

金融社交媒体信息、股票收益率与资本市场异象

刘汶楚, 王希钰, 王梦涵, 刘研

(武汉大学, 武汉 430000)

摘要: 本文首次构建金融社交媒体信息指标 (FI), 研究了金融社交媒体信息对股票横截面收益率的预测力以及对资本市场异象的影响. 研究发现: 金融社交媒体信息与股票未来横截面收益率间存在明显的单调递增特征, 并且在控制多种风险或者情绪定价因子后上述特征仍然稳健. 金融社交媒体信息对股票收益率的预测力体现了信息经由金融社交平台进入股价的过程. 金融社交媒体信息将捕捉到的股票基本面信息注入市场, 然而受到市场套利成本的阻碍, 信息进入股价的过程被拉长, 导致金融社交媒体信息对股票收益率的预测力呈现持久性. 本文的实证证据进一步发现, 金融社交媒体信息提供的股价信息有助于投资者识别并消除妨碍市场合理运行的市场异象, 从而有效提升资本市场运行效率. 本文研究结论对于进一步推动互联网建设、加速互联网与资本市场融合建设、促进资本市场高质量发展具有重要的理论意义和现实价值.

关键词: 金融社交媒体信息; 资本市场效率; 市场异象; 资本市场高质量发展

1 引言

优化资本市场运行效率, 提升信息传递效率和保证信息准确性是我国宏观调控的重点. 资本市场的运行受信息驱动, 信息的畅通性、及时性和准确性直接影响市场效率. 因此, 监管层和学术界一直关注如何提高信息传递效率和确保信息准确性. 金融社交媒体作为将信息传递给投资者的渠道, 在信息传递过程中起着重要作用. 理想情况下, 金融社交媒体能高效地传递公司价值信息给投资者, 投资者再通过交易反映在股价上. 然而, 这不仅需要投资者准确分析和识别信息, 也需要金融社交媒体提供准确的信息.

金融社交媒体对资本市场的影响体现了信息与股票价格的关系. 金融社交媒体作为将信息传递到股票价格的中介渠道, 加速了信息的扩散, 但却无法对信息进行甄别. 因此, 金融社交媒体在资本市场运行中的作用尚存在争议. 部分学者认为金融社交媒体放大了投资者的乐观预期, 加剧了市场波动. 然而, 更多的研究认为金融社交媒体对市场稳定起到了积极作用. 金融社交媒体的普及有助于优化信息传递效率, 提高公司透明度, 并显著提升投资者的信息获取能力, 从而减少了市场中的信息不对称, 进而稳定了市场.

有效市场假设下股票价格应当及时充分反映信息, 然而当市场存在摩擦时, 搜集信息往往需要付出一定成本. 这种成本拉长了信息进入股价的过程, 使得信息能够预测股票未来收益率. 此时, 金融社交媒体信息与股票未来收益率应当正向相关. 实证上, 我们参考《金融社交媒体信息、股票收益率与资本市场异象》一文的做法, 并根据中国市场特殊的交易制度以及投资者行为特征进行修正, 构造出金融社交媒体信息指标 (FI) 用来刻画个股层面的金融社交媒体信息. 我们的实证结果也确实显示, 相比于金融社交媒体信息消极的股票 (低 FI), 金融社交媒体信息积极 (高 FI) 的股票有着更高的超额收益率. 换言之, 两者存在明显的正向单调递增特征. 并且与理论假设一致, 当股票面临更高的套利成本时, 这一特征明显增强. 进一步, 本文围绕这一发现给出更多证据表明金融社交媒体信息能够捕捉到股票基本面信息, 并因此有利于投资者消除妨碍市场合理运行的市场异象, 进而能够显著提升资本市场定价效率.

本文的关注点在于信息的传播过程. 现有研究主要通过分析公司财务状况或市场交易情况构建的公司特征指标, 来判断不同公司特征所蕴含的信息对股票收益率的影响. 然而, 公司特征一方面通常变化较慢, 无法充分反映动态信息, 且存在较大的信号噪音, 难以准确提取有效信息. 另一方面, 由于无法准确测量公司特征所包含的信息, 现有研究无法直接观测信息进入股价的过程, 只能通过间接手段观测信息含量的提升效果, 易受其他潜在因素的干扰. 与现有研究不同, 本文利用金融社交媒体信息数据, 聚焦于信息传播过程, 不仅提供了金融社交媒体准确测量基本面信息的经验证据, 而且在此基础上直接揭示了信息进入股价的过程, 从而精准识别了信息与股价之间的联系.

本文剩余部分结构安排如下：第二部分描述了数据来源与变量构造；第三部展示了实证分析与讨论；第四部分提供了进一步分析：市场异象在不同金融社交媒体信息股票中的表现；第五部分是本文结论和政策含义。

2 数据来源、变量构造与研究设计

2.1 金融社交媒体信息指标

本文的金融社交媒体信息数据来源于 CNRDS 数据库，该数据库的金融社交媒体信息数据取自东方财富网。通过爬虫并采用机器学习算法，该数据库识别得到发表帖子的情感信息，然后将帖子标签归类为看涨、看跌以及中性三类。数据起始于 2008 年 1 月，鉴于早期样本数过少且横截面检验需要用到前一个月数据，因此文中正式检验样本期为 2009 年 1 月到 2019 年 12 月。

