.0282** (2.39)	0.0284** (2.40)	0.0279** (2.36)
Panel D: 控制 $Analyst$					
低	0.0202** (2.04)	0.0201** (1.99)	0.0185* (1.80)	0.0184* (1.79)	0.0195* (1.90)
中	0.0197** (2.26)	0.0197** (2.22)	0.0196** (2.20)	0.0196** (2.20)	0.0194** (2.18)
高	0.0161* (1.81)	0.0162* (1.76)	0.0172* (1.82)	0.0171* (1.82)	0.0170* (1.83)

(续表)

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel E: 控制 $ASVI^{STK}$					
低	0.0142 (1.56)	0.0143 (1.62)	0.0132 (1.48)	0.0135 (1.51)	0.0125 (1.40)
中	0.0173** (2.10)	0.0172** (2.03)	0.0178** (2.06)	0.0177** (2.04)	0.0164* (1.89)
高	0.0227** (2.11)	0.0227** (2.08)	0.0244** (2.23)	0.0243** (2.22)	0.0247** (2.25)

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

综上所述，双变量组合分析的结果表明行业关注度与股票横截面收益率之间存在稳健的正向关系，这种关系不受公司特征和个股关注度的影响，假说1得到进一步验证。

(二) 检验关注度假说

1. 卖空约束

关注度本身是没有方向的，投资者可以关注正面或负面的新闻。然而，卖空约束会阻碍悲观的投资者卖出股票，导致吸引注意力的股票出现净的购买压力 (Baber and Odean, 2008)。因此，在关注度假说的框架下，我们预期价格压力在卖空约束强的股票中会更显著。

表5的Panel A中，参考Nagel (2005)，我们用机构投资者持股比例 ($Inst_ownership$) 作为卖空约束的代理指标，进行双变量排序分组。可以看到，多空组合的收益率在机构投资者持股比例较低的股票中更显著，这些股票的卖空约束更强。Panel B中，我们根据股票是否是融券标的股 ($Short_available$) 分为两组，再将其按照 $ASVI^{IND}$ 分为三组，发现价格压力主要在于非融券标的股中显著，与Panel A的结果一致。从而假说2得到验证。

表5 卖空约束与价格压力

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel A: 控制 $Inst_ownership$					
低	0.0236** (2.45)	0.0236** (2.37)	0.0234** (2.35)	0.0235** (2.35)	0.0237** (2.38)
中	0.0201** (2.23)	0.0202** (2.23)	0.0209** (2.28)	0.0208** (2.27)	0.0208** (2.26)
高	0.0168* (1.82)	0.0168* (1.84)	0.0165* (1.79)	0.0165* (1.78)	0.0163* (1.78)

(续表)

	原始收益率	α^{CAPM}	α^{FF3}	$\alpha^{Carhart4}$	α^{FF5}
	(1)	(2)	(3)	(4)	(5)
Panel B: 控制 <i>Short_available</i>					
是	-0.0123 (-1.21)	-0.0123 (-1.21)	-0.0103 (-1.01)	-0.0101 (-0.99)	-0.0119 (-1.15)
否	0.0304*** (4.11)	0.0305*** (4.11)	0.0298*** (3.96)	0.0297*** (3.95)	0.0295*** (3.91)

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

2. 长期反转

在 Barber and Odean (2008) 的关注度假说解释框架下, 如果价格上涨是由于对吸引注意力的股票的净买入压力而非基本面向好导致的, 那么这种价格上涨会出现反转。为了克服组合分析无法同时控制多个变量的缺点, 我们采用 Fama and MacBeth (1973) 回归方法对此进行检验。

参考 Hou and Robinson (2006) 的做法, 每一天, 用 $t+n$ 天的行业收益率对个股特征变量和个股关注度的行业均值进行横截面回归, 再将得到的系数在时间序列上平均, 得到最终的 Fama-MacBeth 回归系数。更为直观地, 横截面回归的式子如下:

$$R_{i,t+n} = \alpha_{0,t} + \alpha_{1,t} ASVI_{i,t}^{IND} + \sum_{n=1}^N \beta_{n,t} X_{i,t} + \epsilon_{i,t+n}, \quad (3)$$

