

doi: 10.19920/j.cnki.jmsc.2021.05.003

股吧个体信息交互对股价联动关系的影响研究^①

陈张杭健¹, 吴 粤², 李世炳³, 任 飞^{1*}

(1. 华东理工大学商学院, 上海 200237; 2. 华东理工大学信息科学与工程学院, 上海 200237;
3. “中央研究院物理研究所”, 中国台北 11529)

摘要: 以往股价联动的影响因素研究主要基于社会嵌入理论, 认为机构持股等上市公司社会关系的信息扩散功能是重要因素之一。但中国沪深股票市场中小投资者居多, 个体投资者在社交媒体上的信息交互行为对股市联动性的影响日益凸显。采用东方财富网股吧用户数据, 基于有效信息理论和社会嵌入理论, 深入探讨微观个体行为对股价联动的影响。研究结果表明: 在基于信息交互行为构造的信息扩散网络中, 连边的权重越大, 对应股票间的价格相关系数越大; 社交媒体中个体大量、频繁的发帖或回帖等信息交互行为, 可以促进股票信息扩散, 从而影响股价联动关系; 此外, 个体的信息流量对股价联动具有预测作用。上述结论表明基于个体信息交互行为的股票信息扩散, 是影响股价联动的另一重要因素。该结论有助于深化对股票市场风险传导根本原因的认识, 为我国沪深股票市场的风险管理提供理论指导。

关键词: 股吧; 信息扩散; 信息交互; 股价联动; 复杂网络

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2021)05-0047-23

0 引 言

近年来, 随着计算机和互联网通信技术的快速发展, 股吧等互联网平台开始成为投资者获取信息和交互信息的重要媒介之一^[1]。股票市场风险事件经过股吧等互联网平台上个体间的信息交互得以迅速传播, 并可能引发股价的大范围、快速联动甚至金融危机。尤其在以中小投资者为主导力量的中国沪深股票市场上, 个体信息交互给市场带来的冲击更为明显。例如, 2015 年 12 月 10 日, “复星集团董事长郭广昌被警察带走”的消息在社交媒体上扩散, 次日复星系控股、参股的公司股价全线下跌^②。2016 年 2 月 24 日 13 点 21 分, 一则关于“创业板全面停止审核”的虚假信息在新浪微博发布并扩散后, 直接导致第二天

沪深两市现千股跌停^③。此类事件说明, 投资者对风险事件的反应和传播速度大大加快, 股票市场的风险传导已经发生深刻的变化, 互联网背景下股价联动关系的研究已十分必要。

风险传导研究对于认识复杂系统的风险演化规律和防范金融风险蔓延具有重要意义。股票市场风险传导通常表现为股价波动的蔓延, 即不同市场或上市公司间的股价出现联动^[2, 3]。针对股价联动的决定因素研究, 现代金融理论主要基于经济基本面进行解释, 但是大量研究表明, 经济基本面因素无法完全解释市场内不同证券的过度联动^[4]。最近, 有文献开始关注基于互联网信息的股价波动或股价联动关系研究。从信息本身的角度出发, 现有文献探讨了互联网信息的情感倾

① 收稿日期: 2019-04-08; 修订日期: 2019-11-18.

基金项目: 国家自然科学基金资助项目(71790594; 71871094); 教育部人文社会科学研究资助项目(17YJAZH067).

通讯作者: 任 飞(1980—), 女, 浙江杭州人, 博士, 教授, 博士生导师. Email: fren@ecust.edu.cn

② 新闻来源于网易新闻: http://news.163.com/15/1214/00/BAONDIP300014AED_mobile.html

③ 新闻来源于新华网: http://www.xinhuanet.com/fortune/2016-02/26/c_128753608.htm

向^[5,6]、数量^[7]、内容^[8]和质量^[9]等对股价波动或股价联动性的影响;从信息扩散的角度出发,由于成熟市场主要以机构投资者为主导,已有文献大多关注成熟市场上机构投资者之间的信息扩散与股价联动的关系^[10-14],但在中国沪深股票市场上个体投资者居多,却鲜有文献探究互联网平台上微观个体间的信息交互行为对股价联动的影响。

与上述文献不同的是,本文利用东方财富网股吧用户数据,构建股票信息扩散网络,分析了股票信息扩散网络与股价联动网络的一致性,研究发现同时活跃在两个股吧中的用户越多,对应股票间的相关系数越大,并且这个结果在考虑样本区间和其它控制变量等因素的影响后依旧存在。此外,在上述结论的基础上进一步考察了此类用户的微观行为对股价联动的影响,研究结果显示,同时活跃在不同股吧中的用户可以通过信息交互促进股票信息扩散,从而影响股价联动关系,并且用户产生或引起的信息流量能够预测股价联动。

本文有如下贡献:第一,首次利用互联网平台数据,构建股票信息扩散网络,发现信息扩散对股价联动存在显著正影响,是对该领域研究的重要拓展;第二,以信息扩散为桥梁,发现个体信息交互行为能够通过促进信息扩散进而影响股价联动,丰富了以往关于股价联动成因探索的研究;第三,从个体信息流量的角度出发,对股价关联进行预测,证明了信息流量的预测性。本文的研究结论有助于揭示股票市场风险传导的根本原因,为监管层有效阻断风险事件进一步扩大、恶化提供了新的启示。

1 文献回顾与研究假设

股票市场风险传导问题是金融风险研究领域的重点课题,一直以来引起学术界的高度关注,风险传导问题研究对认识复杂金融系统的风险演化规律、加强市场风险管理等具有重要意义。随着互联网技术和行为金融学理论的发展,学者们开始基于有效信息理论和社会嵌入理论,从社交媒体信息和市场参与主体间的信息扩散等角度解

释资产价格联动,探究金融市场风险传导的影响因素。

1.1 有效信息理论的观点

市场有效信息理论将资产价格过度联动中,无法由经济基本面因素解释的部分归因于异质信息,且异质信息来源于行为层面和公司层面两类信息^[15]。

行为层面的资产价格联动研究,包括投资者关注和新闻热点信息等^[9]。Preis等^[16]利用谷歌搜索引擎上关键词的搜索量作为投资者关注度指标,研究投资者行为对市场的影响。刘海飞等^[9]通过投资者信息关注度、信赖度和更新频率等3层维度构建微博信息质量指标,研究发现社交网络信息质量与股价同步性存在关联关系。Turan等^[17]研究显示异常新闻流量会导致投资者对公司价值的判断产生分歧,从而影响股价波动。黄俊和郭照蕊^[18]研究表明,新闻信息对公司股价同步性的影响主要通过负面报道产生,而且网络新闻媒体报道也有助于公司股价同步性降低。

公司层面的资产价格联动研究,早期从社交媒体信息数量出发,认为上市公司在社交媒体上发布的信息数量越多,其披露的异质性信息越多,对股价同步性影响越大^[19]。胡军和王甄^[7]进一步指出股价同步性不仅取决于上市公司在社交媒体上披露的异质性信息多寡,还取决于投资者是否能准确理解此类信息,并研究发现此类信息主要通过分析师的信息解读作用进入股价。基于社交媒体信息内容,何贤杰等^[8]发现只有当上市公司透明度、在社交媒体上披露的信息质量较高时,公司层面的信息才能有效地通过社交媒体传递给投资者,从而影响股价同步性。

综上所述,现有文献从社交媒体信息的来源、数量和质量等角度,解释了行为层面和公司层面的异质信息对股价联动的影响,但是上述研究忽略了异质信息在社交媒体上的扩散对资产价格行为的影响。东方财富网股吧是国内最大的股票在线讨论社区,用户作为信息扩散的载体,可能同时活跃在不同的股吧中,将某一股吧内的信息扩散至另一股吧,另一股吧中其他个体基于共同活跃用户提供的相同信息,对于各自关注的股票可能做出相同的预期或决策,最终导致两个股吧对应股票的价格联动。同时活跃在不同股吧中

的用户越多,在股吧间扩散的信息越多且受众范围越广,即信息扩散越容易,进入股价的信息越多,对金融资产行为的影响可能越大.基于上述分析,本文提出第 1 个假设:

H1 股吧间的信息扩散越容易,对应个股间的价格联动效应越明显.

1.2 社会嵌入理论的观点

基于上市公司社会关系的资产价格联动研究的立论基础是社会嵌入理论,该理论认为企业的经营行为会受到其所嵌入社会关系的影响^[20].基金共同持股关系、大股东联结关系、董事联结关系和上市公司交叉持股关系等上市公司间的社会关系,具有资源价值和信息扩散功能,可以被看作上市公司的某类价值属性^[21].投资者会根据上市公司的此类属性进行分类投资,进而导致资产价格联动^[22].

现有文献大多从信息扩散的角度,实证分析了上市公司间的社会关系对资产价格联动的影响. Anton 和 Polk^[10]指出基金共同持股关系能够预测股票间的协方差.这种持股关系体现了群体内部的信息扩散,以及群体成员的协调和信息的影响^[23].董大勇等^[11]针对上市公司前 10 大股东的联结关系进行了研究,结果表明关联交易和信息扩散的存在导致股东联结下的股票存在价格联动.陆贤伟等^[21]考察了企业间董事联结关系对股价联动的影响,结果表明股价的联动性是由董事联结的内在资源价值和信息扩散功能所形成的.类似的研究还包括上市公司交叉持股网络等与股价联动网络的对比研究^[13],企业集团关系型交易对股价同步性的影响^[24].

综上,机构投资者、大股东和董事等市场参与者的持股关系或联结关系所具备的信息扩散功能,在资产价格联动中发挥了重要作用,但是上述文献尚未考虑到个体投资者在上市公司社会关系形成中的作用.中国沪深证券市场主要以个体投资者为主导,且随着社交媒体的迅猛发展,股民与网民加速融合^[25].社交媒体用户对于某类信息的交互,会引起更多投资者的关注,从而引起股票价格联动^[26],因为不同上市公司可能因该类信息而具有共同的价值属性,例如,“复星集团董事长郭广昌被警察带走”的消息引起复星系控股、参股的公司股价全线下跌.此外,社交

媒体用户数量越多、产生或引起的信息流量越多、互动越频繁,信息扩散越容易,个体能接受到的此类信息越多且扩散速度越快,对金融资产行为的影响可能越大.因此,本文在厘清信息扩散对股价联动影响的基础上,进一步深入到微观层面探究个体信息交互如何影响股价联动,提出第 2 个假设:

H2 个体信息交互行为有助于促进股吧间的信息扩散,进而影响股价联动关系.

2 研究设计

2.1 股票信息扩散与股价联动

2.1.1 股价联动网络构建

采用 2017 年股票日度对数收益率序列间的相关系数矩阵来刻画股价联动关系^[27],股价联动网络的构建参考黄玮强等^[28]的做法,具体步骤如下: 1) 计算股票收益率序列相关系数矩阵; 2) 为了便于分析股价联动网络的内在性质,将相关系数矩阵转化为距离矩阵^[29].股票 s_i 和 s_j 之间的距离的计算式为

$$DIS_{s_i s_j} = [2(1 - COR_{s_i s_j})]^{1/2} \quad (1)$$

式中 $COR_{s_i s_j}$ 为两者日度收益率序列的皮尔逊相关系数,相关系数越大,相应的距离越小; 3) 在距离矩阵基础上应用最小生成树(MST)^[30]和平面极大过滤图(PMFG)^[31]算法构建最终的股价联动网络. MST 要求图通过 $M-1$ 条边将 M 个股票节点连接在一起,且满足所选择的边距离之和最小且不出现环^[32]; PMFG 仅要求平面图上所有的边不能交叉,节点数量 M 和连边数量 Q 满足^[33] $M \geq 3$ $Q \leq 3M - 6$.

最终构建的股价联动网络 $S-S$,其模型为

$$NET_S = (S, E_{S-S})$$

其中 $S = \{s_1, s_2, \dots, s_n\}$ 为不同股票集合; $E_{S-S} = \{(s_i, s_j) | s_i, s_j \in S\}$ 表示该网络连边的集合. (s_i, s_j) 为股票 s_i 和 s_j 之间的连边,表示股票 s_i 和 s_j 存在关联关系,连边的权重为相关系数.