在金融社交媒体信息指标的构造上，Antweiler 和 Frank 使用了看涨看跌数量之差与两者之和的比值作为金融社交媒体信息指标。对于中国市场研究而言，简单使用这一指标进行检验存在一定偏颇。这主要是因为中国资本市场存在严重的卖空限制，同时，投资者自身又存在诸如处置效应等心理认知偏差，导致负面信息极为可能完全无法进入股价。为此，本文只关注正面乐观帖子（看涨帖子），提出一个新指标，即金融社交媒体看涨信息指数（金融社交媒体信息指数）FI：

$$FI_{i,t} = \frac{Positive_{i,t}}{Total_{i,t}} \quad (1)$$

其中，分子 $Positive_{i,t}$ 表示股票 i 在 t 月看涨帖子数之和，分母 $Total_{i,t}$ 表示股票 i 在 t 月总帖子数之和。通过比例形式尽量排除数据噪声的影响，聚焦于刻画金融社交媒体看涨信息。考虑到本文构造的金融社交媒体看涨信息指数与已有类似指数有一定的相似度，因此本文参考相关研究构造三种类似指标：看涨看跌帖子数量之差与两者之和的比值作为金融社交媒体信息指标、看跌信息指数（看跌帖子数量与总帖子数之比）以及看涨看跌帖子数量之差与总帖子数的比值，并发现由于上述指标并未考虑中国市场独有的交易制度以及投资者行为特征，因此均无法准确刻画中国金融社交媒体信息，从而不具备预测力。

2.2 其他数据来源

本文选取的股票样本为 2009 年 1 月到 2019 年 12 月年间所有在上交所和深交所上市的 A 股股票（包括主板、中小企业板和创业板）。其中，股票交易数据和财务数据均来自 CSMAR 数据库，剔除非经常性损益的净利润和无风险利率取自 RESSET 数据库。为保证检验结果的有效性，本文剔除了每只上个月交易日小于 7 天的股票，并剔除过去 12 个月交易日小于 120 天的股票。最后，本文参考 Liu 等的做法剔除市值最小 30% 的股票。

2.3 盈余公告后漂移指标

本文采用原文的方法，分别构造两种盈余公告后漂移指标 PEAD。第一种指标是盈余公告发布日累计异常收益率（CAR），测量方式如下：

$$CAR_{i,t} = \sum_{d=-2}^{d=1} (R_{i,d} - R_{m,d}) \quad (2)$$

其中， d 表示以盈余公告发布日为基准的窗口期， $d=0$ 表示盈余公告发布日， $d=(-2, -1$ 和 $1)$ 分别表示盈余公告发布日前 2 日、前一日和后一日， $R_{i,d}$ 表示股票 i 在 d 日的收益率， $R_{m,d}$ 表示 d 日的市场综合市值加权收益率。使用每个月最新公告日计算出的 CAR 代表在股票 i 在 t 月的 CAR。

第二种指标是未预期盈余（SUE），测量方式如下：

$$SUE_{i,t} = \Delta_{i,t} / \delta(\Delta_{i,t}) \quad (3)$$

其中， $\Delta_{i,t}$ 表示股票 i 在 t 月的季度净利润相比去年的变化值， $\delta(\Delta_{i,t})$ 是 $\Delta_{i,t}$ 在过去两年中的标准差。

2.4 其他指标

参考现有文献的研究来构造其他指标，并结合中国市场特征进行适当调整.由于换手率（TO）在中国市场表现尤为显著，本文根据换手率构造两种新的投资者情绪指标 TOMAX（最高换手率）和 TOMIN（最低换手率），结合 TO（换手率）和 TOVOL（换手率波动率）共有四种换手率指标.所有变量定义参见表 1.

表 1 变量定义

简称	变量含义	变量定义
ME	市值	个股月末的总市值
EP	市盈率倒数	净利润/总市值 净利润：最近可获取的剔除非经常性损益的净利润； 总市值：ME
BM	账面市值比	账面价值/总市值 账面价值：最近可获取的剔除优先股的股东权益； 总市值：ME.剔除 BM 小于等于零项
ILL	非流动性指标	$ILL_{i,d} = \frac{ ret_{i,d} }{volume_{i,d}}$ 其中， $ ret_{i,d} $ 是个股 i 在 d 日收益率的绝对值， $volume_{i,d}$ 是个股 i 在 d 日的成交量.每个月取 $ILL_{i,d}$ 均值得到 ILL
TO	换手率	个股月换手率
REV	反转效应	个股月收益率
MOM	动量效应	个股过去 11 个月的收益率之和（不包括最近一个月）
β	贝塔	$R_{i,t} = \alpha + \beta_{i,t}^1 MKT_t + \beta_{i,t}^2 MKT_{t-1} + \varepsilon_{i,t}$ 其中， $R_{i,t}$ 代表个股 i 在 t 月的超额收益率， MKT_t 是 t 月剔除了 30%小股票后的市场超额收益； $\beta_{i,t} = \beta_{i,t}^1 + \beta_{i,t}^2$ 使用 48 个月的滚动窗口回归计算出 $\beta_{i,t}^1$ 和 $\beta_{i,t}^2$ ，两者求和得到 β
INST	机构投资者持股比率	个股最近可获得的机构投资者持股占比，对缺失值取 0
COSKEW	斜偏度	$COSKEW_{i,t} = \frac{Avg[\varepsilon_{i,d} MKT_d^2]}{\sqrt{Avg[\varepsilon_{i,d}^2] Avg[MKT_d^2]}}$ 其中， $\varepsilon_{i,d}$ 是个股 i 在 d 日的 Fama-French 三因子残差，残差由每月日度数据回归求得， MKT_d 是 d 日的市场超额收益
ISKEW	特质偏度	个股每月日度 Fama-French 三因子模型残差偏度
IVOL	特质波动率	个股每月日度 Fama-French 三因子模型残差的标准差
VOL	波动率	个股每月日收益率的标准差
MAX	最高收益率	个股每月最高日收益率
SKEW	偏度	个股每月日收益率偏度
MIN	最低收益率	个股每月最低日收益率
TOVOL	换手率波动率	个股每月日换手率的标准差
TOMAX	最高换手率	个股每月最高日换手率
TOMIN	最低换手率	个股每月最低日换手率
AGE	上市年龄	公司上市月数
Read	阅读量	相关股票信息的阅读量
Comment	评论数	相关股票信息的评论数量

2.5 主要变量描述性统计结果

表 2 展示了主要变量描述性统计结果.可以发现，金融社交媒体信息指标 FI 介于 0 到 0.886 之间，均值为 31.6%，标准差为 7.5%，意味着金融社交媒体信息对市场的平均看涨率为 31.6%，并且近似符合正态分布.其余变量的结果也有呈现，统计结果与以往的研究结果类似.