其中 $R_{i,t+n}$ 为行业 i 在 $t+n$ 天的收益率, 为该行业所有股票的收益率的平均, 单位为 %。 $ASVI_{i,t}^{IND}$ 为行业 i 在 t 天的异常搜索量。 $X_{i,t}$ 为行业 i 在 t 天的一系列控制变量, 为该行业所有股票的某一变量的均值。控制变量包括前文涉及的所有公司特征变量和个股关注度变量。⁸

为了进一步控制前期搜索量的影响, 我们还在回归中加入了异常搜索量的滞后项。表 6 的结果显示, 价格上涨是短期的。当因变量为 $t+2$ 天的收益率时, $ASVI^{IND}$ 的系数不再显著且变为负值; 在因变量为 $t+4$ 天的收益率时, $ASVI^{IND}$ 的系数显著为负, 表明最初的价格压力在 4 天内出现了反转, 假说 3 得到验证。

表 6 检验反转效应

	$t+1$ 天	$t+2$ 天	$t+3$ 天	$t+4$ 天
	(1)	(2)	(3)	(4)
$ASVI_{i,t}^{IND}$	0.0096** (2.30)	-0.0017 (-0.43)	-0.0032 (-0.85)	-0.0076* (-1.94)

⁸ 限于篇幅, 控制变量的回归系数未在后文的表格中展示, 可向作者索取。

(续表)

	$t+1$ 天	$t+2$ 天	$t+3$ 天	$t+4$ 天
	(1)	(2)	(3)	(4)
$ASVI_{t-1}^{IND}$	-0.0005 (-0.13)	0.0016 (0.39)	-0.0051 (-1.27)	0.0020 (0.48)
$ASVI_{t-2}^{IND}$	-0.0008 (-0.20)	-0.0060 (-1.45)	-0.0005 (-0.12)	-0.0003 (-0.08)
$ASVI_{t-3}^{IND}$	-0.0055 (-1.44)	0.0003 (0.09)	0.0007 (0.19)	-0.0015 (-0.36)
控制变量	是	是	是	是
<i>Intercept</i>	0.9372*** (3.86)	0.9136*** (3.63)	0.9016*** (3.58)	0.9637*** (3.72)
<i>N</i>	229 512	229 400	229 286	229 168
R^2	0.2890	0.2814	0.2783	0.2739

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

综合以上结果，我们发现行业关注度与股票横截面收益率显著正相关，这种关系在卖空约束强的股票中更突出，且关注度导致的正向价格压力会在接下来几天反转，与 Barber and Odean (2008) 的关注度假说一致。

(三) 行业关注度与个股关注度预测能力对比

1. 行业和个股关注度与股票收益率

为了比较行业关注度与个股关注度在预测股票收益率上的差别，我们首先分别检验行业和个股关注度与下一期股票收益率的关系，再检验当回归中同时纳入两种关注度指标时系数的变化情况。延续上一节的做法，我们用下一天的行业收益率对个股特征变量和个股关注度的行业均值进行 Fama-MacBeth 回归，以此来检验 $ASVI^{IND}$ 与股票收益率的关系。在检验 $ASVI^{STK}$ 与股票收益率的关系时，回归方法类似，但是是在个股层面上进行，即因变量和自变量都是个股变量。

表 7 的第 (1) 列中，与 Da *et al.* (2011) 的结果一致， $ASVI^{STK}$ 的系数为 0.0823， t 值为 7.74，表明个股异常搜索量与收益率显著正相关。第 (2) 列 $ASVI^{IND}$ 的系数显著为正，再次验证了假说 1。然而，将两个关注度指标同时加入回归中时， $ASVI^{STK}$ 的系数变得不显著，说明在控制了行业关注度后，个股关注度对股票收益率的影响消失了，表明行业关注度对股价的预测能力强于个股关注度，假说 4 得到验证。

表7 行业和个股关注度与股票收益率

	(1)	(2)	(3)
$ASVI^{IND}$		0.0093**	0.0090**
		(2.57)	(2.51)
$ASVI^{STK}$	0.0823***		0.0215
	(7.74)		(0.83)
控制变量	是	是	是
<i>Intercept</i>	0.8743***	0.9746***	0.9260***
	(4.35)	(4.08)	(3.85)
<i>N</i>	2 713 631	229 732	229 732
R^2	0.0837	0.2489	0.2641

