

**本科毕业设计[论文]**

**投资者情绪、文本质量和股价同步性**

**——基于互动易的研究**

院 系 管理学院财务金融系

专业班级 财务管理1801

姓 名 何卓扬

学 号 U201815934

指导教师 李安泰

2022年 6 月 1 日

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

作者签名： 2022年 4 月 4 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密 囗 ，在 年解密后适用本授权书。

2、不保密 囗 。

（请在以上相应方框内打“√”）

作者签名： 年 月 日

导师签名： 年 月 日

# 摘要

股价同步性是衡量投资市场定价效率与资金配置效率的重要指标，党的十九大报告中指出，“深化经济体制改革，推动金融支持实体经济”是未来金融市场的变革方向。探索如何降低股价同步性，能够响应十九大报告的号召，对金融市场改革具有重要意义。

本文搜集了国内外互动性信息与股价同步性的相关文献。互动性信息平台即上交所深交所互动平台等允许投资者与上市公司进行交流的媒体平台。先前学者认为，股价同步性代表私有信息的披露程度，投资者提问数量越多，其提供的私有信息越多，能够显著促进市场信息流动，降低股价同步性。

本文以2013-2019年中国上交所投资者互动平台上互动性信息作为研究对象，搜集了将近1200家公司共约50万条信息，研究互动性信息中投资者情绪、信息质量与股价同步性的关系。

传统研究认为投资者提问的数量与股价同步性呈现负相关，但是本文通过回归证明了这两个变量之间的相关关系并不明显。研究发现，投资者提问中含有负面情绪的提问数量越多，所关注公司的股价同步性越低；由于文本质量影响，平均投资者提问长度与股价同步性相关性不显著。以上主要发现在变更主要变量模型、更改控制变量后依然稳健。这意味着投资者能够通过传递负面情绪来传播信息，但是大部分人提问质量堪忧。本文启示监管机构应该着重激励投资者以发挥其信息发现能力，同时也要加强投资者教育，提升其信息交流水平。

**关键词：**“互动易”平台，负面语气，信息含量

# ABSTRACT

The synchronization of stock prices is an important indicator to measure the pricing efficiency and capital allocation efficiency of the investment market. The report of the 19th National Congress of the Communist Party of China pointed out that "deepening the reform of the economic system and strengthening the ability of finance to support the industry" is the direction of future financial market reform. Exploring how to reduce the synchronization of stock prices can respond to the call of the report of the 19th National Congress of the Communist Party of China, which is of great significance to the reform of the financial market.

This paper collects domestic and foreign literature on interactive information and stock price synchronization. Interactive information platforms are media platforms that allow investors to communicate with listed companies, such as the Shanghai Stock Exchange and Shenzhen Stock Exchange Interactive Platform. Previous scholars believed that stock price synchronization represents the degree of disclosure of private information. The more questions investors ask, the more private information they provide, which can significantly promote information communication and reduce on-hook synchronization.

This paper takes the interactive information on the investor interaction platform of the Shanghai Stock Exchange from 2013 to 2019 as the research object, and collects about 500,000 pieces of information from nearly 1,200 companies to study the relationship between investor sentiment, information quality and stock price synchronization in interactive information.

Traditional research believes that the number of investor questions is negatively correlated with stock price synchronicity, but this paper proves that the correlation between these two variables is not obvious through regression. The study found that the more questions investors asked with negative emotions, the lower the stock price synchronization of the companies concerned; due to the influence of text quality, the average investor question length was not significantly correlated with stock price synchronization. The above main findings are still robust after changing the main variable model and changing the control variable. This means that investors are able to spread the word by conveying negative sentiment, but the quality of the questions asked by most of them is of poor quality. This paper suggests that regulators should repeatedly incentivize retail investors to exert their information discovery capabilities, and at the same time, strengthen investor education and improve their level of information exchange.