表 2 描述性统计

	Mean	S.D.	Min	0.250	Mdn	0.750	Max
FI	0.316	0.0750	0	0.265	0.312	0.361	0.886
Log(ME)	22.91	0.936	20.96	22.24	22.74	23.38	28.57
EP	1.710	7.718	-187.4	0.340	1.151	2.539	575.2
BM	0.491	0.719	0	0.220	0.376	0.626	33.13
TO	0.310	0.319	0	0.108	0.206	0.397	5.366
REV	1.751	16.12	-68.97	-6.439	0.207	8.090	2205
MOM	15.78	53.64	-178.9	-16.75	8.420	40.02	2229
IVOL	1.817	1.043	0	1.112	1.607	2.301	20.22
VOL	2.730	1.475	0	1.797	2.434	3.304	44.79
MAX	5.517	3.595	-10	3.188	4.749	7.382	73.13
MIN	-5.095	2.572	-12.63	-6.653	-4.552	-3.186	10.07
TOMAX	3.483	3.568	0.00300	1.180	2.308	4.478	51.98
TOMIN	0.675	0.749	0	0.228	0.437	0.840	18.80
TOVOL	0.757	0.862	0	0.237	0.476	0.947	20.88
SKEW	2.971	74.30	-412.9	-41.34	1.771	46.40	447.5
AGE	125.0	83.37	0	51	115	193	348

注：EP、REV、MOM、IVOL、VOL、MAX、MIN、TOMAX、TOMIN、TOVOL 和 SKEW 展示百分比结果

2.6 因子模型

在因子模型的选取上，本文主要选取 Liu 等的 CH3 和 CH4 模型作为检验用的因子模型.通常研究采用的 Fama-French 模型主要存在两个问题：其一是没有考虑“借壳上市”现象导致的“壳污染”；其二是 Fama-French 三因子和五因子模型都是基于美国市场特点所构造，并不符合中国市场.例如，Fama-French 五因子模型中新添加的投资因子和盈利因子都不适合中国市场，导致 Fama-French 五因子模型在中国市场的表现甚至不及 Fama-French 三因子模型.相比于 Fama-French 模型，CH 模型基本解决了上述问题，在中国市场表现出很强的解释力.有鉴于此，本文按照 Liu 等的方法构造 CH3 和 CH4 模型.

3 实证分析与讨论

本部分将对金融社交媒体信息是否涵盖股票基本面信息进行检验.具体而言，我们认为金融社交媒体信息将缓慢地进入股票价格，从而导致股票收益率可以被预测.为此，本文基于四个维度进行检验，首先运用组合分析检验金融社交媒体信息对股票横截面收益率的预测力，然后分别通过套利成本分析以及长期预测力分析，深入挖掘其预测力背后的经济学逻辑，最后直接检验金融社交媒体信息与公司经营表现的联系，基于多个维度提供金融社交媒体信息涵盖股票基本面信息的证据.

3.1 单变量组合分析

接着本文检验金融社交媒体信息能否预测股票收益率.使用类似的分组分析，分别计算出每个组合在 t+1 月的超额总市

值加权收益率和超额等权收益率，以及卖出低 FI 组合（L）买入高 FI 组合（H）构建的投资组合（H-L）的收益率，结果呈现于表 3。从表 3 的左半部分的结果可以发现，超额收益率（RET）随着 FI 的增加而逐步上升，表现出近似单调上升特征，月平均超额收益率从 0.69% 上升到 1.43%，（H-L）每月可以获取 0.73% 的平均收益率。使用 CH3 和 CH4 模型调整后，CH3 和 CH4 alpha 均高度显著，分别达到 0.63%（Newey-West t 值为 2.02）和 0.66%（Newey-West t 值为 2.22）。类比来看，可以发现表 3 右半部分的超额等权收益率结果也是颇为相似。上述检验结果说明 FI 与股票横截面收益率存在极为显著的正向关系，并且这一关系无法被现有的定价因子模型所解释。换言之，上述结果说明金融社交媒体信息能够正向预测股票收益率且这种预测力是独有的。