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

2. 行业与个股关注度的领先滞后关系

为了解释行业关注度更强的股价预测能力，我们对行业和个股信息的领先滞后关系进行检验。表8的第(1)列，我们用 $t+1$ 天的 $ASVI^{STK}$ 对 t 天的 $ASVI^{IND}$ 以及一系列个股控制变量回归；第(2)列则将自变量和因变量调换，即用下一天的 $ASVI^{IND}$ 对行业平均的 $ASVI^{STK}$ 以及控制变量回归。如果第(1)列 $ASVI^{IND}$ 的系数显著为正，而第(2)列 $ASVI^{STK}$ 的系数不显著，则表明行业关注度能够预测个股关注度，但个股关注度无法预测行业关注度，意味着行业层面的信息领先于个股信息，因而行业关注度有更强的股价预测能力。实证结果与预期一致，行业搜索量是领先于个股搜索量的，假说4得到进一步验证。

表8 行业和个股关注度的领先滞后关系

	$ASVI_{t+1}^{STK}$	$ASVI_{t+1}^{IND}$
	(1)	(2)
$ASVI^{IND}$	0.0017***	0.3473***
	(4.67)	(79.23)
$ASVI^{STK}$	0.4102***	0.0116
	(109.95)	(1.21)
控制变量	是	是
<i>Intercept</i>	0.1448***	-0.1602**
	(4.56)	(-2.16)
<i>N</i>	4 087 308	344 749
R^2	0.2959	0.2435

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

五、稳健性检验

（一）排除内生性

为了排除遗漏变量等潜在的内生性问题，我们以上一级行业（二级行业）分类下其他三级行业的关注度均值（记为 $ASVI^{IND}_{other}$ ）作为工具变量，对行业关注度与股票收益率的关系进行检验。同一个二级行业分类下的子行业的业务模式和营业收入的构成具有很大的相似性，因此很容易被投资者关联搜索。从而对任一行业而言，同一上级行业下的其他行业的搜索量与该行业的搜索量存在相关性，表9第一阶段的回归系数显著为正验证了这个假设。第二阶段的结果表明， $ASVI^{IND}$ 与股票收益率之间存在显著的正相关关系，因而本文的结果在控制了内生性之后依然是稳健的。

表9 两阶段最小二乘回归

	第一阶段	第二阶段
	$ASVI^{IND}$	Returns
	(1)	(2)
Predicted $ASVI^{IND}$		0.2846*** (7.83)
$ASVI^{IND}_{other}$	0.2369*** (9.98)	
控制变量	是	是
Intercept	0.1676 (0.53)	1.6639*** (3.85)
N	159 699	159 699
Day-of-the-week fixed effects	是	是
Industry fixed effects	是	是
Standard error cluster	Industry	Industry

注：括号内为 Newey-West 调整后的 t 值。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。

（二）去除股灾样本

前文对多空组合净值曲线的分析表明，基于行业关注度的策略会受到市场情绪的影响。为了排除极端市场情况对结果的影响，我们将2015年6月15日至2016年2月29日股灾期间的样本去除，进行Fama-MacBeth回归。表10第(1)列的结果显示， $ASVI^{IND}$ 的系数显著为正，表明我们的结果在去除了股灾样本之后依然是稳健的。

(三) 区分电脑和移动端搜索量

前文都是以电脑端和移动端的搜索量之和作为搜索指数,本小节将电脑端与移动端的搜索量加以区分并分别研究其预测能力。表10第(2)–(3)列的结果表明,仅移动端的搜索量表现出显著的股价预测能力。因而,在样本期内,行业搜索量对股票收益率的预测能力主要是由移动端搜索量驱动的。⁹

(四) 其他异常关注度指标

到目前为止,ASVI^{IND}都是以过去7天的搜索量均值为基准计算的,内在的假设为7天的搜索量均值代表了一个正常的搜索量水平。表10的第(4)–(6)列,我们将计算ASVI^{IND}(ASVI^{STK})的基准值调整为过去14天、21天、28天的SVI^{IND}(SVI^{STK})均值,结果显示ASVI^{IND}的系数均显著为正,表明我们的结果对不同的异常关注度定义方法依然稳健。