2.1.2 信息扩散网络构建

在以个体投资者居多的中国沪深股票市场上,个体投资者间的信息交互对股价联动的影响可能更明显,股吧用户数据较好地呈现了个体投

投资者间的信息交互情况，例如，股吧用户可以通过发帖或回帖进行信息交互，且同时活跃在两个股吧中的用户越多，信息越容易在不同股吧间的扩散。因此，本文利用东方财富网股吧用户数据，基于个体信息交互行为构建如下股票信息扩散网络：

按不同个股股吧间共同活跃用户的数量，形成股票信息扩散网络 $G-G$ ，其模型为

$$NET_G = (P, E_{G-G})$$

式中 E_{G-G} 表示该网络连边的集合， $E_{G-G} = \{(g_k, g_l) \mid g_k, g_l \in G\}$ ， $G = \{g_1, g_2, \dots, g_n\}$ 为不同股吧的集合。若某个用户与股吧 g_k 和 g_l 均存在链接，则股吧 g_k 和 g_l 存在连边 (g_k, g_l) ，且此类用户越多，连边 (g_k, g_l) 的权重越大，以下简称为股吧对的权重。例如，在图 1 的个股信息扩散网络中，用户 p_1 和 p_2 均在股吧 g_1 和 g_3 中发帖，与股吧 g_1 和 g_3 建立直接链接，用虚线表示，则股吧 g_1 和 g_3 存在连边，且连边的权重为 2；用户 p_2 通过回复用户 p_4 在股吧 g_2 中的帖子，与股吧 g_2 建立间接链接，用点划线表示，同时该用户也在股吧 g_4 中的发帖，因此股吧 g_2 和 g_4 存在连边，但连边的权重为 1。计算股吧共同活跃用户数量矩阵，同样采用 MST 和 PMFG 算法构建最终的股票信息扩散网络。

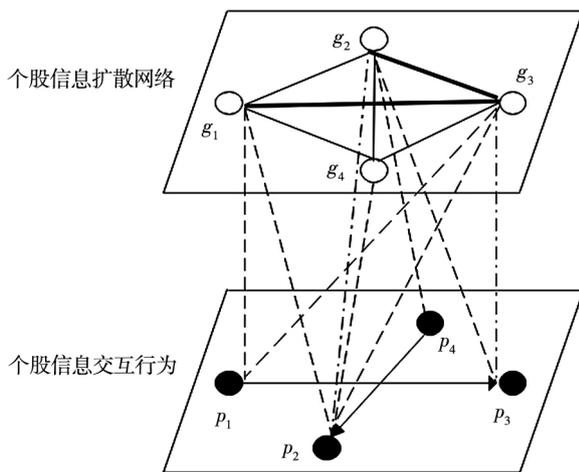


图 1 股票信息扩散网络构建

Fig.1 Construction of stock information diffusion network

2.1.3 回归模型构建

为了验证 H1，以股吧的共同活跃用户数量矩阵作为表征个股间信息扩散的指标，用股票相关系数矩阵 (COR) 作为刻画股票间联动关系的指

标，控制变量的选取参考董大勇等^[11]的做法，建立如下分析模型

$$COR = \alpha_0 + \alpha_1 STK + \alpha_2 DHY + \alpha_3 DMV + \alpha_4 DP + \alpha_5 DPE + \varepsilon \quad (2)$$

式中 STK 为共同活跃用户数量矩阵； DHY 为股票行业关系矩阵，其元素取值为 1 或 0，1 表示股票 s_i 和 s_j 属于同一行业，否则非同一行业； DMV 为规模差异矩阵，表示公司规模差异程度； DP 为价格差异矩阵； DPE 为市盈率相似性矩阵。

DMV 的元素 $DMV_{s_i s_j}$ 越小表明股票 s_i 和 s_j 的公司规模越接近，其取值为

$$DMV_{s_i s_j} = \left| \ln \left(\frac{MV_{s_i}}{MV_{s_j}} \right) \right| \quad (3)$$

式中 MV 为股票的流通市值。 DP 的元素 $DP_{s_i s_j}$ 越小表明股票 s_i 和 s_j 的市场价格越接近，其取值为

$$DP_{s_i s_j} = \left| \ln \left(\frac{PRI_{s_i}}{PRI_{s_j}} \right) \right| \quad (4)$$

式中 PRI 为股票的价格。

DPE 的元素 $DPE_{s_i s_j}$ 越大表明股票 s_i 和 s_j 的市盈率越接近，其取值为

$$DPE_{s_i s_j} = \begin{cases} \min \left(PEB_{s_i s_j}, \frac{1}{PEB_{s_i s_j}} \right) & PEB_{s_i s_j} > 0 \\ \max \left(PEB_{s_i s_j}, \frac{1}{PEB_{s_i s_j}} \right) & PEB_{s_i s_j} \leq 0 \end{cases} \quad (5)$$

式中 $PEB_{s_i s_j}$ 为股票 s_i 和 s_j 市盈率的比值，市盈率为每股价格 (PRI_{s_i}) 与每股收益 (EPS_{s_i}) 的比值。式 (2) 的参数估计，采用以重新抽样为基础的非参数检验 (QAP) 方法^[34]，该网络分析方法可以检验两种网络联结模式是否存在关联^[35]。

最后，为了验证同一股票在信息扩散和股价联动网络中是否均处于主导地位，以信息扩散网络中股票的中心度作为网络位置的替代指标，以该股票与其他股票平均相关系数衡量其在群体联动关系中的相对主导位置，并结合相关控制变量^[11]，建立如下回归模型

$$MCOR_{s_i} = \beta_0 + \beta_1 DEG_{g_i} + \beta_2 MV_{s_i} + \beta_3 PRI_{s_i} + \beta_4 EPS_{s_i} + \beta_5 PE_{s_i} + \varepsilon \quad (6)$$

式中 $MCOR_{s_i}$ 为股票 s_i 与其他股票相关系数的平均值； DEG_{g_i} 为与股票 s_i 相对应的股吧 g_i 在信息扩散网络中的度； MV_{s_i} 、 PRI_{s_i} 、 EPS_{s_i} 和 PE_{s_i} 分别

为股票 s_i 的流通市值、价格、每股收益和市盈率。

2.2 个体行为与股价联动

上述研究方法虽然有助于从中观层面理解信息扩散如何影响股票间的联动关系,但是依然未知信息扩散的微观原因,进而无法深入了解导致股票间出现联动现象的根本原因. 为了验证假设 2, 首先探究个体信息交互行为在股票信息扩散中的作用, 包括分析用户的信息流量特征以及这些信息流的时间间隔分布等; 其次, 分别利用个体产生和引起的信息流量作为信息扩散指标, 直接考察个体行为对股价联动的影响。

2.2.1 个体信息交互与信息扩散

考虑到股吧用户行为的多样性, 不同类型的用户可能对信息扩散的影响不同, 例如在低权重的股吧对中, 可能只发帖而没有回复的用户较多, 因为此类股吧可能极不活跃, 大部分帖子没有用户或很少有用户回复. 而在高权重的股吧对中, 用户较多且信息交互频繁, 只发帖无回复的用户可能较少, 大部分用户会通过发帖或回帖来参与信息交互. 因此, 有必要根据用户发帖和回帖的行为特征, 将共同活跃用户分为 B、C、D 3 种类型, 另外 C 类用户还可细分为 C1 和 C2 两种类型, 具体含义如图 2。

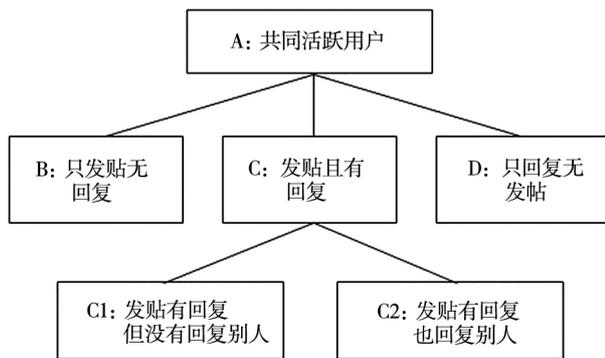


图 2 共同活跃用户分类

Fig. 2 Classification of common users

两个股吧中共同活跃用户的信息流量越多, 关于股票的异质性信息可能越多, 接受到此类信息的其他用户也可能越多, 共同活跃用户对股吧信息扩散的促进作用越大. 因此, 以两个股吧中 C、C1 和 C2 类用户的平均被回复次数度量 3 类用户的平均信息流量, 以 B 类用户的平均发帖数和 D 类用户的平均回复次数度量这两类用户的平均

信息流量, 计算公式如下

$$\overline{f_{-B}^{(g_k, g_l)}} = \frac{\sum_i^{n_B} ft_{-p_{i,B}^{(g_k, g_l)}}}{n_B} \quad (7)$$

$$\overline{f_{-C}^{(g_k, g_l)}} = \frac{\sum_i^{n_C} pf_{-p_{i,C}^{(g_k, g_l)}}}{n_C} \quad (8)$$

$$\overline{f_{-D}^{(g_k, g_l)}} = \frac{\sum_i^{n_D} hf_{-p_{i,D}^{(g_k, g_l)}}}{n_D} \quad (9)$$

式中 $\overline{f_{-B}^{(g_k, g_l)}}$ 为该股吧对中 B 类用户的平均发帖量; $ft_{-p_{i,B}^{(g_k, g_l)}}$ 为同时活跃在股吧 g_k 和 g_l 中的 B 类用户 $p_{i,B}$ 的发帖量; n_B 为 B 类用户的数量; $\overline{f_{-C}^{(g_k, g_l)}}$ 为该股吧对中 C 类用户的平均被回复次数; $pf_{-p_{i,C}^{(g_k, g_l)}}$ 表示 C 类用户 $p_{i,C}$ 在股吧 g_k 和 g_l 中所有帖子的被回复次数; $\overline{f_{-D}^{(g_k, g_l)}}$ 为该股吧对中 D 类用户的平均回帖次数; $hf_{-p_{i,D}^{(g_k, g_l)}}$ 为 D 类用户 $p_{i,D}$ 在股吧 g_k 和 g_l 中的回复次数. 同时, 利用式 (8) 计算任意股吧对中 C1 和 C2 类用户的平均信息流量。

为了进一步探究不同权重的股吧对共同活跃用户发帖或回帖时间间隔的分布特征, 首先根据股吧对的权重, 对连边进行降序, 并取排名在前 1% 和后 1% 的股吧对作为实验组(高权重组)和对照组(低权重组); 其次, 计算实验组和对照组中 B 类用户的发帖时间间隔、C 类用户发的每个帖子的回复时间间隔和 D 类用户在每个帖子中的回复时间间隔; 最后, 分析各类用户行为的时间间隔分布情况。

2.2.2 个体信息流量与股价联动

共同活跃用户在两个股吧中产生的信息流量是衡量股吧间信息扩散程度最直接的指标, 信息流量越多, 说明流动在两个股吧中的信息越多, 越有可能引起对应个股间的价格联动. 因此, 以两个股吧中共同活跃用户在每天产生的信息流量作为信息扩散指标, 包括共同活跃用户在两个股吧中的发帖数量和回帖数量; 其次, 为了考察日度信息流量对股价联动关系的影响及预测效果, 需要计算对应两个股票收益率的日度相关系数序列, 采用 DCC-GARCH 方法^[36]; 最后, 利用 2017 年每个交易日的截面数据进行 1 次回归分析, 考察

基于微观个体行为的信息扩散指标对股价联动的影响,回归模型如下

$$DCOR_{t, s_i s_j} = \gamma_0 + \gamma_1 FLOW_{t, g_k g_l} + \gamma_2 DHY'_{s_i s_j} + \gamma_3 DP'_{t, s_i s_j} + \gamma_4 DMV'_{t, s_i s_j} + \gamma_5 FLOW_{t, s_i s_j} \times DHY'_{s_i s_j} + \varepsilon \quad (10)$$

式中 $DCOR_{t, s_i s_j}$ 为第 t 天股票 s_i 和 s_j 对应的相关系数; $FLOW_{t, g_k g_l}$ 为当天对应两个股吧中共同活跃用户产生的信息流量; $DHY'_{s_i s_j}$ 为行业虚拟变量,表示股票 s_i 和 s_j 是否属于同一行业,是则取 1,否则取 0; $DP'_{t, s_i s_j}$ 为股票 s_i 和 s_j 的价格差异,利用式(4)采用当天的价格计算; $DMV'_{t, s_i s_j}$ 为股票 s_i 和 s_j 的规模差异,利用式(3)采用当天的流通市值计算。