表 3 单变量分组分析

FI	Value_weighted			Equal_weighted		
	RET	CH3	CH4	RET	CH3	CH4
L	0.69	-0.08	-0.20	1.67	0.57***	0.47***
	(0.90)	(-0.53)	(-1.45)	(1.58)	(5.71)	(6.10)
2	0.78	0.01	-0.08	1.65	0.69***	0.64***
	(1.02)	(0.05)	(-0.37)	(1.65)	(6.04)	(5.83)
3	0.60	-0.07	-0.16	1.29	0.33**	0.30**
	(0.83)	(-0.22)	(-0.57)	(1.39)	(2.39)	(2.55)
4	0.39	-0.26***	-0.38***	1.34	0.39***	0.35***
	(0.62)	(-3.11)	(-3.21)	(1.45)	(5.78)	(5.92)
5	0.36	-0.32**	-0.33**	1.51	0.51***	0.54***
	(0.55)	(-2.00)	(-2.20)	(1.57)	(4.88)	(4.69)
6	0.30	-0.43**	-0.38*	1.53*	0.50***	0.53***
	(0.44)	(-2.30)	(-1.95)	(1.68)	(6.03)	(6.37)
7	0.81	0.34	0.40*	1.80*	0.82***	0.88***
	(1.22)	(1.48)	(1.70)	(1.90)	(6.94)	(6.93)
8	0.99	0.17*	0.21**	1.90**	0.91***	0.92***
	(1.42)	(1.66)	(2.09)	(2.02)	(6.95)	(6.84)
9	1.10*	0.43**	0.37**	2.00**	0.96***	0.93***
	(1.70)	(2.28)	(2.03)	(2.08)	(6.00)	(5.70)
H	1.43**	0.55**	0.46**	3.28***	2.09***	1.71***
	(2.18)	(2.60)	(2.01)	(2.77)	(5.88)	(6.96)
H-L	0.73**	0.63**	0.66**	1.62***	1.52***	1.24***
	(2.35)	(2.02)	(2.22)	(7.57)	(5.17)	(4.86)

注：***、**、*分别表示在 1%、5%、10%水平上显著；括号内的 t 值经过 Newey-West 调整。

H-L 组合的获利来自于买入 H 还是卖出 L 呢？答案是买入 H。组合 H 的平均每月 CH4 市值加权 alpha 高达 0.46%（Newey-West t 值为 2.01），作为对比组合 L 的 alpha 仅为 -0.20%（Newey-West t 值为 -1.45），可见金融社交媒体信息的预测力主要是由代表利好消息的高 FI 股票所驱动。这符合上一节中的发现，即高 FI 股票是经历过利好消息的优质股票。综合以上实证结果，我们可以间接地发现金融社交媒体信息涵盖了股票的基本面信息，由于市场对这类信息反应缓慢，从而使得金融社交媒体信息能够预测股票横截面收益率。

3.2 套利成本限制分析

在确认了金融社交媒体信息能够预测股票横截面收益率后，下一步我们分析这种预测力是否会受到套利成本限制的影响。理论上，信息的传播存在着一定的阻力。投资者在通过交易将信息注入股价时，面临着诸如不确定性等套利成本限制的约

束.套利成本限制一方面增加了投资者的信息获取成本,另一方面也阻碍了交易的发生.因此在套利成本高的股票中,金融社交媒体信息进入股价的速度会变得更缓慢,FI 的预测力也应当更强,而在套利成本低的股票中信息传播速度加快,FI 预测力会随之相对减弱.

在上述理论的指引下,本文使用双变量分组分析法检验套利成本对 FI 预测力的影响.按照 t 月套利成本指标的大小以 30%和 70%为分位点将股票分为三个投资组合,在每个组中进一步按照 FI 分成十组并求出超额市值加权收益率和超额等权收益率.在套利成本指标的选取上,本文借鉴 Lam 等和 Gu 等的做法,选取 ME、EP、IVOL 和 INST 度量套利交易难易程度.股票的市值(ME)越小越难以交易、受到的关注越少,套利成本也就越高;EP 衡量了股票的价值高低,Fama 和 French 认为成长型股票(低 EP)相较价值型股票(高 EP)风险更高,因此 EP 越小套利成本也就越高;IVOL 代表特质风险,其值越大套利成本越高;INST 刻画有经验的交易者占比,或者卖空交易的难易程度,所以 INST 越高,套利成本应当越低.

表 4 给出了按照套利成本大小和 FI 进行双变量分组的组合超额收益率结果.结果显而易见(见表 4 右侧的等权结果),在套利成本高的组合(小 ME,低 EP,高 IVOL 和低 INST)中组合 H-L 的收益率分别为 2.43%,1.06%,2.18%和 2.08%,Newey-West t 值分别为 2.22,5.28,3.68 和 3.45,均显著为正并且远大于对应的低套利成本组(大 ME,高 EP,低 IVOL 和高 INST).作为稳健性检验,表 4 左侧市值加权的结果虽然有所减弱,但也说明组合 H-L 在套利成本高的股票中能取得更高收益.比较不同组中 CH3(CH4) alpha 可以进一步确认这一发现.这些发现支持了套利成本越高,市场对信息反应越为缓慢的理论假设,进一步证明 FI 的预测力来源于股票基本面信息的理论假设.