表10 其他稳健性检验

	去股灾样本	区分移动和电脑端		其他异常关注度定义方法		
	(1)	移动端 (2)	电脑端 (3)	14天 (4)	21天 (5)	28天 (6)
ASVI ^{IND}	0.0096*** (2.72)	0.0103*** (3.63)	0.0047 (1.00)	0.0083** (2.33)	0.0076** (2.17)	0.0070** (2.00)
控制变量	是	是	是	是	是	是
Intercept	0.6449*** (2.77)	0.8774*** (3.66)	0.9138*** (3.77)	0.9191*** (3.81)	0.9055*** (3.75)	0.8901*** (3.68)
N	211 011	229 732	230 165	229 187	228 642	228 097
R ²	0.2582	0.2630	0.2629	0.2646	0.2649	0.2651

注:括号内为Newey-West调整后的t值。*** $p < 0.01$,** $p < 0.05$,* $p < 0.1$ 。

六、结 论

尽管行业在资本市场中的重要作用已经得到广泛认可,尚未有学者研究如何利用投资者行业关注度构建投资策略。本文以百度搜索指数作为关注度的代理变量,实证研究发现行业关注度能正向预测股票横截面收益率。基于行业关注度构建的多空策略可获得显著的超额收益率,这在控制了个股特征

⁹ 我们也检验了2011年之前电脑端搜索量的股价预测能力,发现2011年之前电脑端搜索量有显著的预测能力。这与我国移动互联网的发展规律是一致的:2009年3G牌照正式发放后,我国移动互联网进入飞速发展阶段。据《第47次中国互联网络发展状况统计报告》,2020年网民中使用手机上网的比例已经高达99.7%。电脑端网民占比的急剧下跌解释了2011年之后电脑端搜索量的预测能力不显著,而移动端搜索量却表现出很强的预测能力。限于篇幅,该部分结果未放入文中。

和个股关注度之后依然稳健。借助 Barber and Odean (2008) 的投资者关注度假说分析框架, 后续分析表明行业关注度导致的价格压力效应在卖空约束强的股票中更突出, 并且会在接下来几天反转。更为重要的, 我们对比了个股关注度与行业关注度在预测股票收益率上的相对强弱, 发现行业关注度具有更强的股价预测能力, 这是因为行业关注度与个股关注度之间存在一个领先滞后关系。稳健性检验表明我们的结果不受内生性、极端市场情况、搜索指数构成以及异常关注度构建方法的影响。

本文的结果具有重要的实践启示。一方面, 我们的结果表明, 基于行业关注度的投资策略能获得显著且平稳的超额收益率, 并且行业关注度相比个股关注度而言对股票收益率的影响更强。因此对于个人投资者而言, 可以利用百度指数量化市场情绪, 识别热门行业, 构建基于行业关注度的投资策略以获取超额回报。对于机构投资者而言, 应加强对行业关注度相关指标的重视程度, 将其纳入跟踪范畴, 作为预测散户投资动态以及股票估值的参考指标, 开发基于行业关注度的量化投资产品。但另一方面, 我们发现行业搜索量对股票收益率的影响是短期的, 表明基于关注度的交易蕴含着很强的投机性。个人投资者由于信息获取、信息处理和认知能力等方面的限制, 容易掉入盲目“炒热点”的投资陷阱。因此, 对于监管者而言, 应加强对网络搜索指数的监测, 以此预判市场资金流向、改进监管手段, 从而实现对市场的精准调控, 以提升监管效率。

当然, 本文还有很大的拓展空间。一是关注度的度量方面。如何准确地度量投资者关注度一直是相关研究的重点和难点(贾春新等, 2010)。虽然以搜索指数衡量投资者关注度的有效性已经被诸多学者证明(如: Da *et al.*, 2011; 赵龙凯等, 2013), 不可否认该指标还存在很多缺陷, 比如贡献搜索量的不一定是股民, 以及搜索关键词可能存在噪音等。因此如何科学有效地度量关注度是后续可以改进的方向。二是行业的定义方面。以往关于行业的研究大多是基于固有的行业分类(如证监会、中信证券行业分类), 近年来越来越多的学者从上市公司年报中挖掘行业信息, 这种“隐性”行业信息在资产定价中的作用甚至强于固有的行业分类(Hoberg and Phillips, 2018)。因此, 后续研究可以参考胡聪慧等(2015)以及Hoberg and Phillips(2018)等学者的研究从上市公司年报中提取行业信息, 据此构建行业关注度指标, 并探究这种基于隐性行业分类的指标与传统指标的差异。

参 考 文 献

- [1] Barber, B. M., and T. Odean, "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors", *Review of Financial Studies*, 2008, 21 (2), 785-818.