**Key words** : "Easy Interaction " platform, negative tone, information content

**目录**

**[摘要](#_Toc11030)** [iii](#_Toc11030)

**[ABSTRACT](#_Toc24324)** [ii](#_Toc24324)

**[1. 绪论](#_Toc18903)** [6](#_Toc18903)

[1.1 研究背景和研究问题 6](#_Toc31996)

[1.2 研究目的和研究问题 8](#_Toc12896)

[1.3 研究内容和研究贡献 9](#_Toc26792)

**[2. 文献综述与理论分析](#_Toc14757)** [12](#_Toc14757)

[2.1 股价同步性 12](#_Toc17828)

[2.2 股价同步性与部分信息披露渠道 13](#_Toc21881)

[2.3 股价同步性与互动性平台 14](#_Toc14448)

[2.4 股价同步性与提问数量或字符长度 15](#_Toc3669)

[2.5 股价同步性与投资者情绪 15](#_Toc11708)

[2.6 本章小结 17](#_Toc21493)

**[3. 基于爬虫技术的投资者负面情绪指标生成](#_Toc23794)** [19](#_Toc23794)

[3.1 数据库基本介绍 19](#_Toc29748)

[3.2 爬虫介绍与选择 19](#_Toc26674)

[3.3 具体数据搜集整理方法 20](#_Toc11675)

[3.4 情感分析方法 21](#_Toc19746)

[3.5 本章小结 22](#_Toc25461)

**[4. 假设提出与研究设计](#_Toc13626)** [23](#_Toc13626)

[4.1 假设提出 23](#_Toc26466)

[4.2 变量定义与数据描述 24](#_Toc26998)

[4.3 被解释变量 25](#_Toc11464)

[4.4 实证模型 27](#_Toc3219)

[4.5 本章小结 28](#_Toc7321)

**[5. 主要结果](#_Toc6215)** [29](#_Toc6215)

[5.1 描述性统计 29](#_Toc13359)

[5.2 对传统结论的检验 33](#_Toc15628)

[5.3 投资者提问的负面语气与股价同步性 35](#_Toc12389)

[5.4 投资者提问平均长度与股价同步性 37](#_Toc379)

[5.5 稳健性检验 38](#_Toc19981)

[5.6 本章小结 41](#_Toc9544)

**[6. 结束语](#_Toc13168)** [42](#_Toc13168)

**[致谢](#_Toc11617)** [44](#_Toc11617)

**[参考文献](#_Toc29312)** [45](#_Toc29312)

**[成 绩 评 定](#_Toc27070)** [55](#_Toc27070)

# 绪论

## 研究背景和研究问题

1. **研究问题**

随着我国经济的高速增长与国际化程度的进一步提升，信息披露越来越重要。资产的价格同时受到信息与噪音的影响，有效的信息披露不仅能够使得资产价格配置更加有效，同时也能降低噪音对资产价格的扭曲，信息效率是证券市场运行的基础（谭松涛等，2016），如何提升信息效率也成为了资本市场进一步深化改革的重要难题。

股价同步性指的是公司个体股价波动受到市场总体股价波动影响的程度，是学者们用来衡量资本配置效率的重要指标。股价同步性越低意味着公司股价受到市场指数影响越少，基于有效市场假说我们可以认为公司股价中包含有更多的私有信息而非市场总体信息，这也就意味着市场中信息流通性水平较高。学者们普遍认为，我国股价同步性较高，股票价格中包含公司私有信息较少，这降低了金融体系中资本配置的效率。（Morck et al.,2000；Gul et al.,2010;黄俊和郭照蕊,2014）。Durnev等（2004）研究发现市场资金分配效率与股价同步性负相关，个体股价的异常波动是私有信息的体现，而更加丰富的异质信息能够提升股票市场的投资效率。此外，从信息透明度来看，股价异质波动率负向相关于资本市场信息定价效率（王亚平等，2009）。党的十九大报告中指出，“促进经济体系变革，推动金融支持实业”是未来金融市场的变革方向。只有在资本市场充分反映私有信息的条件下，我国金融市场才能发挥配置经济资源的作用，进一步助推实体经济发展。因此，资本市场信息效率是我国金融研究的重点，关注股价同步性的相关影响因素并且探究如何使其下降到合理水平是我们未来的工作目标，它对提升资本市场定价效率，促进我国进一步深化金融改革有重大意义。

目前已经有大量研究开始关注社会责任报告、问询函等传统非财务信息披露渠道。王艳艳等（2014）认为第一次非强制公开社会责任报告能够改善资本市场信息环境，李新丽等（2018）发现强制性信息披露能够提升信息分配效率，官方日常的公告语气对企业未来收益状况也有较强的预测力（Loughran and McDonald,2011；Li，2008）。Bushee et al.（2017）认为交流会议能够使投资者与管理层进行非公开会谈，在这些时间段中交易规模会显著增大，但是相较于投资者会议，分析师引发的市场反应更强（Kirk and Markov,2016）。