表 4 套利成本分组分析

	Panel A:ME						Panel B:EP					
	Value-weighted			Equal-weighted			Value-weighted			Equal-weighted		
	S	M	B	S	M	B	S	M	B	S	M	B
L	2.98*** (2.66)	0.75 (0.72)	0.28 (0.48)	3.58*** (3.08)	0.79 (0.74)	0.35 (0.48)	0.28 (0.3)	0.77 (0.99)	1.19* (1.87)	1.34 (1.26)	1.4 (1.55)	2.06** (2.29)
H	5.04*** (2.9)	1.69* (1.86)	1.61** (2.25)	6*** (2.95)	1.7* (1.85)	2.05 (1.64)	0.76 (0.84)	1.18* (1.67)	1.52*** (2.73)	2.4** (2.12)	2.45** (2.49)	2.58*** (3.03)
H-L	2.06** (2.51)	0.94*** (3.43)	1.32*** (3.22)	2.43** (2.22)	0.92*** (3.59)	1.7** (2.25)	0.48* (1.95)	0.41* (1.82)	0.34 (0.94)	1.06*** (5.28)	1.05*** (5.89)	0.52* (1.88)
Alpha	2.09*** (9.99)	-0.13*** (-2.62)	-0.1 (-1.5)	2.64*** (9.82)	-0.13*** (-3.19)	-0.03 (-0.29)	0.18 (1.4)	0.17 (1.18)	-0.1 (-1.04)	0.78*** (7.38)	0.78*** (7.13)	0.5*** (3.36)
	Panel C:IVOL						Panel D:INST					
	Value-weighted			Equal-weighted			Value-weighted			Equal-weighted		
	S	M	B	S	M	B	S	M	B	S	M	B
L	0.71 (0.93)	1.17 (1.44)	0.37 (0.41)	1.91* (1.93)	1.65* (1.75)	1.09 (1.15)	2.34** (2.3)	1.87** (2.29)	1.52* (1.9)	2.27** (2.03)	1.51* (1.66)	1.22 (1.21)
H	1.49** (2.38)	1.37** (2.16)	1.42 (1.65)	2.67** (2.32)	2.63*** (2.95)	3.28** (2.6)	4.37** (2.51)	4.74*** (2.67)	2.6*** (4)	4.35** (2.61)	3.16*** (2.97)	2.37*** (2.99)
H-L	0.78** (2.28)	0.2 (0.52)	1.05** (2.35)	0.76* (1.89)	0.98*** (6.2)	2.18*** (3.68)	2.03** (2.29)	2.86** (2.31)	1.07*** (3.14)	2.08*** (3.45)	1.66*** (7.67)	1.15*** (3.7)
Alpha	-0.07 (-0.61)	0.04 (0.48)	-0.01 (-0.03)	0.64*** (2.88)	0.82*** (11.91)	0.72*** (4.53)	1.79*** (11.49)	1.2*** (4.74)	0.56*** (4.19)	1.37*** (9.28)	0.67*** (8.5)	0.33*** (2.73)

注:***、**、*分别表示在 1%、5%、10%水平上显著;括号内的 t 值经过 Newey-West 调整。

3.3 长期表现检验

尽管我们的上述检验都支持 FI 的预测力来源于其对基本面信息的捕捉,但仍然存在另一种可能的解释,即金融社交媒体

体信息 FI 可能捕捉了投资者情绪.行为金融学认为投资者情绪会造成股价在短期偏离基本面,但在长期随着市场认识到价格偏误会出现修正过程,股价呈现“先涨后跌”的态势.与之形成鲜明对比,市场对信息的反应缓慢则会造成股价保持增长并随时间增加而逐渐减弱,维持涨的态势.综合上述研究结果可以直观的发现,通过检验金融社交媒体信息 FI 的长期预测力表现,就可以排除 FI 度量投资者情绪的可能.为此,本文按照 t 月的 FI 将股票分为十组构建组合 H-L,分别计算出从 t+2 到 t+13 月组合 H-L 的收益率和 CH4 调整后的收益率.表 5 给出了 FI 长期预测力分析的检验结果.纵向对比表 5 中每一列的分析结果,可以发现从 t+2 期一直到 t+13 期,6 组合 H-L 的收益率 (CH4alpha) 均为正值.同时,也可发现 FI 对股票收益率的预测力随着期限的拉长呈现明显的下降趋势,在 t+13 月收益率下降至 0.44% (t 值为 2.19),显著程度下降,标志着 FI 的预测力在一年期后降低.总结以上实证结果可以发现 FI 的预测力表现与投资者情绪的理论假设完全相悖,基于此,我们排除了 FI 度量投资者情绪的可能,并进一步肯定金融社交媒体信息蕴含股票基本面信息.

表 5 长期预测力分析

	Value_weighted		Equal_weighted	
	H-L	Alpha	H-L	Alpha
t+2	0.40 (1.33)	0.28* (1.68)	0.34** (2.25)	0.22*** (3.51)
t+3	0.51*** (2.74)	0.54*** (4.90)	0.38** (2.40)	0.37*** (6.13)
t+4	0.53 (1.18)	0.38* (1.90)	0.47 (1.60)	0.30* (1.69)
t+5	0.49 (1.54)	0.73*** (3.64)	0.40 (1.52)	0.34 (1.52)
t+6	0.36 (0.74)	0.34 (1.06)	0.45 (1.30)	0.42* (1.66)
t+7	0.55 (1.31)	0.81*** (2.77)	0.07 (0.26)	0.08 (0.36)
t+8	0.65** (2.47)	0.53** (2.04)	0.22 (0.98)	0.03 (0.12)
t+9	0.58*** (2.74)	0.46 (1.55)	0.24** (2.10)	0.05 (0.24)
t+10	0.50*** (2.87)	0.36** (2.23)	0.26*** (3.69)	0.12 (0.96)
t+11	0.52*** (2.67)	0.33 (1.46)	0.17 (1.01)	0.01 (0.06)
t+12	0.30* (1.91)	-0.03 (-0.21)	0.06 (0.48)	-0.16 (-0.79)
t+13	0.44** (2.19)	0.45** (1.99)	0.03 (0.17)	0.04 (0.31)

注: ***, **, *分别表示在 1%、5%、10%水平上显著; 括号内的 t 值经过 Newey-West 调整。

3.4 金融社交媒体信息反映基本面信息了吗？

上文中,我们通过分析金融社交媒体信息 FI 对股票横截面收益率的预测力给出了金融社交媒体信息蕴含股票基本面信息的间接证据.接下来,本文通过检验 FI 与公司经营表现之间的联系,给出 FI 蕴含股票基本面信息的直接证据.参考 Daniel 等的做法,我们选取两种的指标代表公司的经营表现变化以度量股票的基本面信息,分别是 CAR 和 SUE.CAR 使用市场交易数据构建; SUE 则由财务指标数据所构建.通过市场交易数据和财务数据这两个维度构造公司经营表现变化,可以让我们

尽可能完全地捕捉到股票的基本面信息.在 t 月我们分别按照 CAR 和 SUE 将股票分成十组, 低的组为 D1, 高的组为 D10, 再计算每个组的 t 月平均 FI 值.为了不使每个组中样本过少, 类似的, 我们也尝试了五分组法, 低的组为 Q1, 高的组为 Q5. 两种方法的结果分别在表 6 的 Panel A 和 B 中展现.