考虑到信息流量可能对收益率序列间的相关性具有预测作用,利用面板数据,采用如下混合回归模型,考察信息流量的预测效果

$$DCOR_{s_i s_j}(t) = \psi_1 + \psi_2 \sum_{m=1}^T FLOW_{g_k g_l}(t-m) + \psi_3 DHY'_{s_i s_j} + \psi_4 DMV'_{s_i s_j}(t) + \psi_5 DP'_{s_i s_j}(t) + \varepsilon \quad (11)$$

式中 $FLOW_{g_k g_l}(t-m)$ 为股吧 g_k 和 g_l 滞后 m 期的信息流量。其它变量的定义与模型(10)相同。此外,共同活跃用户的帖子或回复还可能被其他非共同活跃用户回复,因此在共同活跃用户产生的信息流量的基础上,加上这一部分非共同活跃用户产生的信息流量来衡量由共同活跃用户引起的信息流量,记为 $FLOW'_{g_k g_l}$ 。 $FLOW'_{g_k g_l}$ 更加全面地反映了共同活跃用户带来的信息扩散效果,用其替换式(10)和式(11)中的信息流量变量,进一步考察由共同活跃用户引起的信息流量对股价联动性的影响。

3 实证分析与稳健性检验

3.1 数据说明

本文的研究区间为 2017-01-03 ~ 2017-12-29,研究对象为上证 A 股。由于采用 2017 年年度收益率序列的相关系数矩阵来度量股价联动,收益率序列长度有限,2017 年共有 244 个交易日,为避免当样本容量较大时,估计噪声起主导作用^[37, 38] 将样本容量控制在 300 个以内。个体

投资者的行为对不同市值股票的价格联动影响存在差异,且不同市值股票的波动率不同^[39],因此,在样本的选择过程中,首先根据上证 A 股 2017 年日度流通市值的均值,将其分为 3 组,在每一组内随机选择 100 支股票,共得到 300 支股票。其次,为了确保股票的流动性,剔除了 300 支股票中非连续交易的个股,最终,得到 267 支样本股。样本股行业分布情况如表 1 所示,按一级行业名称划分。

表 1 样本股行业分布

Table 1 Industrial sector distribution of sample stocks

行业名称	样本数量
制造业	131
批发和零售业	24
交通运输、仓储和邮政业	22
电力、热力、燃气及水生产和供应业	20
金融业	12
房地产业	14
采矿业	11
建筑业	7
农、林、牧、渔业	6
文化、体育和娱乐业	5
信息传输、软件和信息技术服务业	4
综合	4
水利、环境和公共设施管理业	2
住宿和餐饮业	2
租赁和商务服务业	2
教育	1

相关系数矩阵(COR)的计算采用 2017 全年日度对数收益率序列,行业关系矩阵(DHY)根据证监会《上市公司行业分类指引》2012 版计算,价格差异(DP)、规模差异(DMV)和市盈率相似性(DPE)矩阵分别采用 2016 年最后一个交易日的收盘价、流通市值和 2016 年年报中的市盈率计算。研究个体行为与股价联动关系时,变量 DP' 和 DMV' 分别采用 2017 年全年日度的收盘价和流通市值计算,数据均来自 Wind 数据库。股票信息扩散网络数据,来源于东方财富网个股股吧中个人投资者的发帖记录,通过 Python 编程采集,采集字段包括帖子 ID、用户 ID、吧龄、用户发帖和回帖的内容以及时间等,共收集了 2017 年 1 357 673 个用户的 7 636 107 条发帖或回帖记录,以及 2016 年 347 018 个用户的 2 707 045 条

发帖或回帖记录. 同时活跃在两个股吧中的用户为当年同时在这两个股吧中有发帖或回帖行为的用户, 以下简称共同活跃用户.

3.2 信息扩散与股价联动网络研究

中观层面的股票信息扩散与股价联动研究, 有助于理清互联网时代下的信息传播变革如何引起金融风险传导途径和方式的改变. 首先通过信息扩散网络与股价联动网络的形态和拓扑结构指标的描述性统计分析, 包括网络的度、介数、网络密度、平均路径长度等, 初步了解两层网络的结构. 进一步, 通过两层网络相关指标的差异性检验、度分布等方法, 对两层网络的一致性进行深入评价. 其次, 通过 QAP 和线性回归方法, 定量研究股票信息扩散对股价联动关系的影响.

3.2.1 两个网络拓扑结构指标对比

图 3 给出了基于 PMFG 算法构建的个股信息

扩散和股价联动网络.

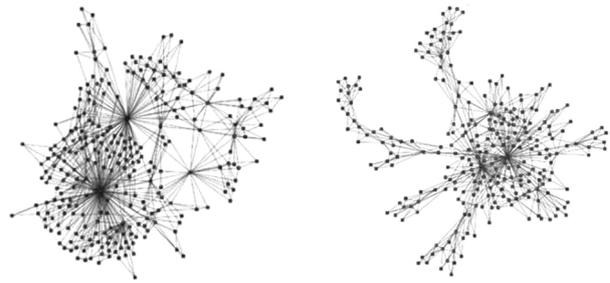


图 3 信息扩散(左)和股价联动(右)网络

Fig. 3 Information diffusion network (left) and stock return co-movement network (right)

表 2 为不同算法构建的两个网络的相关拓扑结构指标. 基于 PMFG 算法构建的信息扩散网络和股价联动网络在平均度、平均聚类系数和网络密度上表现出较强的一致性, 在平均介数和平均路径长度上, 股价联动网络稍高于信息扩散网络. 但基于 MST 构建的两个网络在平均介数和平均路径长度上相差较多.

表 2 网络拓扑结构指标对比

Table 2 Lists of topological structure parameters of networks

指标	PMFG		MST	
	信息扩散网络	股价联动网络	信息扩散网络	股价联动网络
平均度	5.955	5.955	1.993	1.993
平均介数	210.206	336.004	414.101	782.378
平均聚类系数	0.782	0.757	0.000	0.000
平均路径长度	2.580	3.526	4.114	6.883
网络密度	0.224	0.224	0.008	0.008

表 3 为差异性检验统计量, 从中可看出, 基于 PMFG 算法构建的两个网络, 仅节点度和介数分别在方差和中位数差异性检验上显著拒绝了原假设, 其它检验均接受了原假设, 即不存在差异; 但基于 MST 构建的两个网络, 仅节点度和节点介数分别在均值和方差差异性上接受了原假设.

由此可见, 基于 MST 算法构建的两个网络在网络结构上存在一定差异, 原因在于 MST 算法要求新加入的边与现有的边不出现环, 约束较为严格, 而 PMFG 仅要求最终形成的图是平面图即可, 最大限度的保留了距离网络中的有效信息^[40].

表 3 差异性检验结果

Table 3 Results of difference test

参数	PMFG		MST	
	节点度	节点介数	节点度	节点介数
均值差异性	-1.99×10^{-15} (1.000 0)	-0.901 5 (0.367 7)	9.81×10^{-15} (1.000 0)	-1.766 4* (0.077 9)
中位数差异性	0.924 1 (0.355 4)	2.220 7** (0.026 4)	2.406 7** (0.040 3)	2.479 8** (0.013 1)
方差差异性	2.010 9*** (0.000 0)	1.074 1 (0.560 2)	4.059 8*** (0.000 0)	1.223 5 (0.100 6)

注: 括弧内为 p 值, *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下拒绝原假设.

图4给出了信息扩散网络和股价联动网络的度分布,从中可以看出两者的节点度分布大体相同,分别服从指数为2.08和2.07的幂律分布,均具有明显的无标度特性,再次验证了两个网络在结构上的相似性,同时也表明两个网络中均存在少数关键节点对网络的运行起着主导作用,这一点从图3中也可以看出。

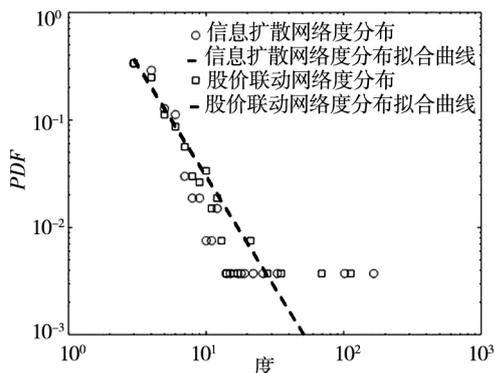


图4 网络的度分布

Fig. 4 Degree distribution of networks

上述结论表明,信息扩散网络与股价联动网络

在网络结构上呈现出一定的相似性,股票间的信息扩散在一定程度上能够解释股价的联动现象。

3.2.2 两个网络的定量研究

采用式(2)研究信息扩散对股价联动的影响,回归结果见表4,其中概率A为随机置换产生的回归系数的绝对值不小于观察到的回归系数的概率,概率B为随机置换产生的回归系数的绝对值不大于观察到的回归系数的概率。从表4可以看出,在5%水平上,股吧共同活跃用户数量矩阵对股票相关系数矩阵具有显著正影响,说明股吧间的共同活跃用户越多,股票间的价格联动效应越明显。此外,在1%水平上,行业关系矩阵对股票相关系数矩阵具有显著正影响,说明同行业股票间的价格联动效应更明显;在1%水平上,规模差异矩阵和价格差异矩阵对股票相关系数矩阵具有显著负影响,说明股票间的规模和价格差异越大,股票间的联动效应越小。市盈率相似矩阵与股票相关系数矩阵正相关,但并不显著。

表4 共同活跃用户数量矩阵与股价相关系数矩阵的QAP回归

Table 4 QAP regression of *STK* and *COR*

矩阵	回归系数		显著性 概率	概率	
	非标准化	标准化		A	B
<i>STK</i>	0.000 1 **	0.044 4 **	0.045 0	0.045 0	0.956 0
<i>DHY</i>	0.069 8 ***	0.080 6 ***	0.000 0	0.000 0	1.000 0
<i>DMV</i>	-0.097 6 ***	-0.534 2 ***	0.000 0	1.000 0	0.000 0
<i>DP</i>	-0.039 0 ***	-0.106 1 ***	0.000 0	1.000 0	0.000 0
<i>DPE</i>	0.006 1	0.016 3	0.290 0	0.290 0	0.710 0
常数项	0.349 9	0.000 0	—		
调整 R^2	0.324 0		0.000 0	—	

注: *、**和***分别表示系数在10%、5%和1%水平下显著。

针对式(6),分别采用MST和PMFG算法构造的信息扩散网络中的节点度值作为自变量进行回归分析,回归结果见表5。从表5中可以看出,不论是基于MST还是PMFG算法,信息扩散网络中股票的中心度对该股票与其他股票的平均相关系数均存在显著正影响,说明股票在信息扩散网络中越处于中心位置,该股票在群体联动关系中越处于主导地位。此外,流通市值和每股收益对平均相关系数存在显著负影响,说明小规模、业绩差的股票联动效应更大。

基于MST算法时,价格对平均相关系数有显著负影响且系数较大,而基于PMFG算法时,价格无显著影响且系数较小,可以认为低价股的联动效应更大。市盈率与平均相关系数正相关,但并不显著。

综上所述,信息扩散网络与股价联动网络存在较高的一致性,共同活跃用户数量对股票收益率序列间的相关系数存在显著正影响,这一结果验证了假设1,即两个股吧间的信息扩散越容易,对应个股间的价格联动效应越明显。

表 5 信息扩散网络中心度与股票平均相关系数的回归分析

Table 5 Regression analysis of *DEG* and *MCOR*

算法	<i>DEG</i>	<i>MV</i>	<i>PRI</i>	<i>EPS</i>	<i>PE</i>	常数项	调整 R^2
MST	0.774 *** (4.832)	-0.371 *** (-6.229)	-0.485 ** (-2.524)	-0.383 *** (-3.325)	0.021 (0.418)	2.27×10^{-10}	0.325
PMFG	0.225 ** (2.350)	-0.397 *** (-6.297)	0.100 (0.731)	-0.412 *** (-3.467)	0.012 (0.219)	2.96×10^{-11}	0.280