然而上述研究首先局限于政府等传统监管机构，忽视了新兴互联网信息披露渠道。其次，这些研究受限于研究角度，只能观察到信息披露带来的影响，无法研究投资者获取信息的具体过程，也无法深入探讨投资者与其他主体信息交互的过程。随着互联网的普及，基于互联网的沟通方式引起学者们的广泛关注，通过对社交平台的研究发现，上市公司开通微博能够降低其股价同步性，并且发布微博数量和股价同步性呈现负相关（胡军和王甄，2015；周冬华，2016）。就微信而言，刘姗（2017）发现，上市公司官方号的综合影响力越大，企业的股价同步性水平越小（张洁，2018）。然而传统研究主要关注点要么是上市公司单方面的信息披露，要么就是投资者之间的交流，缺乏对上市公司与投资者互动机制的研究。

传统研究认为，互动频率越高，字数越多，信息效率越高，越能够缓解资本市场的信息不对称（杨凡和张玉明，2020；丁慧，2018；蒋艳辉和郑佳尔，2017），与传统结果不同，我们的研究发现，信息发布的数量和股价同步性之间并不具有显著的负相关，并且信息字数越多并不意味着股价同步性会显著降低。由于信息收集手段的缺陷与投资者视角的缺失，传统研究更多集中于公司视角下的信息效率研究，缺乏对公司与个体之间的互动，更重要的是缺乏群众情绪传染的识别与探究。

2013年上交所推出的交流平台是投资者与企业交流信息的重要渠道，也为学者们展现了投资者与上市企业在网络空间交流的研究场景。并且，由于信息获取技术如爬虫的发展，研究者们能够大量搜集相关数据。这些变化让学者有机会去探究个人与企业以及个人与个人之间的信息传播对资本市场的影响。谭松涛等（2016）将2010年深交所互动易平台的推出作为外生冲击，研究发现平台为投资者提供了新的获取信息渠道并且改善了信息效率。岑维等（2014）发现网络平台上投资者注意力的增加能够显著消除投资市场信息不均衡，并且上市公司回复数量和个股收益情况或换手量存在显著正相关关系（蒋艳辉，郑佳尔;2017）。这些都证明了沟通平台能够促进投资者交流与资本市场资产优化配置。

近年来，相当数量的学者开始对互动易平台上投资者情绪传染进行分析，孟庆斌等（2020）发现“互动易”上投资者提问与上市公司相关负责人答疑的负面语气能够有效传递公司信息并且显著负向估计股票未来累计超额利益。而且互动平台上的投资者情绪跟踪指数对交易量呈现显著影响，这意味着互动平台上的情绪能够显著影响资本市场定价（陈晓娟，2020）。

1. **研究问题**

基于以上研究，本文探究下面两个问题。

1. **积极/消极提问对股价同步性的影响**

在互动易平台上投资者对公司的积极/消极提问会如何影响股价同步性？引起的反应是正向的还是负向的？其背后的机制是怎样的？提问字数和数量是否会影响股价同步性？

1. **信息质量与股价同步性**

我们知道提问中可能存在大量质量低下的提问，我们能否筛选出那些质量低下的提问并且检验出信息质量与股价同步性之间的关联？

## 研究目的和研究问题

1. **研究目的**

基于我们的研究背景，本文研究目的如下：

1. **研究股价同步性与信息数量/平均长度的关系**

本文首先通过实证分析来探究传统结论：信息数量/平均长度越高，股价同步性越低这一结论。

1. **研究股价同步性与投资者情绪的关系**

如果传统结论不成立，我们进一步探究股价同步性与投资者情绪的关系，希望通过本文的补充，对投资者情绪-股价同步性这一相关关系进行有益的补充。

1. **研究股价同步性与提问平均长度相关关系不显著的原因**

探究投资者提问与文本质量的关系，探究股价同步性与文本质量的相关关系，给出进一步衡量文本质量的方法。

1. **给出政策启示**

基于当前我国上市公司信息披露普遍较不明朗的情况，以及在互联网迅速增长情况下，投资者与董秘沟通越来越重要的基本条件，我们希望本文的研究能够给监管机构一定的政策辅助，同样也给投资者以参考，提升双方沟通质量。

1. **研究方法**

本文主要使用以下两种方法。

1. **文献分析法**

我们首先对股价同步性，新媒体平台等等相关的文献进行分析与总结，以了解和探讨目前关于股价同步性研究的成果与缺陷；其次，根据所梳理的文献，我们来分析相关的理论基础，并且根据业界最新的研究提出我们的具体的研究问题。