金融社交媒体信息蕴含股票基本面信息了吗? 从表 6 的结果来看, 答案是无疑的.横向比较 Panel A 中的 D10-D1 结果, 可以发现 CAR 与 SUE 的数值分别为 0.73% (Newey-West t 值为 4.88) 和 1.71% (Newey-West t 值为 5.89), 证明 FI 分别与 CAR 和 SUE 高度正向相关.类比来看, 将分组方式替换为五分组法也可以明确发现 CAR 和 SUE 分别与 FI 间存在的正向关系.由此, 本文证明了金融社交媒体信息能够捕捉到公司经营表现, 即股票基本面信息.

表 6 金融社交媒体信息与股票基本面信息

	CAR	SUE
PanelA: 十分组法		
D1	31.25*** (28.87)	30.60*** (31.58)
D10	31.99*** (29.52)	32.31*** (28.21)
D10-D1	0.73*** (4.88)	1.71*** (5.89)
PanelB: 五分组法		
Q1	31.53*** (30.64)	30.85*** (31.45)
Q5	32.22*** (30.71)	32.24*** (28.16)
Q5-Q1	0.69*** (6.44)	1.39*** (5.61)

注: ***, **, *分别表示在 1%、5%、10%水平上显著; 括号内的 t 值经过 Newey-West 调整。

总之, 以上所有发现都与本文的理论假设相吻合, 金融社交媒体信息捕捉到了股票的基本面信息, 由于存在套利成本限制, 导致市场对这类信息反馈缓慢, 导致了金融社交媒体信息能够预测股票横截面收益率.因此金融社交媒体信息有助于投资者和监管层的优化自身的各类决策部署, 增强市场层面信息传递的准确性.

3.5 进一步的异质性检验（对原论文的创新点）

理性预期均衡(REE)模型是金融市场微观结构理论研究的基础框架。根据该理论, 若价格是市场中分散信息的充分统计量, 即价格能够精确地反映公司信息, 则该均衡是强有效的(Fama,1970;Grossman,1978);若理性预期均的市场不是完全有效的(如市场存在噪音或经济中存在不确定性),则称均衡为弱有效的。弱有效理性预期均衡主要讨论市场微观主体的学习行为及其对金融市场的影响(Crossman 和 Stiglitz,1980;Admati,1985)。在这一理论框架下, 微观主体的私有信息可以通过交易行为影响资产价格, 市场均衡价格加总投资者的私有信息, 形成关于资产基本面的模糊信号。因此, 资产价格的波动很大程度上反映了知情交易者将私有信息注入价格的过程(Mcinish 和 Wood, 1992)。

为了识别社交媒体信息影响股票收益率的途径, 我们将阅读量(Read)和评论数(Comment)纳入了我们的分组分析中。根据投资者情绪理论来推测, 阅读量越高的帖子更能激发出投资者情绪, 因此对收益率的预测能力也应更强。但是, 基于行为金融理论的分析思路提供了另外一种推测。该理论认为投资者是有限理性的, 从而导致媒体报道对股票价格的影响不在于其为投资者提供了多少有用信息, 而在于其对投资者注意力的影响 (Barber 和 Odean, 2008)。例如, Odean (1999) 认为, 对于投资者而言, 特别是中小投资者, 由于受到搜寻成本和处理信息能力的约束, 他们不可能搜集到所有股票的信息, 而只能关注几支股票的信息, 从而倾向于购买能够吸引自己注意力的股票, 即导致所谓的“注意力驱动购买行为”(attention-driven

buying)。沿袭这一思路，我们认为阅读量和评论数背后反应的并不是信息量的大小，而是投资者（特别是中小投资者）的注意力的大小。较高的阅读量和评论数往往意味着投资者们较高的注意力，从而驱动这些投资者买卖这些股票，导致相关信息能够很快的反映到股价中。相反，那些不被关注的帖子所反应的信息在市场中将会以较慢的速度传导。

表 7 阅读量与评论量分组分析

	Read			Comment		
	RET	CH3	CH4	RET	CH3	CH4
L	1.90** (2.48)	1.03*** (4.47)	0.71** (2.51)	1.80** (2.40)	0.84*** (4.42)	0.54* (1.95)
2	1.61** (2.13)	0.71*** (4.17)	0.46** (2.05)	1.67** (2.28)	0.79*** (4.87)	0.57*** (2.84)
3	1.49* (1.93)	0.58*** (3.67)	0.35** (2.59)	1.46* (1.79)	0.47** (2.31)	0.27 (1.40)
4	1.15 (1.58)	0.25 (1.23)	-0.02 (-0.07)	1.04 (1.50)	0.13 (0.66)	-0.03 (-0.11)
5	0.94 (1.29)	0.06 (0.37)	-0.08 (-0.36)	1.10 (1.34)	0.22 (1.47)	0.11 (0.84)
6	1.24 (1.57)	0.47** (2.49)	0.35* (1.89)	0.84 (1.18)	-0.23* (-1.88)	-0.31** (-2.00)
7	0.83 (1.09)	-0.06 (-0.38)	-0.12 (-0.76)	0.86 (1.20)	0.13 (1.09)	0.06 (0.43)
8	0.95 (1.25)	0.18 (0.89)	0.10 (0.54)	0.82 (1.04)	0.13 (0.58)	0.04 (0.19)
9	0.47 (0.69)	-0.31** (-2.60)	-0.24** (-2.01)	0.53 (0.75)	-0.27*** (-2.63)	-0.23** (-2.23)
H	0.10 (0.14)	-0.42*** (-2.65)	-0.27 (-1.51)	0.12 (0.18)	-0.33** (-2.11)	-0.19 (-1.11)
H-L	-1.80*** (-3.83)	-1.45*** (-4.32)	-0.98** (-2.49)	-1.68*** (-3.51)	-1.17*** (-3.98)	-0.73* (-1.93)