注: 括弧内为系数的 t 统计量的值; *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。

3.3 个体行为与股价联动研究

在上述研究的基础上, 首先从以下两点分析了个体信息交互对股票信息扩散的影响: 一方面, 用户数量和用户在股吧中产生或引起的信息流量多少决定了信息扩散的广度和深度^[41]; 另一方面, 若用户在短时间内频繁发帖或回帖, 说明股吧中用户信息交互的效率较高, 信息越容易在股吧间扩散。因此, 有必要对用户数量、用户的信息流量特征和这些信息流的时间间隔进行分析, 理清个体信息交互如何通过促进股票信息扩散, 进而影响股价联动。其次, 以个体产生或引起的信息流量作为衡量信息扩散的指标, 直接考察个体行为对股票间价格联动的影响。

3.3.1 不同类型个体数量及其信息流量分析

图 5 给出了不同权重的股吧对中, 各类共同

活跃用户数量趋势图。从图中可以看出, B 类用户最少且无明显趋势, 由于 B 类用户发布的帖子大多是在简单描述股票价格变动或其它与证券市场无关的信息等, 信息含量较低, 与宏观经济政策、市场或行业信息相关的帖子极少^④, 可能是导致其帖子没有回复的主要原因之一, 因此对股吧间的信息流动促进作用较弱。D 类用户数量最多, 其次是 C 类用户, 且两者均随权重的增加而上升, C 类用户既有发帖行为且其帖子也有回复, 而 D 类用户在不同股吧中均有回帖行为, 两者都有助于信息在不同股吧间的流动。此外, 由于股吧是供用户交流意见的互动型在线讨论社区, 回帖的用户多于发帖的用户, 这可能是导致 D 类用户数量高于 C 类用户的主要原因。C1 类用户的数量稍高于 C2 类用户, 两者均随权重的增加而上升。

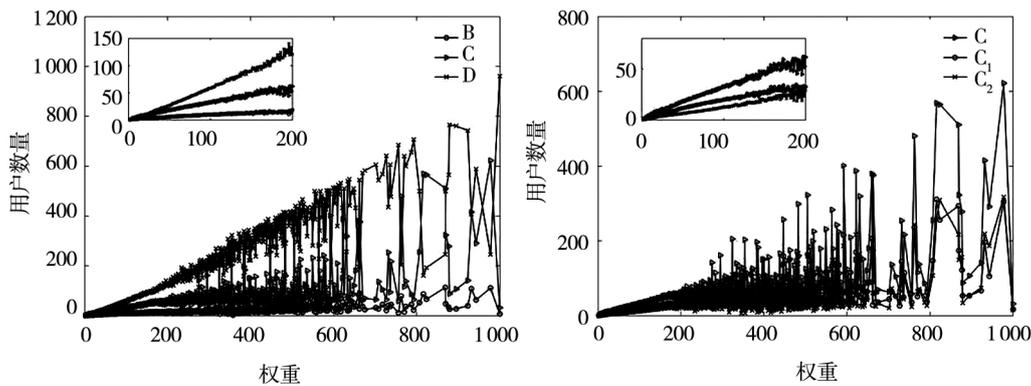


图 5 不同权重的股吧对中各类共同活跃用户数量

Fig. 5 Number of various common users in pairs of stock bars with different weights

图 6 给出了不同权重的股吧对中, B 类用户的平均发帖量、C 类用户的平均被回复次数以及 D 类用户的平均回复次数走势图, 这里统称为各

类用户的平均信息流量。从图中可以看出, C 类用户的平均信息流量明显高于 B 类和 D 类用户, 在权重大于 50 之后, 存在明显的上升趋势, 说明

④ 本文采用支持向量机训练分类模型将 B 类用户的所有帖子分为与宏观经济政策等信息相关和不相关两类, 发现相关的帖子仅占 0.04% 左右, 模型的预测准确率在 95% 左右, 限于篇幅原因未列出。

C类用户是促进股吧间信息流动的关键用户，并且C类用户在股吧中引起的信息流量越多，对信息流动的促进作用越强。从图6的右图中可以看出，在不同权重的股吧对中，C2类用户的平均信息流量均高于C1类用户，在权重大于50之后，两者均存在明显的上升趋势。与C1类用户的区别在于，C2类用户不仅发帖并有回复，且也回复

了自己或其他用户的帖子，是4类用户中最为活跃的用户，对信息扩散的作用最为显著。权重小于50时，C、C1和C2类用户引起的信息流量呈现先上升后下降的趋势，是由于不同股吧的活跃程度存在差异导致，在剔除活跃程度差异较大的股吧对后该趋势消失，并且在利用2016年的股吧数据做稳健性检验时，并没有出现类似趋势。

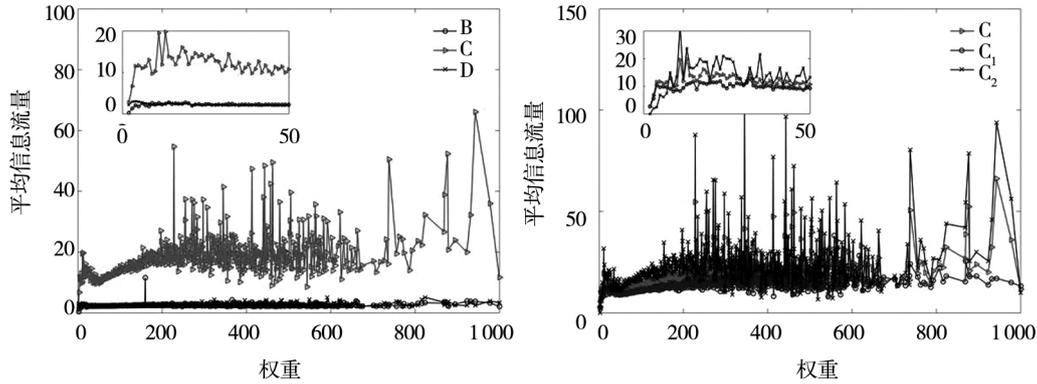


图6 不同权重的股吧对中各类共同活跃用户的平均信息流量

Fig. 6 Average information flows generated by various common users in pairs of stock bars with different weights

3.3.2 个体信息流的时间间隔分布

图7给出了实验组和对照组中各类用户行为的时间间隔分布情况，第1行为概率密度分布(PDF)，第2行为累积概率分布(CDF)。从概率密度分布上看，间隔时间在24h内，基本呈直线下降；24h后，C类用户和D类用户行为的时间间隔分布呈波浪式下降，且相邻的波峰与波谷之间大约相差12h，这是因为用户的发帖或回帖行为大部分集中于白天，尤其是交易时段，股吧内用户的

信息交互更为频繁，从而使用户行为的时间间隔呈现出周期性特征。从累积概率分布上看，高权重组中各类用户行为间隔时间的累积概率分布曲线均高于低权重组，以时间间隔小于1h为例，不论是共同活跃用户的发帖时间间隔(B类用户)还是回帖时间间隔(C类和D类用户)，实验组的概率均高于对照组，表明在共同活跃用户较多的两个股吧中，用户在短时间内发帖或回帖的概率更高，用户间的信息交互效率更高，信息扩散越快。

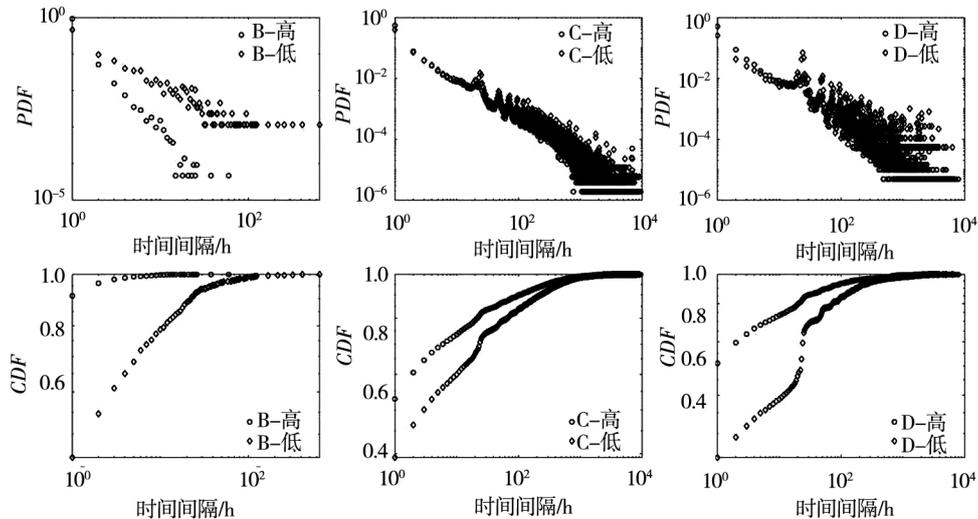


图7 各类共同活跃用户行为的时间间隔分布

Fig. 7 Time interval distribution of the behavior of various common users

3.3.3 个体信息流量对股价联动的影响

2017 年共 244 个交易日, 利用式(10)对每个交易日的截面数据进行回归分析, 表 6 统计了各自变量在 244 次回归分析中, 10% 水平下显著的正、负回归系数的占比. 其中, 模型 1 和模型 2 中的自变量是用户产生的信息流量, 模型 3 和模型 4 中的自变量是用户引起的信息流量.

从模型 1 的结果来看, 在 244 次回归中, 共同活跃用户产生的信息流量的回归系数大多显著为负, 占 71.31%. 但在模型 2 中, 进一步考虑了价格差异和规模差异因素后, 显著为正的的比例由 2.05% 上升至 27.05%, 而显著为负的比例从 71.31% 下降至 8.20%, 并且信息流量

与行业虚拟变量交互项的回归系数显著为正的的比例高达 94.26%. 上述结论表明, 对同行业股票而言, 信息流量对股价联动的显著正影响更明显, 对于非同行业股票, 这种显著正影响较弱. 价格差异和规模差异对股票相关系数的影响与上述 QAP 回归的结果一致. 由于共同活跃用户引起的信息流量更加全面的反映了股票间的信息扩散程度, 所以相较于模型 2, 模型 4 中 $FLOW'$ 的回归系数显著为正的的比例更高, 显著为负的比例更低. 整体来看, 信息流量与股票收益率序列间的相关系数基本呈显著正相关关系, 且对同行业股票来说, 这种显著的正相关关系更为明显.

表 6 式(10)中 10% 水平下显著的正、负回归系数占比 (%)

Table 6 Proportion of positive and negative regression coefficients at 10% significance level in Eq. (10) (%)

A 组: 共同活跃用户产生的信息流量						
模型		$FLOW$	DHY'	$FLOW \times DHY'$	DP'	DMV'
1	正	2.05	100	94.26	—	—
	负	71.31	0.00	0.00	—	—
	总和	73.36	100	94.26	—	—
2	正	27.05	100	89.75	2.05	0.00
	负	8.20	0.00	0.00	73.36	100
	总和	35.25	100	89.75	75.41	100
B 组: 共同活跃用户引起的信息流量						
模型		$FLOW'$	DHY'	$FLOW' \times DHY'$	DP'	DMV'
3	正	4.51	100	94.26	—	—
	负	72.13	0.00	0.00	—	—
	总和	76.64	100	94.26	—	—
4	正	34.43	100	94.67	1.23	0.00
	负	6.56	0.00	0.00	79.51	100

图 8 给出了模型 1 和模型 2 各自变量在 244 次回归中, 回归系数的概率分布. 对比发现, 模型 2 中信息流量的回归系数分布整体右偏, 说明在控制其它影响因素后, 信息流量对股票收益率序列间相关系数的显著正影响更大. 模型 1 中, 信息流量与行业虚拟变量交互项系数较大的概率稍高, 因为在加入控制变量后, 交互项的解释力稍有下降. 行业虚拟变量与交互项的回归结果与上述类似. 此外, 与价格差异相比, 规模差异对相关系数的影响更大.