- [2] Barberis, N., and A. Shleifer, "Style Investing", *Journal of Financial Economics*, 2003, 68 (2), 161-199.
- [3] 包锋、徐建国, "异质信念的变动与股票收益", 《经济学》(季刊), 2015年第14卷第4期, 第1591—1610页。
- [4] Carhart, M. M., "On Persistence in Mutual Fund Performance", *Journal of Finance*, 1997, 52 (1), 57-82.
- [5] Chan, K., and A. Hameed, "Stock Price Synchronicity and Analyst Coverage in Emerging Markets", *Journal of Financial Economics*, 2006, 80 (1), 115-147.
- [6] Chan, K., A. J. Menkveld, and Z. Yang, "Information Asymmetry and Asset Prices: Evidence from the China Foreign Share Discount", *Journal of Finance*, 2008, 63 (1), 159-196.
- [7] Choi, N., and R. W. Sias, "Institutional Industry Herding", *Journal of Financial Economics*, 2009, 94 (3), 469-491.
- [8] Chou, P., P. Ho, and K. Ko, "Do Industries Matter in Explaining Stock Returns and Asset-pricing Anomalies?", *Journal of Banking & Finance*, 2012, 36 (2), 355-370.
- [9] Da, Z., J. Engelberg, and P. Gao, "In Search of Attention", *Journal of Finance*, 2011, 66 (5), 1461-1499.
- [10] Da, Z., J. Engelberg, and P. Gao, "The Sum of All FEARS Investor Sentiment and Asset Prices", *Review of Financial Studies*, 2015, 28 (1), 1-32.
- [11] Drake, M. S., J. Jennings, D. T. Roulstone, and J. R. Thornock, "The Comovement of Investor Attention", *Management Science*, 2017, 63 (9), 2773-3145.
- [12] Fama, E. F., and K. R. French, "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, 1993, 33 (1), 3-56.
- [13] Fama, E. F., and K. R. French, "A Five-factor Asset Pricing Model", *Journal of Financial Economics*, 2015, 116 (1), 1-22.
- [14] Fama, E. F., and K. R. French, "The Cross-section of Expected Stock Returns", *Journal of Finance*, 1992, 47 (2), 427-465.
- [15] Fama, E. F., and J. D. MacBeth, "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy*, 1973, 81 (3), 607-636.
- [16] 冯旭南, "注意力影响投资者的股票交易行为吗? ——来自‘股票交易龙虎榜’的证据", 《经济学》(季刊), 2017年第16卷第1期, 第255—274页。
- [17] Fritzscheier, L. H., "Relative Price Fluctuations of Industrial Stocks in Different Price Groups", *Journal of Business of the University of Chicago*, 1936, 9 (2), 133-154.
- [18] Froot, K., and M. Teo, "Style Investing and Institutional Investors", *Journal of Financial and Quantitative Analysis*, 2008, 43 (4), 883-906.
- [19] Gilbert, T., S. Kogan, L. Lochstoer, and A. Ozyildirim, "Investor Inattention and the Market Impact of Summary Statistics", *Management Science*, 2012, 58 (2), 336-350.
- [20] Hirshleifer, D., and S. H. Teoh, "Limited Attention, Information Disclosure, and Financial Reporting", *Journal of Accounting and Economics*, 2003, 36 (1-3), 337-386.
- [21] Hoberg, G., and G. M. Phillips, "Text-based Industry Momentum", *Journal of Financial and Quantitative Analysis*, 2018, 53 (6), 2355-2388.
- [22] Hou, K., and D. T. Robinson, "Industry Concentration and Average Stock Returns", *Journal of Finance*, 2006, 61 (4), 1927-1956.
- [23] 胡聪慧、刘玉珍、吴天琪、郑建明, "有限注意、行业信息扩散与股票收益", 《经济学》(季刊), 2015年第14卷第3期, 第1173—1192页。