注：***、**、*分别表示在 1%、5%、10%水平上显著；括号内的 t 值经过 Newey-West 调整。

使用与上文类似的分组分析，我们分别根据阅读量和评论数计算出每个组合在 t+1 月的超额总市值加权收益率和超额等权收益率，以及卖出低阅读量（评论数）组合（L）买入高阅读量（评论数）组合（H）构建的投资组合（H-L）的收益率，结果呈现于表 7。可以看出，根据股票反映基本面信息的理论，阅读量和评论数较大帖子已经及时、准确、充分地反映在对应公司的股价走势当中，而阅读量少的帖子尚未被股价完全反映，因此仍具有对股价的预测能力。这与投资者情绪理论推导的结果相悖，我们认为这与上文基于行为金融理论的推测相关。同时，这也再次证明了金融社交媒体信息的确含有基本面信息。

4 进一步分析：市场异象在不同金融社交媒体信息股票中的表现

在前文的检验中，已经看到金融社交媒体信息具有丰富股票基本面信息含量，这里本文主要分析投资者能否利用金融社交媒体信息进行交易以提高资本市场效率.已有文献发现大量能够预测股票收益率的变量（市场异象），并认为这些市场异象反映了价格与基本面的偏离，因此严重损害了市场效率.如果金融社交媒体信息能够有助于投资者正确认识价格与基本面的偏差，那么它将削弱市场异象.为此，本文检验市场异象在不同金融社交媒体信息 FI 股票中的表现，确认 FI 能够削弱市场异象的解释力。

行为金融学理论认为市场异象是资本市场效率缺失的直接体现.本文参考已有的研究成果，选取了 12 个中国市场最具代

表性的异象变量分别是 IVOL（特质波动率）、VOL（波动率）、MAX（最高收益率）、SKEW（偏度）、EP（市盈率倒数）、BM（账面市值比）、MIN（最低收益率）、TO（换手率）、TOVOL（换手率波动率）、TOMAX（最高换手率）、TOMIN（最低换手率）和 AGE（上市年龄）用以检验金融社交媒体信息对资本市场效率的影响.如果金融社交媒体信息能够提高资本市场效率，那么市场异象会在高 FI 股票中变弱，反之，如果金融社交媒体信息不利于资本市场，异象则会在高 FI 股票中变强.

在每个月末，首先，将股票按照 FI 以 30%和 70%为分位点分为三组，分别是 LFI（低 FI 组），和 HFI（高 FI 组）；其次，在每个组中将股票分别按照每个异象变量分成十组，分别计算每个组合的市值加权收益率；最后，分别在三个 FI 组中，根据每个异象构造出一个买入卖空的零成本投资组合（Long-Short）.表 8 中 LFI 代表了各个异象买入卖空零成本组合在低 FI 组中的表现，L-H 则代表了 LFI 和 HFI 组中异象买入卖空的零成本组合收益率之差。

表 8 金融社交媒体信息与市场异象

	Long-Short				Alpha			
	LFI	2	HFI	L-H	LFI	2	HFI	L-H
IVOL	-1.32***	-1.01*	0.37	-1.68***	-0.76	-0.37	1.07*	-1.83***
	(-2.89)	(-1.78)	(0.71)	(-4.56)	(-1.46)	(-0.43)	(1.94)	(-4.23)
VOL	-0.71	-0.76	0.68	-1.39***	-0.39	-0.32	0.70	-1.09***
	(-1.24)	(-1.31)	(1.05)	(-4.84)	(-0.81)	(-0.38)	(1.21)	(-3.14)
MAX	-0.60	-0.64	0.23	-0.83*	-0.33	-0.19	0.15	-0.48
	(-1.11)	(-1.27)	(0.58)	(-1.94)	(-0.62)	(-0.30)	(0.19)	(-0.77)
SKEW	-1.04***	-0.40	0.36	-1.40***	-0.88**	-0.07	0.74	-1.62***
	(-3.31)	(-1.26)	(0.96)	(-5.27)	(-2.15)	(-0.17)	(1.19)	(-3.44)
EP	0.96*	0.47	0.29	0.68*	0.44	-0.53	-0.61*	1.05**
	(1.80)	(0.99)	(0.47)	(1.73)	(1.45)	(-1.63)	(-1.96)	(2.51)
BM	0.51	-0.38	-0.12	0.63	-0.62	-1.52*	-1.44*	0.81
	(0.74)	(-0.37)	(-0.16)	(0.82)	(-0.73)	(-1.77)	(-1.75)	(1.21)
MIN	0.24	-0.33	-0.19	0.43	0.16	-0.55	-0.15	0.31
	(0.45)	(-0.60)	(-0.34)	(1.05)	(0.37)	(-0.97)	(-0.29)	(0.81)
TO	-1.10**	-0.37	-0.32	-0.78*	-0.75	-0.46	-0.31	-0.44
	(-2.11)	(-0.71)	(-0.54)	(-1.70)	(-1.50)	(-0.65)	(-0.42)	(-0.84)
TOVOL	-1.17*	-0.65	-0.52	-0.65*	-1.06**	-0.79	-0.53	-0.54
	(-1.92)	(-1.17)	(-0.85)	(-1.67)	(-2.56)	(-1.45)	(-0.83)	(-1.18)
TOMAX	-1.04*	-0.51	-0.81	-0.23	-0.86**	-0.60	-0.75	-0.11
	(-1.85)	(-0.95)	(-1.36)	(-0.59)	(-2.21)	(-1.16)	(-1.23)	(-0.23)
TOMIN	-0.77	-0.23	-0.27	-0.50	-0.45	-0.26	-0.27	-0.19
	(-1.44)	(-0.46)	(-0.44)	(-0.98)	(-0.88)	(-0.38)	(-0.39)	(-0.31)
AGE	0.63*	0.45	-0.87**	1.50***	0.21	-0.59*	-1.41***	1.62***
	(1.95)	(1.00)	(-2.10)	(2.97)	(0.69)	(-1.68)	(-2.85)	(3.05)