表 7 给出了式(11)的回归结果, 其中, A 组中模型 5 ~ 模型 8 的自变量为共同活跃用户产生的信

息流量的滞后项, B 组中模型 9 ~ 模型 12 的自变量为共同活跃用户引起的信息流量的滞后项. 从回归结果中可以看出, 在 A 组中, 模型 7 的拟合优度最高且各阶滞后项均显著, 模型 8 的拟合优度有所下降且仅滞后 2 阶项显著, 可以认为模型 7 为最优模型; 在 B 组中, 在滞后 4 阶的范围内, 各滞后项的系数均在 10% 水平下显著, 但模型 11 的拟合优度最高. 上述结论表明, 信息流量确实能够预测股价联动. 此外, 由于共同活跃用户引起的信息流量更全面的反映了信息扩散程度, 所以相较于模型 7, 模型 11 的拟合优度更高, 并且由共同活跃用户引起的信息流量在滞后 4 阶后依然显著为正.

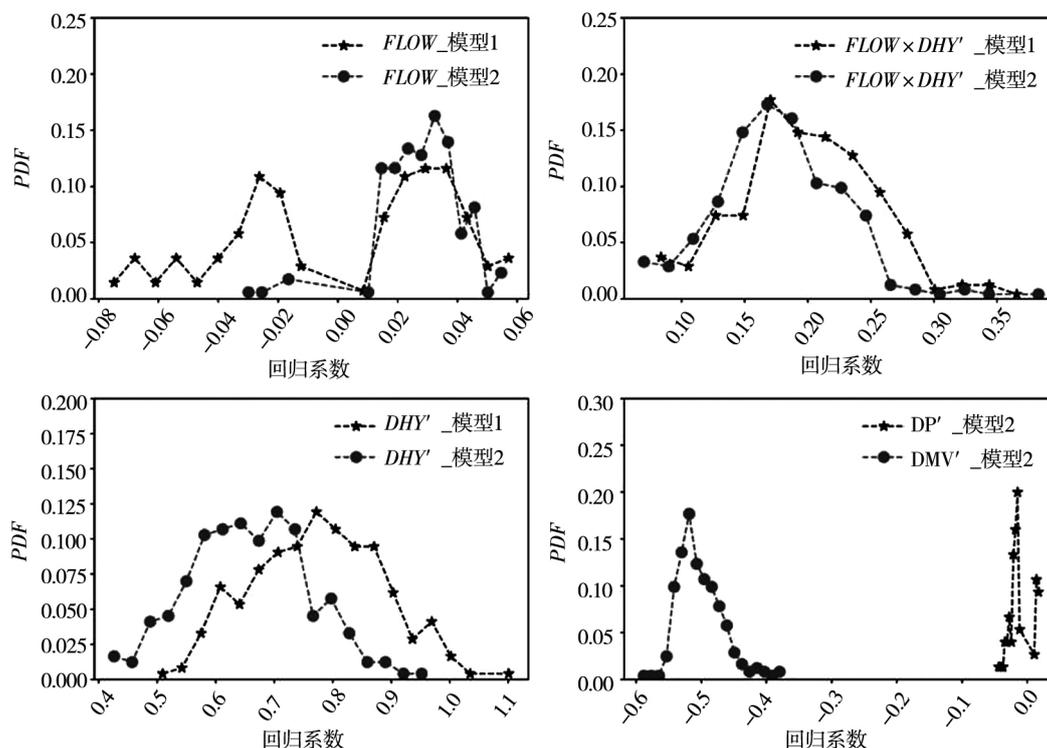


图8 式(10)中10%水平下显著的回归系数概率分布

Fig.8 Probability distribution of regression coefficients at 10% significance level in Eq. (10)

表7 信息流量对相关系数的预测回归

Table 7 Prediction regression of information flow to correlation coefficient

A组: 共同活跃用户产生的信息流量				
参数	模型			
	5	6	7	8
$FLOW(t-1)$	0.000 52*** (3.520 00)	0.000 34** (3.120 00)	0.000 31*** (3.010 00)	0.000 11 (1.170 00)
$FLOW(t-2)$	—	0.000 33*** (3.82000)	0.000 32*** (4.490 00)	0.000 50*** (5.370 00)
$FLOW(t-3)$	—	—	0.000 12* (1.790 00)	0.000 08 (1.160 00)
$FLOW(t-4)$	—	—	—	0.000 06 (0.930 00)
DHY'	0.069 05*** (13.600 00)	0.069 02*** (13.630 00)	0.069 10*** (13.680 00)	0.069 22*** (13.710 00)
$DP'(t)$	-0.017 01*** (-7.910 00)	-0.016 86*** (-7.860 00)	-0.016 75*** (-7.840 00)	-0.016 72*** (-7.830 00)
$DMV'(t)$	-0.099 82*** (-117.150 00)	-0.099 63*** (-117.030 00)	-0.099 21*** (-116.700 00)	-0.099 03*** (-116.440 00)
常数项	0.355 23*** (140.910 00)	0.354 11*** (140.780 00)	0.353 04*** (140.860 00)	0.352 81*** (140.950 00)
调整 R^2	0.258 78	0.259 56	0.261 20	0.258 58

续表 7

Table 7 Continues

B 组: 共同活跃用户引起的信息流量				
参数	模型			
	9	10	11	12
$FLOW'(t-1)$	0.000 53 *** (4.290 00)	0.000 30 *** (3.770 00)	0.000 22 *** (3.480 00)	0.000 17 *** (3.160 00)
$FLOW'(t-2)$	—	0.000 35 *** (4.550 00)	0.000 20 *** (3.940 00)	0.000 15 *** (3.740 00)
$FLOW'(t-3)$	—	—	0.000 30 *** (4.870 00)	0.000 18 *** (4.280 00)
$FLOW'(t-4)$	—	—	—	0.000 27 *** (5.060 00)
DHY'	0.068 70 *** (13.590 00)	0.068 51 *** (13.560 00)	0.068 39 *** (13.550 00)	0.068 30 *** (13.540 00)
$DP'(t)$	-0.017 07 *** (-7.950 00)	-0.017 07 *** (-7.950 00)	-0.017 06 *** (-7.940 00)	-0.017 05 *** (-7.940 00)
$DMV'(t)$	-0.099 78 *** (-117.180 00)	-0.099 85 *** (-117.050 00)	-0.099 89 *** (-116.940 00)	-0.099 93 *** (-116.840 00)
常数项	0.354 74 *** (140.920 00)	0.354 63 *** (140.820 00)	0.354 56 *** (140.740 00)	0.354 52 *** (141.680 00)
调整 R^2	0.258 12	0.259 62	0.261 78	0.258 11

注: 括弧内为系数的 t 统计量的值; *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。

综上所述, 针对个体微观行为的研究结论显示, 不同类型的共同活跃用户对信息扩散的促进作用不同, 频繁发帖或回帖的共同活跃用户是促进股吧间信息扩散的主力军; 此外, 共同活跃用户在两个股吧中产生的信息流量能够预测对应股票间的相关性。结合信息扩散与股价联动网络的研究结论, 上述结果表明个体信息交互能够通过促进股票信息扩散, 进而影响股价联动, 假设 2 得证。

3.4 稳健性检验

3.4.1 内生性问题

鉴于信息扩散与股价联动的关系存在一定的内生性问题, 即股票价格波动可能会影响共同活跃用户的发帖或回帖行为, 从而影响股吧间的信息扩散。因此, 采用两阶段最小二乘法 (2SLS) 对两者的关系进行进一步考察。参考何贤杰等^[8]的方法, 选取共同活跃用户的吧龄, 即用户的网站注册时长, 来考察用户特征对信息扩散的影响。用户吧龄与股价波动无关, 且吧龄越长, 作为股吧的老用户, 其对信息扩散的促进作用可能越强, 是解决内生性问题较好地工具变量^[42]。因

此, 在 2SLS 的第 1 阶段采用附录 A 中的式 (A1) 进行回归, 第 2 阶段采用表 7 中的模型 (5) 探究第 1 阶段拟合值对股价联动的影响, 回归模型和结果见附录 A。

附录 A 中的表 A1 报告了考虑内生性问题后共同活跃用户产生的信息流量对股价联动的影响, 从第 1 阶段的回归结果中可以看出, 在 1% 的水平下, 吧龄 (CY) 与信息流量显著正相关, 说明共同活跃用户的吧龄越长, 对信息扩散的促进作用越强。第 2 阶段的回归结果显示, 在考虑内生性问题后, 滞后 1 阶的信息流量对相关系数仍然存在显著正影响, 说明共同活跃用户产生的信息流量能够预测股票收益率序列间的相关系数, 并且在考虑内生性问题后结果依然稳健。此外, 还分别考察了滞后 2 阶和 3 阶信息流量, 在考虑内生性问题后, 对相关系数的影响, 结论依然稳健。

3.4.2 样本股筛选

由于在样本股的筛选过程中作了随机抽样, 可能导致实证结果存在一定的偶然性, 为避免该问题, 按照 3.1 节中的抽样过程进行多次取样,

并根据新样本对文章的主要结论再次进行验证. 附录 B 中的表 B1 给出了样本 1 ~ 样本 5 关于式 (2) 的回归结果, 其中样本 1 为上述 267 支样本股, 样本 2 和样本 3 均为重新随机抽样获取. 另外, 考虑到市值对结果的影响, 还从小市值组和高市值组分别随机抽取 100 支股票, 组成样本 4 和样本 5. 从表中可以看出, 样本 2 和样本 3 的回归结果与样本 1 基本一致, 共同活跃用户数量矩阵在 5% 的水平下依然显著为正. 样本 4 中, 共同活跃用户数量矩阵不显著, 而样本 5 中该变量在 1% 的水平下显著为正且系数较大. 附录 B 中的表 B2 为各个样本关于模型 (7) 的回归结果, 结果显示, 样本 2 和样本 3 的回归结果依然与样本 1 基本一致. 另外, 共同活跃用户产生的信息流量的各阶滞后项在样本 4 中均不显著, 而在样本 5 中均显著为正且系数较大.

小市值和高市值股票的结果存在差异的可能原因在于, 宏观或市场环境往往会对股吧用户的信息交互行为产生影响, 加速股票信息扩散, 且与小市值股票相比, 高市值股票受宏观或市场环境的影响更大^[19, 43]. 为了分析在剔除宏观或市场环境等因素的影响后, 信息扩散对小市值和高市值股票价格联动的影响, 采用五因子模型的残差计算相关系数矩阵和日度动态相关系数^[44, 45], 并对上述结果再次进行验证. 结果显示, 剔除后样本 4 与样本 5 的结果基本一致, 共同活跃用户数量矩阵均在 1% 的水平下显著为正, 共同活跃用户产生的信息流量的各阶滞后项也都显著为正^⑤. 综上, 可以认为文章的主要结论在重新筛选样本的情况下, 依然是稳健的.

3.4.3 样本区间

鉴于本文仅采用 2017 年股吧数据进行研究, 为了确保样本区间对研究结论的稳健性没有影响, 采用 2016 年股吧用户数据对本文的主要结论再次进行验证. 同时, 考虑到 2016 年中国股市波动较大, 为避免市场整体环境对主要结论的影响, 在计算股票相关系数矩阵和动态相关系数序列前, 将市场整体因素从股票收益率序列中剔除, 剔除方法^[39]见附录 C. 此外, 将 2017 年的数据剔除市场整体因素后, 上述结果基本不变.

附录 C 中的表 C1 给出了式 (2) 的回归结果, 结果显示股吧共同活跃用户数量矩阵的系数略有下降, 但对股票相关系数矩阵依然存在显著正影响, 行业虚拟变量、规模差异和价格差异对相关系数的影响依旧不变, 而且此时市盈率相似矩阵的系数由不显著变为在 5% 的水平下显著为正, 说明股票间的市盈率越接近, 股票间联动效应越大. 附录 C 中的图 C1 为 2016 年股吧中各类共同活跃用户的平均信息流量, C 类用户的平均信息流量依然最高且随权重的增加而增加. 此外, 在权重较小时, C、C1 和 C2 类用户的平均信息流量并没有出现先上升后下降的趋势. 附录 C 中的表 C2 给出了式 (11) 的回归结果, 相较于表 7 中 A 组的结果, 滞后 1 期至滞后 4 期的信息流量系数和显著性水平均有所提高, 再次说明共同活跃用户产生的信息流量能够预测股票收益率序列间的相关系数.