- [24] Jame, R., and Q. Tong, "Industry-based Style Investing", *Journal of Financial Markets*, 2014, 19, 110-130.
- [25] Jegadeesh, N., and S. Titman, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *Journal of Finance*, 1993, 48 (1), 65-91.
- [26] Jiang, H., S. Z. Li, and H. Wang, "Pervasive Underreaction: Evidence from High-frequency Data", *Journal of Financial Economics*, 2021, 141 (2), 573-599.
- [27] Jones, C. M., and O. A. Lamont, "Short-sale Constraints and Stock Returns", *Journal of Financial Economics*, 2002, 66 (2-3), 207-239.
- [28] Jones, C. M., D. Shi, X. Zhang, and X. Zhang, "Heterogeneity in Retail Investors: Evidence from Comprehensive Account-level Trading and Holdings Data", Working Paper, Columbia University, 2020.
- [29] 贾春新、赵宇、孙萌、汪博, "投资者有限关注与限售股解禁", 《金融研究》, 2010年第11期, 第108—122页。
- [30] Li, J., and J. Yu, "Investor Attention, Psychological Anchors, and Stock Return Predictability", *Journal of Financial Economics*, 2012, 104 (2), 401-419.
- [31] Lou, D., "Attracting Investor Attention through Advertising", *Review of Financial Studies*, 2014, 27 (6), 1797-1829.
- [32] MacKay, P., and G. M. Phillips, "How Does Industry Affect Firm Financial Structure?", *Review of Financial Studies*, 2005, 18 (4), 1433-1466.
- [33] Moskowitz, T. J., and M. Grinblatt, "Do Industries Explain Momentum?", *Journal of Finance*, 1999, 54 (4), 1249-1290.
- [34] Nagel, S., "Short Sales, Institutional Investors and the Cross-section of Stock Returns", *Journal of Financial Economics*, 2005, 78 (2), 277-309.
- [35] Newey, W., and K. West, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 1987, 55, 703-708.
- [36] Peng, L., and W. Xiong, "Investor Attention, Overconfidence and Category Learning", *Journal of Financial Economics*, 2006, 80 (3), 563-602.
- [37] Seasholes, M. S., and G. Wu, "Predictable Behavior, Profits, and Attention", *Journal of Empirical Finance*, 2007, 14 (5), 590-610.
- [38] Vozlyublennaia, N., "Investor Attention, Index Performance, and Return Predictability", *Journal of Banking & Finance*, 2014, 41, 17-35.
- [39] Yuan, Y., "Market-wide Attention, Trading, and Stock Returns", *Journal of Financial Economics*, 2015, 116 (3), 548-564.
- [40] 杨涛、郭萌萌, "投资者关注度与股票市场——以PM2.5概念股为例", 《金融研究》, 2019年第5期, 第190—206页。
- [41] 杨晓兰、沈翰彬、祝宇, "本地偏好、投资者情绪与股票收益率: 来自网络论坛的经验证据", 《金融研究》, 2016年第12期, 第143—158页。
- [42] 俞庆进、张兵, "投资者有限关注与股票收益——以百度指数作为关注度的一项实证研究", 《金融研究》, 2012年第8期, 第152—165页。
- [43] 赵龙凯、陆子昱、王致远, "众里寻'股'千百度——股票收益率与百度搜索量关系的实证探究", 《金融研究》, 2013年第4期, 第183—195页。

Industry-wide Attention Based on Internet Searches and the Cross-section of Stock Returns

XUEYONG ZHANG* GUOMEI TANG

(Central University of Finance and Economics)

Abstract This study investigates how investor attention on industries affects the cross-section of stock returns, using aggregate search frequency from the Baidu Index based on industry names as a proxy for industry-specific attention. We find that an increase in the index predicts higher stock returns the next day. A long-short strategy based on this attention proxy yields an equal-weighted (value-weighted) annual return of 4.89% (4.28%). The buying pressure which will revert is stronger among stocks with more binding short-sales constraints. Further analysis shows that industry-level attention proxy is dominant over individual stock-level search volume index in predicting stock returns.

Keywords Baidu search volume index, industries, limited attention

JEL Classification G11, G12, G14

* Corresponding Author: Xueyong Zhang, School of Finance, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing 100081, China; Tel: 86-13701249262; E-mail: zhangxueyong@cufe.edu.cn.