1. **实证分析法**

利用固定效应面板模型，控制了年度和行业效应等等，探讨消极提问数量和股价同步性之间的关系，还探讨了虚词数量对文本质量的影响。最后我们检验了结果的稳健性

## 研究内容和研究贡献

1. **研究内容**

互动式信息沟通是否能够缓解股票市场的信息不对称，从而有效降低股价同步性？为了回答上述问题，本文以2013年1月1日至2019年12月31日上交所上市公司为研究对象，搜集了1200余家公司共约50万条交流数据，包括投资者提问内容，提问时间，提问长度，提问积极消极程度，董秘答复内容，答复时间，答复长度，答复积极消极程度等指标。

本文基于上交所 “互动易”平台上投资者提问和董秘回复进行情感文本分析，考察二者的负面语气是否具有信息含量。并且对文本质量进行度量，检验文本质量能否提升信息质量。研究发现与传统研究不同，投资者提问长度与股价同步性相关性不明显，这可能是因为没有考虑到情绪与文本质量的原因。本文发现消极提问能够显著降低股价同步性，并且投资者提问长度和文本质量无关，无法显著影响股价同步性。

1. **研究贡献**

此次研究的边际贡献主要体现以下三点：其一，本文的研究为现有投资者与上市公司互动方面的研究补充了个人投资者层面的内容。过往研究主要集中于传统信息渠道（何贤杰等，2016），本文则将研究拓展到“互动易”平台这一新兴互联网沟通渠道。并且重点关注投资者行为而不是公司行为。其二， 本文一定程度上弥补了现有投资者与公司互动行为的研究缺失，由于数据量较大，前人主要研究上市公司是否与投资者沟通或者沟通的数量（杨凡和张玉明，2020；丁慧，2018；蒋艳辉和郑佳尔，2017），而本文采用了程序大数据分析的方法，能够更加关注信息质量与情绪质量。其三，本文的研究在一定程度上缓解了行为金融与市场有效性之间的冲突。在行为金融的研究中，投资者情绪往往被认为是非理性的，有时被视作噪音（West, 1998）。但是本文研究发现， 投资者的负面情绪能够吸引投资者关注，从而传递出特质信息，有效提升市场有效性。其四，本文证明了信息长度不一定代表信息质量，越长的信息中中含有越多的虚词，根据信息长度分组能够发现信息长度与信息质量相关性较弱。

本文剩余部分的结构如下，第二部分进行了文献综述和理论分析，并提出研究假设；第三部分介绍主要的变量、数据描述以及研究方法；第四部分对研究假设进行检验并分析其结果；最后一部分为研究结论并提出相应政策启示，下面图1-1是我们的基本论文结构。