3.4.4 解释变量

虽然针对式 (2) 的回归结果已表明, 股吧间的共同活跃用户数量越多, 股价的联动效应越明显, 但可能会存在如下的情况: 股吧间共同活跃用户数量较多, 但是这些共同活跃用户产生的信息流量较少, 对股吧间信息扩散的促进作用较弱. 为了避免此类状况对回归结果造成影响, 统计了任意股吧对的共同活跃用户在 2017 年产生的所有信息流量, 包括此类用户在两个股吧中的发帖和回帖数量, 构成共同活跃用户信息流量矩阵 (CFW), 替代共同活跃用户数量矩阵, 进一步对本文的假设 1 进行验证. 从附录 D 中的表 D1 可以看出, 共同活跃用户信息流量矩阵在 5% 的水平下显著为正, 其它控制变量的显著性水平和系数正负依然保持不变. 此外, 上述结论在利用 2016 年数据再次验证后, 依然稳健.

3.4.5 其他控制变量

在控制行业关系、价格差异、规模差异和市盈率差异的基础上, 还考虑了两个股票动态相关系数的滞后 1 期、地域关系、财务杠杆比率差异、是否为指数成分股和超额交易量相关性等因素, 进一步验证本文的主要结论. 此外, 在信息流量对相关系数的预测回归中, 还控制了信息流的情感

⑤ 限于篇幅原因, 回归结果未列出.

倾向, 社交媒体情绪是影响股票市场的重要因素之一^[6], 共同活跃用户在两个股吧中呈现出不同的情感倾向, 可能会导致对应个股的价格反方向变动, 从而影响股价联动。因此, 计算了共同活跃用户在两个股吧中的情感倾向差异指标, 计算方法见附录 E。

从附录 E 中的表 E1 中可以看出, 在进一步控制地域关系、财务杠杆比率差异、是否为指数成分股和超额交易量相关性等因素后, 共同活跃用户数量矩阵依然在 10% 水平下对股票相关系数矩阵存在显著正影响, 模型的解释力从 0.342 上升至 0.452。从表 E1 中还可以看出, 超额交易量序列间的相关系数对股价相关系数存在显著正影响, 与 Landsman 和 Maydew^[46] 的结论一致; 财务杠杆差异的系数显著为负, 表明上市公司间的财务杠杆差异越大, 股价间的联动效应越不明显; DLIS 的系数显著为正, 说明指数成分股之间的联动效应更强; 股票间的地域关系对股价联动效应存在正影响, 但不显著。此外, 在以共同活跃用户的信息流量矩阵(CFW)为核心解释变量时, 上述结论依然稳健。附录 E 中的表 E2 为控制所有因素后, 包括滞后 1 期动态相关系数和共同活跃用户的情感倾向差异, 共同活跃用户产生的信息流量对动态相关系数的回归结果, 在 1% 显著性水平下, 滞后 2 期的信息流量对股价动态相关系数依然存在显著正影响且比较稳定。情感倾向差异在 1% 的水平下显著为负, 说明共同活跃用户在两个股吧中的情绪差异越大, 对应个股的相关系数越低, 同时也表明社交媒体的情绪确实对股票市场存在影响^[6]。

4 结束语

本文采用东方财富网股吧用户数据, 考察互联网信息扩散对股价联动性的影响, 并进一步深入到微观个体信息交互层面, 探究个体信息交互如何通过促进信息扩散进而影响股价联动。通过网络拓扑结构指标的对比研究, 发现基于个体信息交互行为的信息扩散网络与股价联动网络在网络结构上呈现出一定的相似性。定量研究结果表明, 同时活跃在两个股吧间的用户数量对股票收

益率序列间的相关系数存在显著正影响, 说明信息扩散越容易, 对应个股间的价格联动效应越明显; 信息扩散网络中股票的中心度对该股票与其他股票的平均相关系数存在显著正影响, 说明股票在信息扩散网络中越处于中心位置, 该股票在群体联动关系中越处于主导地位。基于微观个体信息交互行为的研究表明, 不同类型的共同活跃用户对股吧间信息扩散的促进作用不同, 频繁发帖或回帖的共同活跃用户是促进股吧间信息扩散的主力军; 此外, 共同活跃用户产生的或引起的信息流量对股票收益率序列间的相关系数存在显著正影响, 并具有预测作用。上述结论表明, 个体信息交互能够通过促进股票信息扩散, 进而影响股价联动, 并且这些结论在考虑内生性问题、样本选择、样本区间、解释变量和其他控制变量的影响后, 依然稳健。

本文的研究结论揭示了互联网平台的发展给投资者带来的信息交互、资源交换的便利, 这种便利促进了信息在不同股票间的扩散, 并最终通过交易影响资本市场。与现有文献相比, 本文深入到社交媒体上个体的信息交互行为, 探究信息扩散对股价联动的影响, 从更加微观的层面解释了金融资产行为的根本原因, 不仅是对风险传导影响因素的重要补充, 也是对有效信息理论和社会嵌入理论的重要拓展。

研究结果有助于深化对股票市场风险传导根本原因的认识, 对监管层加强股票市场风险管理有一定的借鉴意义。例如, 监管层可以重点监控互联网社交媒体平台中的关键用户, 如本文中在短时间内大量、频繁发帖或回帖的 C2 类用户, 以防止虚假信息等通过此类用户加速扩散。监管层还可以依据股吧信息流量的预测作用, 重点监控有异常信息流量的股吧, 防止个股风险通过信息扩散网络蔓延, 从而导致股价大范围、快速联动。当然, 本文的研究还有值得深入的地方, 例如利用投资者真实交易数据, 分析个体信息交互行为对投资者交易决策的影响, 厘清互联网信息扩散影响股市风险传导的渠道、机制; 探究不同市值股票对应的股吧中市场、行业和公司特征等各类信息的占比差异, 以及不同类型信息的扩散对不同市值股票的价格联动的影响是否存在差别。

参考文献:

- [1]于李胜,王成龙,王艳艳. 分析师社交媒体在信息传播效率中的作用——基于分析师微博的研究[J]. 管理科学学报,2019,22(7): 107-126.
Yu Lisheng, Wang Chenglong, Wang Yanyan. The role of analysts' social media in information dissemination: Evidence from analysts' microblogs [J]. Journal of Management Sciences in China, 2019, 22(7): 107-126. (in Chinese)
- [2]Forbes K J, Rigobon R. No contagion, only interdependence: Measuring stock market comovements [J]. The Journal of Finance, 2002, 57(5): 2223-2261.
- [3]裴 茜,朱书尚. 中国股票市场金融传染及渠道——基于行业数据的实证研究[J]. 管理科学学报,2019,22(3): 90-112.
Pei Xi, Zhu Shushang. Financial contagion in China's stock market: A study based on industry-level data [J]. Journal of Management Sciences in China, 2019, 22(3): 90-112. (in Chinese)
- [4]Tjostheim D, Hufthammer K O. Local gaussian correlation: A new measure of dependence [J]. Journal of Econometrics, 2013, 172(1): 33-48.
- [5]Steven C, Wesley G, Bryan R J, et al. What motivates buy-side analysts to share recommendations online? [J]. Management Science, 2018, 64(6): 2574-2589.
- [6]部 慧,解 峥,李佳鸿,等. 基于股评的投资者情绪对股票市场的影响[J]. 管理科学学报,2018,21(4): 86-101.
Bu Hui, Xie Zheng, Li Jiahong, et al. Investor sentiment extracted from internet stock message boards and its effect on Chinese stock market [J]. Journal of Management Sciences in China, 2018, 21(4): 86-101. (in Chinese)
- [7]胡 军,王 甄. 微博、特质性信息披露与股价同步性[J]. 金融研究,2015,(11): 190-206.
Hu Jun, Wang Zhen. Weibo, firm-specific information disclosure and stock price synchronicity [J]. Journal of Financial Research, 2015 (11): 190-206. (in Chinese)
- [8]何贤杰,王孝钰,孙淑伟,等. 网络新媒体信息披露的经济后果研究——基于股价同步性的视角[J]. 管理科学学报,2018,21(6): 43-59.
He Xianjie, Wang Xiaoyu, Sun Shuwei, et al. Economic consequences of new media information disclosure: From the perspective of stock price synchronicity [J]. Journal of Management Sciences in China, 2018, 21(6): 43-59. (in Chinese)
- [9]刘海飞,许金涛,柏 巍,等. 社交网络、投资者关注与股价同步性[J]. 管理科学学报,2017,20(2): 53-62.
Liu Haifei, Xu Jintao, Bai Wei, et al. Social networks, investor attention and stock price synchronicity [J]. Journal of Management Sciences in China, 2017, 20(2): 53-62. (in Chinese)
- [10]Anton M, Polk C. Connected stocks [J]. Journal of Finance, 2014, 69(3): 1099-1127.
- [11]董大勇,刘海斌,胡 杨,等. 股东联结网络影响股价联动关系吗[J]. 管理工程学报,2013,27(3): 20-26.
Dong Dayong, Liu Haibin, Hu Yang, et al. Do shareholders link network influence relation of stock price comovement [J]. Journal of Industrial Engineering and Engineering Management, 2013, 27(3): 20-26. (in Chinese)
- [12]马源源,庄新田,李凌轩. 股市中危机传播的 SIR 模型及其仿真[J]. 管理科学学报,2013,16(7): 80-94.
Ma Yuanyuan, Zhuang Xintian, Li Lingxuan. Susceptible-infected-removed (SIR) model of crisis spreading in the correlated network of listed companies and their main stock-holders [J]. Journal of Management Sciences in China, 2013, 16(7): 80-94. (in Chinese)
- [13]马丽莎,王建琼,董大勇. 交叉持股关系影响股价联动吗[J]. 财贸经济,2014,35(4): 59-68.
Ma Lisha, Wang Jianqiong, Dong Dayong. Does cross-holding affect share price linkage [J]. Finance & Trade Economics, 2014, 35(4): 59-68. (in Chinese)
- [14]肖欣荣,刘 健,赵海健. 机构投资者行为的传染——基于投资者网络视角[J]. 管理世界,2012,28(12): 35-45.
Xiao Xinrong, Liu Jian, Zhao Haijian. The contagion of institutional investor behavior: Based on the perspective of investor network [J]. Management World, 2012, 28(12): 35-45. (in Chinese)