注：***、**、*分别表示在 1%、5%、10%水平上显著；括号内的 t 值经过 Newey-West 调整。

从表 8 左侧（Long-Short）的结果可以看出，异象在 LFI 组中的表现更加显著，基于异象构造的买入卖空组合（Long-Short）收益率都要明显高于 HFI 组中的对应结果，后者中的所有异象买入卖空组合都无法取得显著收益.观察组合 L-H 列的结果能够更为直观地发现异象获利与 FI 之间的联系，LFI 组中的异象获利与 HFI 组中的异象获利之差（L-H）几乎都显著为正.同样的，经过 CH4 因子模型调整后，组合 L-H 收益率（alpha）表现也都显著为正（见表 8 右侧）8，代表市场异象随着 FI 上升逐渐减弱.总结来看，我们发现金融社交媒体信息在削弱市场异象，意味着投资者能够利用金融社交媒体信息套利交易市场异象，这一实证结果证实了金融社交媒体信息可以起到提升资本市场效率的作用。

5 结论与启示

本文首次构造金融社交媒体信息指标 (FI)，通过在市场微观层面探讨金融社交媒体信息对个股价格的影响，揭示金融社交媒体对资本市场的影响。分析结果表明金融社交媒体信息指标 FI 蕴含一定信息，这种信息有助于预测股价。借助于金融社交媒体信息，投资者能够交易获利并推动市场理性，从而带来资本市场定价效率的提升。具体而言，本文发现：第一、金融社交媒体信息指标 FI 能够准确捕捉到股票基本面信息，因此可以正向预测股票横截面收益率。FI 的预测力表现持久稳健且无法被任何已有因子所解释。第二、金融社交媒体信息的预测力与套利成本有关。在难以套利的股票中，金融社交媒体信息的预测力更强，随着套利成本减弱，金融社交媒体信息的预测力也随着变弱。另外，FI 的预测力随期限增长而逐渐减弱但并不存在反转，表明金融社交媒体信息没有涵盖投资者情绪。第三、金融社交媒体信息有助于提高资本市场效率。使用市场异象刻画资本市场效率，本文发现在金融社交媒体信息消极（低 FI）的股票中市场异象显著增强，而在金融社交媒体信息积极（高 FI）股票中市场异象消失，异象的表现与 FI 呈现单调下降递减特征，表明投资者可以利用金融社交媒体信息准确识别市场异象并通过交易将之消除，进而带来资本市场效率的提升。为了保证上述的结果可靠性，我们进行了如下的稳健性检验：（1）排除公司层面投资者情绪和关注度的影响；（2）排除市场层面投资者情绪的影响；（3）排除市值因素的干扰；（4）剔除小市值股票；（5）剔除流动性差的股票；（6）WLS；（7）子样本检验；（8）时间维度上的进一步分析。一系列稳健性检验都证明本文的发现稳健。

本文的发现具有以下几点启示：第一，进一步提升金融社交媒体信息的有效性。金融社交媒体信息的预测力实质上体现了信息经由金融社交平台进入股票价格的过程。细节上，有两个关键点值得深入讨论，一是信息的可靠性，二是信息进入股价的高效性。针对第一点，及时防范注意互联网平台发布信息的真实性，通过排除虚假信息的干扰并及时更新有效信息，完善金融社交平台的信息披露功能，有助于投资者掌握可靠的信息。针对第二点，完善资本市场的套利机制，一方面通过完善各类如融资融券在内的套利工具，另一方面引入更多的如对冲基金在内的套利交易者，通过完善的套利手段加快信息在市场上的传播速度，以便信息更快速地进入股价。第二，有效的信息传播中介有助于缓解资本市场内信息不对称问题，减弱股票市场与经济基本面的偏离，从而强化资本市场对实体经济的服务功能。因此，进一步推动互联网建设、加速互联网与资本市场融合建设，不仅可以促进资本市场的高质量发展，同时也有助于中国经济的高质量增长。第三，阅读量较少或者尚未被广泛传播信息的股票评论仍然具有一定的股价预测能力，这启示我们投资者通过社交媒体不仅仅是信息传递，而且可以进行信息交互和学习，并提高投资者信息获取与信息处理能力，从而提高市场的定价效率和股价预测能力。

参考文献：

- [1] 郑建东,吕晓亮,吕斌,郭峰. 社交媒体平台信息交互与资本市场定价效率——基于股吧论坛亿级大数据的证据[J]. 数量经济技术经济研究, 2022, 39(11): 91-112.
- [2] 罗进辉,蔡地. 媒体报道能够提高股价的信息含量吗?[J]. 投资研究, 2013, 32(05): 38-53.
- [3] Liu J, Stambaugh R F, Yuan Y. Size and value in China[J]. Journal of financial economics, 2019, 134(1): 48-69.