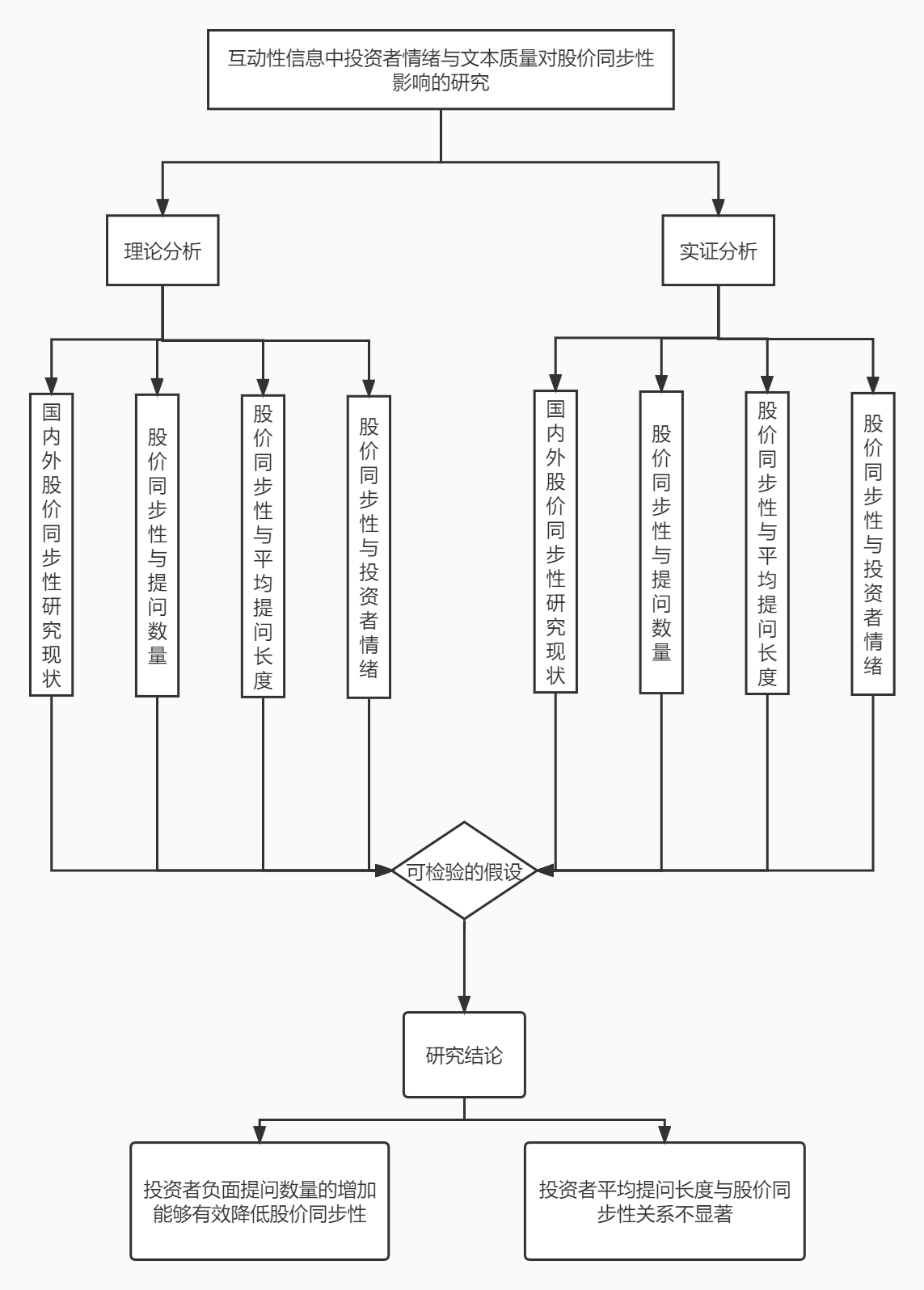


图1-1 论文结构展示

# 文献综述与理论分析

股价同步性指的是公司股价波动与总体股价波动的相关性水平，有时也被称为股票价格的同涨同跌性（Piotroski and Roulstone，2004；许年行等，2011）。学者研究发现，个股收益波动中只有很少一部分能够被市场或者行业因素来解释，个股的特质性波动可能是公司特质性信息或者是其他突发性事件导致的。过往研究股价同步性主要关注于研究报告等对公司私有信息的披露（Liu,2011），近年来也有关注于网络社交平台与网络互动平台的。

近年来，随着大数据技术的广泛应用，越来越多的学者开始搜集和关注投资者情绪这一指标，有研究表明媒体负面语气的减少能够有效提升承销商费用，提升超募资金比例，这也就意味着机构有动机通过信息渠道推荐IPO标的，提升投资者对该股票好感情绪（汪昌云等,2015），不过大部分研究主要专注于通过各类媒体平台来传播投资者情绪，没有研究在直接互动平台上的情绪传染问题。

新的技术也拓展了测量私有信息的方法，过往大部分研究主要以投资者-公司交流数量或者字数作为代理变量，但是近年来越来越多的研究开始着手于使用计算机软件来对文本进行更加丰富的处理和质量评测（周耿等，2020）。不过值得注意的是大部分研究还是缺乏对文本质量有更加深入和直观的认识，在进一步应用计算机领域如NLP等相关技术后可能能够解决这些问题。

## 股价同步性

West（1988）认为理论上的波动程度远低于股票价格的真实波动，增量波动很有可能是非理性噪音交易导致的。这一结论为股价同步性差异基于投资者非理性的假设巩固了基础。

Eun（2015）等指出，从1993年到2010年，全世界47个市场里面，我国的股价同步性大概是0.549，排名相当靠前。其他研究者也有类似的结论，根据Morck等（2000）的研究，1995年全球资本市场中我国股价同步性约为0.45，为全球第二，而同时其他发达国家的同步性均在0.10以内。这些数据表明在我国研究股价同步性有着重大意义，可以解决我国资本市场的重要问题。