- [15] Morek R , Yu W , Yeung B Y. The information content of stock markets: Why do emerging markets have synchronous stock price movements? [J]. *Journal of Financial Economics* ,2000 ,58(1/2) : 215 - 260.
- [16] Preis T , Moat H S , Stanley H E. Quantifying trading behavior in financial markets using google trends [J]. *Scientific Reports* ,2013 ,3: 1684.
- [17] Turan G B , Bodnaruk A , Scherbina A , et al. Unusual news flow and the cross section of stock returns [J]. *Management Science* ,2018 ,64(9) : 4137 - 4155.
- [18] 黄俊, 郭照蕊. 新闻媒体报道与资本市场定价效率——基于股价同步性的分析 [J]. *管理世界* ,2014 , (5) : 121 - 130.
Huang Jun , Guo Zhaorui. News media report and the pricing efficiency of capital market: Based on the analysis of stock price comovement [J]. *Management World* ,2014 , (5) : 121 - 130. (in Chinese)
- [19] 周冬华, 赵玉洁. 微博信息披露有利于降低股价同步性吗 [J]. *当代财经* ,2016 , (8) : 109 - 120.
Zhou Donghua , Zhao Yujie. Can microblogging information disclosure help to reduce stock price synchronicity [J]. *Contemporary Finance & Economics* ,2016 (8) : 109 - 120. (in Chinese)
- [20] 郑方. 治理与战略的双重嵌入性——基于连锁董事网络的研究 [J]. *中国工业经济* ,2011 (9) : 108 - 118.
Zheng Fang. Dual embeddedness of governance and strategy: Analysis based on interlocking directorate network [J]. *China Industrial Economics* ,2011 , (9) : 108 - 118. (in Chinese)
- [21] 陆贤伟, 王建琼, 董大勇. 董事联结影响股价联动: 关联分类还是资源价值? [J]. *证券市场导报* ,2013 , (10) : 47 - 54.
Lu Xianwei , Wang Jianqiong , Dong Dayong. The impact of director's network on prices comovement: Category-based investment or internal resource value [J]. *Securities Market Herald* ,2013 , (10) : 47 - 54. (in Chinese)
- [22] Barberis N , Shleifer A , Wurgler J. Comovement [J]. *Journal of Financial Economics* ,2005 ,75(2) : 283 - 317.
- [23] Powell W W , White D R , Koput K W , et al. Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences [J]. *American Journal of sociology* ,2005 ,110(04) : 1132 - 1205.
- [24] 李增泉, 叶青, 贺卉. 企业关联、信息透明度与股价特征 [J]. *会计研究* ,2011 , (1) : 44 - 51 + 95.
Li Zengquan , Ye Qing , He Hui. Inter-firm relationship , corporate transparency and stock price characteristics [J]. *Accounting Research* ,2011 , (1) : 44 - 51 + 95. (in Chinese)
- [25] 张永杰, 张维, 金曦, 等. 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 , et al. Does the Internet know more? Open source information and asset pricing [J]. *Systems Engineering: Theory & Practice* ,2011 ,31(4) : 577 - 586. (in Chinese)
- [26] Jiang L , Liu J , Yang B. Communication and comovement: Evidence from online stock forums [J]. *Financial Management* ,2019 ,48(3) : 805 - 847.
- [27] Khanna T , Thomas C. Synchronicity and firm interlocks in an emerging market [J]. *Journal of Financial Economics* ,2009 ,92(2) : 182 - 204.
- [28] 黄玮强, 庄新田, 姚爽. 中国股票关联网络拓扑性质与聚类结构分析 [J]. *管理科学* ,2008 , (3) : 94 - 103.
Huang Weiqiang , Zhuang Xintian , Yao Shuang. Topology property and clustering structure analysis of China's stock correlation network [J]. *Journal of Management Sciences* ,2008 , (3) : 94 - 103. (in Chinese)
- [29] Elton E J , Gruber M J. Improved forecasting through the design of homogeneous groups [J]. *Journal of Business* ,1971 ,44(4) : 432 - 450.
- [30] Onnela J P , Kaski K , Kertész J. Clustering and information in correlation based financial networks [J]. *The European Physical Journal B* ,2004 ,38: 353 - 362.
- [31] Tumminello M , Matteo T D , Aste T , et al. Correlation based networks of equity returns sampled at different time horizons [J]. *The European Physical Journal B* ,2007 ,55: 209 - 217.
- [32] Lee K E , Lee J W , Hong B H. Complex networks in a stock market [J]. *Computer Physics Communications* ,2007 ,177(1) : 186 - 186.
- [33] Tumminello M , Aste T , Matteo T D , et al. A tool for filtering information in complex systems [J]. *Proceedings of the National Academy of Sciences of the United States of America* ,2005 ,102(30) : 10421 - 10426.

- [34] Everett M. Social Network Analysis [M]. Essex: Textbook at Essex Sumer School in SSDA ,2002.
- [35] 奇达夫, 蔡文彬. 社会网络与组织 [M]. 北京: 中国人民大学出版社, 2007.
Qi Dafu, Cai Wenbing. Social Networks and Organizations [M]. Beijing: China Renmin University Press, 2017. (in Chinese)
- [36] Engle R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models [J]. Journal of Business & Economic Statistics, 2002, 20(3): 339 – 350.
- [37] Zhang M, Rubio F, Palomar D P, et al. Finite-sample linear filter optimization in wireless communications and financial systems [J]. IEEE Transactions on Signal Processing, 2013, 61(20): 5014 – 5025.
- [38] 李冰娜, 惠晓峰, 李连江. 基于蒙特卡洛 RMT 去噪法小股票组合风险优化研究 [J]. 管理科学, 2016, 29(2): 134 – 145.
Li Bingna, Hui Xiaofeng, Li Lianjiang. Research on risk optimization of small stock portfolio based on the filtering method of RMT using Monte Carlo simulation [J]. Journal of Management Science, 2016, 29(2): 134 – 145. (in Chinese)
- [39] Kumar A, Charles M C L. Retail investor sentiment and return comovements [J]. Journal of Finance, 2006, 61(5): 2451 – 2486.
- [40] 欧阳红兵, 刘晓东. 基于网络分析的金融机构系统重要性研究 [J]. 管理世界, 2014, 27(8): 171 – 172.
Ouyang Hongbing, Liu Xiaodong. Research on the systematical importance of financial institutions based on network analysis [J]. Management World, 2014, 27(8): 171 – 172. (in Chinese)
- [41] Yang J, Leskovec J. Modeling information diffusion in implicit networks [C]. IEEE International Conference on Data Mining, 2011: 599 – 608.
- [42] 周 静, 沈俏蔚, 涂 平, 等. 社交网络中用户关注类型与发帖类型对发帖行为的影响 [J]. 管理科学, 2019, 32(2): 67 – 76.
Zhou Jing, Shen Qiaowei, Tu Ping, et al. Influence of neighbor and posting type on users' posting behavior in social networks [J]. Journal of Management Science, 2019, 32(2): 67 – 76. (in Chinese)
- [43] 赵汝为, 熊 熊, 沈德华. 投资者情绪与股价崩盘风险: 来自中国市场的经验证据 [J]. 管理评论, 2019, 31(3): 50 – 60.
Zhao Ruwei, Xiong Xiong, Shen Dehua. Investor sentiment and stock price crash risk: Evidence from China [J]. Management Review, 2019, 31(3): 50 – 60. (in Chinese)
- [44] Fama E F, French K R. A five-factor asset pricing model [J]. Journal of Financial Economics, 2015, 116(1): 1 – 22.
- [45] Li J, Zhang Y J, Xu F, et al. Which kind of investor causes comovement? [J] Journal of International Financial Markets, Institutions and Money, 2019, 61(15): 1 – 15.
- [46] Landsman W R, Maydew E L. Has the information content of quarterly earnings announcements declined in the past three decades? [J]. Journal of Accounting Research, 2010, 40(3): 797 – 808.
- [47] 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.

Impact of interaction of individual stock bar information on stock price co-movement

CHEN Zhang-hang-jian¹, WU Yue², LI Shi-bing³, REN Fei^{1*}

1. School of Business, East China University of Science and Technology, Shanghai 200237, China;
2. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China;
3. “Institute of Physics, Academia Sinica”, Taipei 11529, China

Abstract: Previous research on the influencing factors of stock price co-movement, based mainly on the theory

of Social Embeddedness , suggested that the information diffusion function of listed companies' social relations , such as Institutional Ownership , is one important factor. There are a large number of individual investors in Chinese stock market , where the impact of information interaction among individual investors through social media on stock price co-movement is becoming increasingly prominent. Using the data from Eastmoney , this paper investigates the impact of individual investor behavior on stock price co-movement from a more microscopic perspective , based on the Effective Information Theory and the Social Embedding Theory. The empirical results show that the higher the weight of edges in the information diffusion network , which is based on information interaction , the greater the correlation coefficient between stock prices. Furthermore , massive and frequent posts or replies by users through social media can promote the spread of stock information , thus affecting stock price co-movement. In addition , the information flow generated by individuals has a predictive effect on stock price co-movement. The above results show that stock information diffusion through individual information interaction is another important factor affecting stock price co-movement. The conclusion is helpful in deepening the understanding of the underlying causes of risk conduction in the stock market and provides theoretical guidance for risk management in Chinese stock market.

Key words: stock bar; information diffusion; information interaction; stock price comovement; complex network

附录 A

$$FLOW_{g_k g_l}(t-1) = \omega_1 + \omega_2 CY_{g_k g_l} + \varphi_3 DHY'_{s_i s_j} + \omega_4 DMV'_{s_i s_j}(t) + \omega_5 DP'_{s_i s_j}(t) + \varepsilon \tag{A1}$$

$$DCOR_{s_i s_j}(t) = \psi_1 + \psi_2 \widehat{FLOW}_{g_k g_l}(t-1) + \psi_3 DHY'_{s_i s_j} + \psi_4 DMV'_{s_i s_j}(t) + \psi_5 DP'_{s_i s_j}(t) + \varepsilon \tag{A2}$$

式中 $CY_{g_k g_l}$ 为股吧 g_k 和 g_l 共同活跃用户的平均吧龄 , $\widehat{FLOW}_{g_k g_l}(t-1)$ 为式(A1)的拟合值 , 控制变量与式(11)中的相同.

表 A1 信息流量对相关系数的预测回归(2SLS)

Table A1 Prediction regression of information flow to correlation coefficient (2SLS)

变量	第 1 阶段	第 2 阶段
	$FLOW(t-1)$	$DCOR(t)$
$\widehat{FLOW}(t-1)$	—	0.001 5 *** (4.220 0)
CY	0.502 5 *** (87.770 0)	—
DHY'	1.088 9 *** (68.630 0)	0.067 9 *** (117.940 0)
$DP'(t)$	0.114 7 *** (14.210 0)	-0.017 1 *** (-94.840 0)
$DMV'(t)$	0.353 3 *** (122.720 0)	-0.100 2 *** (-641.270 0)
常数项	-1.468 1 *** (-57.720 0)	0.354 7 *** (1287.120 0)
调整 R^2	0.010 0	0.257 4

注: 括弧内为系数的 t 统计量的值, *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著.

附录 B

表 B1 为共同活跃用户数量矩阵与股价相关系数矩阵的 QAP 回归的标准化回归系数 表 B2 是信息流量对相关系数的预测回归结果.

表 B1 共同活跃用户数量矩阵与股价相关系数矩阵的 QAP 回归(重新抽样)
Table B1 QAP regression of *STK* and *COR* (the rebalance of sample stocks)

矩阵	随机抽样			小市值	高市值
	样本				
	1	2	3	4	5
<i>STK</i>	0.044 4 **	0.044 1 **	0.049 7 **	-0.057 4	0.097 0 ***
<i>DHY</i>	0.080 6 ***	0.060 5 **	0.052 3 **	-0.015 8	0.154 2 ***
<i>DMV</i>	-0.534 2 ***	-0.491 4 ***	-0.493 5 ***	-0.117 9 ***	-0.375 8 ***
<i>DP</i>	-0.106 1 ***	-0.087 0 ***	-0.084 0 ***	-0.215 8 ***	-0.092 3 ***
<i>DPE</i>	0.016 3	0.018 7	0.027 8	0.028 3	0.052 1 *
R^2	0.324 0	0.271 0	0.272 0	0.067 0	0.203 0
样本数	267	241	271	100	100

注: *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著.

表 B2 信息流量对相关系数的预测回归(重新抽样)

Table B2 Prediction regression of information flow to correlation coefficient (the rebalance of sample stocks)

信息流量	随机抽样			小市值	高市值
	样本				
	1	2	3	4	5
$FLOW(t-1)$	0.000 31 *** (3.010 00)	0.000 28 *** (3.260 00)	0.000 23 *** (3.12000)	0.000 20 (-1.650 00)	0.000 48 *** (3.580 00)
$FLOW(t-2)$	0.000 32 *** (4.490 00)	0.000 24 *** (4.010 00)	0.000 29 *** (4.270 00)	0.000 03 (0.680 00)	0.000 36 *** (4.050 00)
$FLOW(t-3)$	0.000 12* (1.790 00)	0.000 12 ** (2.050 00)	0.000 13 ** (2.160 00)	0.000 03 (0.460 00)	0.000 23 *** (25.190 00)
DHY_t	0.069 10 *** (13.680 00)	0.029 60 *** (14.940 00)	0.025 00 *** (14.660 00)	-0.009 90 ** (-2.220 00)	0.058 11 *** (109.010 0)
DP_t	-0.016 75 *** (-7.840 00)	-0.031 30 *** (-21.410 00)	-0.030 47 *** (-22.640 00)	-0.076 10 *** (-16.760 00)	-0.024 34 *** (-78.340 00)
DMV_t	-0.099 21 *** (-116.700 00)	-0.083 95 *** (-99.200 00)	-0.084 39 *** (-113.150 00)	-0.115 27 *** (-18.290 00)	-0.062 57 *** (-246.180 00)
常数项	0.353 04 *** (140.860 00)	0.326 31 *** (198.290 00)	0.334 27 *** (225.150 00)	0.425 27 *** (109.530 00)	0.228 29 *** (509.720 0)
调整 R^2	0.261 20	0.245 00	0.240 90	0.061 90	0.173 60

注: 括弧内为系数的 t 统计量的值; *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著.

附录 C

采用如下模型剔除股票收益率序列中的市场整体因素

$$RET_{t, s_i} = \theta_0 + \theta_1 RET_{t, index} + \varepsilon_{t, s_i} \quad (C1)$$

式中 RET_{t, s_i} 为股票 s_i 的对数收益率; $RET_{t, index}$ 为上证指数的对数收益率; ε_{t, s_i} 为残差是股票 s_i 不含市场波动影响的收益率序列, 相关系数矩阵和动态相关系数等均基于残差序列计算.

表 C1 共同活跃用户数量矩阵与股价相关系数矩阵的 QAP 回归(2016)

Table C1 QAP regression of *STK* and *COR* (2016)

矩阵	回归系数		显著性水平	概率	
	非标准化	标准化		A	B
<i>STK</i>	0.000 1*	0.027 6*	0.065 0	0.065 0	0.936
<i>DHY</i>	0.068 7***	0.105 8***	0.000 0	0.000 0	1.000 0
<i>DMV</i>	-0.063 0***	-0.482 3***	0.000 0	1.000 0	0.000 0
<i>DP</i>	-0.027 7***	-0.118 0***	0.000 0	1.000 0	0.000 0
<i>DPE</i>	0.013 1**	0.042 5**	0.018 0	0.018 0	0.982 0
常数项	0.132 1	0.000 0	—	—	—
调整 R^2	0.285 0		0.000 0	—	—

注: *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。

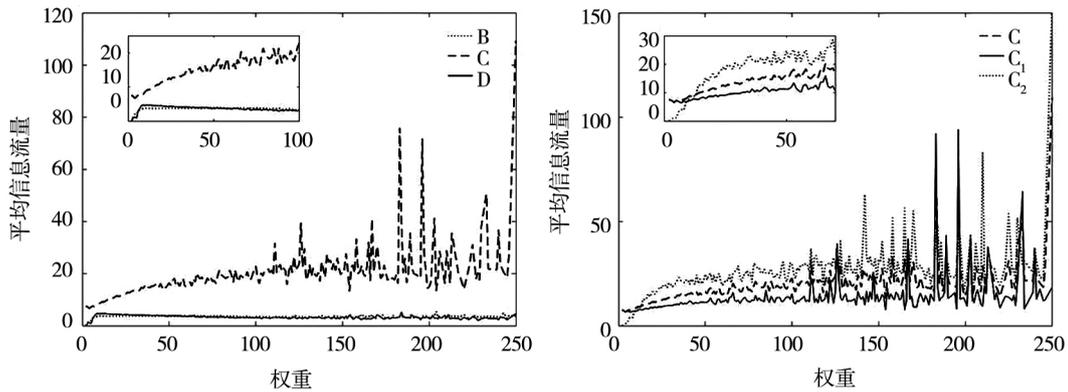


图 C1 不同权重的股吧对中各类共同活跃用户的平均信息流量(2016)

Fig. C1 Average information flows generated by various common users in pairs of stock bars with different weights (2016)

表 C2 信息流量对相关系数的预测回归(2016)

Table C2 Prediction regression of information flow to correlation coefficient (2016)

变量	模型			
	13	14	15	16
$FLOW(t-1)$	0.001 13*** (6.600 00)	0.000 60*** (5.750 00)	0.000 31*** (3.640 00)	0.000 19** (2.480 00)
$FLOW(t-2)$	—	0.000 86*** (7.830 00)	0.000 57*** (7.990 00)	0.000 31*** (4.100 00)
$FLOW(t-3)$	—	—	0.000 79*** (8.420 00)	0.000 63*** (8.220 00)
$FLOW(t-4)$	—	—	—	0.000 67*** (7.140 00)
DHY'	0.073 14*** (19.310 00)	0.072 82*** (19.340 00)	0.072 74*** (19.350 00)	0.072 51*** (19.410 00)
$DP'(t)$	-0.003 82*** (-2.830 00)	-0.003 84*** (-2.850 00)	-0.003 80*** (-2.820 00)	-0.003 84*** (-2.870 00)
$DMV'(t)$	-0.068 94*** (-115.420 00)	-0.068 99*** (-115.420 00)	-0.069 09*** (-115.370 00)	-0.068 84*** (-115.130 00)
常数项	0.135 17*** (83.780 00)	0.134 83*** (83.560 00)	0.134 66*** (83.400 00)	0.134 18*** (83.440 00)
调整 R^2	0.225 00	0.227 00	0.227 00	0.229 00

注: 括弧内为系数的 t 统计量的值; *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。

附录 D

表 D1 共同活跃用户信息流量矩阵与股价相关系数矩阵的 QAP 回归

Table D1 QAP regression of *CFW* and *COR*

矩阵	回归系数		显著性水平	概率	
	非标准化	标准化		A	B
<i>CFW</i>	0.000 0**	0.028 7**	0.033 0	0.033 0	0.967 0
<i>DHY</i>	0.086 5***	0.101 1***	0.000 0	0.000 0	1.000 0
<i>DMV</i>	-0.100 6***	-0.557 9***	0.000 0	1.000 0	0.000 0
<i>DP</i>	-0.025 6***	-0.086 1***	0.000 0	1.000 0	0.000 0
<i>DPE</i>	0.000 4	0.001 1	0.469 0	0.469 0	0.531 0
常数项	0.207 0	0.000 0		—	
调整 R^2	0.285 0		0.000 0	—	

附录 E

$$STI_{t, g_k} = \frac{POS_{t, g_k} - NEG_{t, g_k}}{POS_{t, g_k} + NEG_{t, g_k}} \quad (E1)$$

$$STID'_{t, g_k, g_l} = | STI_{t, g_k} - STI_{t, g_l} | \quad (E2)$$

式中 STI_{t, g_k} 为看涨指标, 代表共同活跃用户的相对看涨程度^[47], 介于 -1 和 +1 之间; POS_{t, g_k} 代表第 t 天股吧 g_k 中共同活跃用户的积极情绪, 为共同活跃用户产生的信息流中的正向词汇数量; NEG_{t, g_k} 代表第 t 天股吧 g_k 中共同活跃用户的消极情绪, 为共同活跃用户产生的信息流中的负向词汇数量; $STID'_{t, g_k, g_l}$ 为情感倾向差异, 其计算方法参考 Anton 和 Polk^[10] 对控制变量的处理, 该指标越大, 对应个股的相关系数可能越低。正、负词汇统计基于 Python 的 Jieba 中文分词模块, 本文在原有词典的基础上加入招金词酷, 该词库由招商证券金融工程研究团队构建, 其中包括社交网络、交易所、调研平台、财经媒体、人民日报和微信公众号等常用词汇, 有助于提高分词的准确性。

DLOC 为股票地域关系矩阵, 表示股票 s_i 和 s_j 的注册地 若为同一省份其元素取值为 1, 否则(非同一省份) 其元素取值为 0; *DLIS* 为股票指数成分股关系矩阵, 表示股票 s_i 和 s_j 的成分股关系 若均为上证 180 成分股其元素取值为 1, 否则(非同上证 180 成分股) 其元素取值为 0; *DLEV* 为财务杠杆比率差异矩阵, 财务杠杆比率为当年长期负债占总资产的比率, *DLEV* 元素取值为股票 s_i 和 s_j 财务杠杆比率之差的绝对值^[10]; *DVOL* 为超额交易量相关系数矩阵, 超额交易量计算参考 Landsman 和 Maydew^[46], 股票超额交易量相关系数矩阵采用 2017 年日度序列的皮尔逊相关系数, 股票超额交易量序列间的动态相关系数采用 DCC-GARCH 方法计算。

表 E1 共同活跃用户数量矩阵与股价相关系数矩阵的 QAP 回归(增加控制变量)

Table E1 QAP regression of *STK* and *COR* (with more controls)

矩阵	回归系数		显著性水平	概率	
	非标准化	标准化		A	B
<i>STK</i>	0.000 1*	0.032 2*	0.084 0	0.084 0	0.916 0
<i>DHY</i>	0.043 3***	0.050 0***	0.000 0	1.000 0	0.000 0
<i>DMV</i>	-0.056 5***	-0.309 6***	0.000 0	1.000 0	0.000 0
<i>DP</i>	-0.025 7***	-0.085 3***	0.001 0	0.999 0	0.001 0
<i>DPE</i>	0.007 9	0.020 9	0.197 0	0.197 0	0.803 0
<i>DVOL</i>	0.290 0***	0.280 0***	0.000 0	0.000 0	1.000 0
<i>DLEV</i>	-0.072 1**	-0.046 7***	0.060 0	0.940 0	0.060 0
<i>DLIS</i>	0.120 7***	0.308 3***	0.000 0	0.000 0	1.000 0
<i>DLOC</i>	0.003 3	0.004 9	0.316 0	0.316 0	0.684 0
常数项	0.374 0	0.000 0		—	
调整 R^2	0.452 0		0.000 0	—	

注: *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。

表 E2 信息流量对相关系数的预测回归(增加控制变量)

Table E2 Prediction regression of information flow to correlation coefficient (with more controls)

变量	模型			
	17	18	19	20
$COR(t)$	0.950 33 ^{***} (1 014.330 00)	0.946 61 ^{***} (955.090 00)	0.949 06 ^{***} (1 020.080 00)	0.950 11 ^{***} (989.690 00)
$FLOW(t-1)$	-0.000 01 (-0.070 00)	-0.000 04 ^{***} (-2.840 00)	-0.000 04 ^{***} (-3.740 00)	-0.000 06 ^{***} (-3.300 00)
$FLOW(t-2)$	—	0.000 07 ^{***} (6.450 00)	0.000 05 ^{***} (4.290 00)	0.000 07 ^{***} (4.380 00)
$FLOW(t-3)$	—	—	0.000 01 (1.420 00)	0.000 02 (0.910 00)
$FLOW(t-4)$	—	—	—	-0.000 04 ^{**} (-2.00 000)
$STID'(t)$	-0.000 83 ^{***} (-11.380 00)	-0.000 32 ^{***} (-3.960 00)	-0.000 79 ^{***} (-8.000 00)	-0.000 98 ^{***} (-6.920 00)
DHY'	0.003 03 ^{***} (12.170 00)	0.003 44 ^{***} (12.840 00)	0.003 25 ^{***} (11.910 00)	0.003 03 ^{***} (9.790 00)
$DP'(t)$	-0.000 79 ^{***} (-7.310 00)	-0.000 79 ^{***} (-6.930 00)	-0.000 83 ^{***} (-7.300 00)	-0.000 97 ^{***} (-6.830 00)
$DMV'(t)$	-0.003 11 ^{***} (-39.490 00)	-0.003 07 ^{***} (-37.300 00)	-0.002 88 ^{***} (-35.050 00)	-0.002 90 ^{***} (-30.000 00)
$VOL(t)$	0.004 77 ^{***} (45.630 00)	0.005 25 ^{***} (44.770 00)	0.004 82 ^{***} (40.510 00)	0.004 55 ^{***} (29.060 00)
LEV	-0.004 06 ^{***} (-10.380 00)	-0.005 15 ^{***} (-12.310 00)	-0.004 00 ^{***} (-9.430 00)	-0.003 44 ^{***} (-6.580 00)
LIS	0.005 73 ^{***} (33.710 00)	0.005 56 ^{***} (31.660 00)	0.004 59 ^{***} (27.29000)	0.004 90 ^{***} (25.250 00)
LOC	0.000 26 (1.450 00)	0.000 38 ^{**} (2.050 00)	0.000 36 [*] (1.880 00)	-0.000 04 (-0.170 00)
常数项	0.011 99 ^{***} (41.690 00)	0.011 73 ^{***} (39.090 00)	0.012 03 ^{***} (39.780 00)	0.012 27 ^{***} (35.480 00)
调整 R^2	0.935 00	0.942 00	0.943 00	0.938 00

注: 括弧内为系数的 t 统计量的值; *、** 和 *** 分别表示系数在 10%、5% 和 1% 水平下显著。