## 股价同步性与部分信息披露渠道

我们可以从几个侧面来观察信息披露对股价同步性的影响。

**1）研究报告**

研究报告在资本市场中充当投资者和标的物之间的沟通桥梁，分析师具有优秀的分析技术与信息渠道，挖掘和利用信息的能力高于普通投资者。理论上来说，通过研究报告，分析师可以披露更多公司的内部消息，揭示不同特质性信息对股价的影响情况，从而增加信息沟通交流水平，降低股价同步性。Liu（2011）认为，研究报告主要提供的是企业方面消息。他认为，之所以分析师会针对公司制作研究报告，很大原因在于其从某些渠道搜集到了公司的特有信息，而且该企业的股价有较充足的市场信息。他的研究发现，研究报告主要影响的是公司特质回报率，而对市场、行业的长期、短期回报率影响不大。Crawford等（2012）的研究则指出了，第一个出现的研究报告主要分析了宏观的行业信息，而之后出现的其他研究报告迫于压力，会更多分析微观的公司特有信息。

**2）社交媒体**

社交媒体主要包含有投资者与投资者，投资者与公司的交流和沟通。伴随着互联网的高速发展，越来越多的学者开始重视社交媒体在信息传播中的作用。

Blankespoor等（2014）以推特上的链接数量作为主要研究对象，发现新媒体平台上，随着新闻的转发数量增加，股票的买卖价差也显著降低。Kim等（2017）主要关注网络社媒平台对证券评级师股票评级决策的影响，研究表明来自投资者的负面信息影响证券研究员的条件相当苛刻，需要有巨大的传播量并得到重要媒体机构的支持，这表明上市公司官方平台所发布的特质信息能够有效缓解信息不对称。

国内研究中，企业在微博平台开通官方号发布私有消息能够有效缓解信息不对称，降低股价同步性（何贤杰，2016）。胡军和王甄（2015） 研究发现上市企业建设微博号能够提升证券研究员预测股价精度，降低股价同步性；同时，社交媒体平台信息主要通过研究员的分析与解读影响股价。微信作为新媒体平台和微博有一定的类似性，例如张洁（2018）研究发现，官方号总体披露消息条数增加会降低股价同步性。刘姗（2017）发现，上市企业官方号的综合影响力增加，公司的股价同步性降低。

## 股价同步性与互动性平台

投资者能够获取及时且准确的信息是市场信息自由流通，资源合理配置的前提。研究表明，投资者与个体股民通过与上市公司董秘等公司高级管理人员直接进行沟通来获取企业盈亏信息、了解企业未来发展战略与相关行业政策( Kimbrough，2005) 。我国共有三大交易所，深交所、上交所与北交所。其中深交所与上交所分别建立的网络互动平台（互动易平台）是我国在公司—股民信息交流，上市公司信息公开方面的重要体制建设。

以新时代的网络平台为基本的沟通渠道，基于官方强力监管的前提下，投资者与上市企业董秘以提问-回答形式进行信息沟通交流，现阶段有关此类互动平台的研究切入点主要有以下几点：

以在交流平台的提问-回复总体数量为投资者注意力的中介变量，关注其对于信息传递效率与资源配置和股价波动的影响。岑维等（2014）发现，更多的交流可以降低股票的总体波动和流动性问题。基于信息交流的方向，投资者注意力能够降低信息不对称情况，从而进一步保护小微股东的所有权益并进一步降低公司直接融资的费用（岑维等，2016）。

从信息交流间隔时长以及是否有回复具体问题的方向探索沟通行为特点及影响。以总体提问-沟通交流字符数量作为关注度的代理变量，交流数量的提升能够促进投资者获取特质性信息并且增加其解读特征信息的能力，这有助于平衡其盈余预期波动，并显著促进盈余公告期的信息沟通与交流（丁慧等，2018）；周耿等（2020）在《上证“e互动”对市场信息效率影响及机制研究》中将文本向量化，利用ICTCLAS3.0系统对回复文本进行中文分词处理以测度文本的信息含量，证实了上市公司与投资者的高互动性水平能够促进市场提升信息质量。

## 股价同步性与提问数量或字符长度

股价同步性代表着股价的同涨同跌性，也代表着公司的私有信息是否能够有效融入到股价中去。从这个角度来说，很多学者认为，信息披露的数量或者是字数越多，则意味着其能够包含更多的私有信息，从而降低股价同步性。

从社交媒体的角度来说，张洁（2018） 通过对2015年-2017年A股上市公司开通微信公众号的情况进行了研究，公众号信息披露数量与股价同步性更低，但是公众号信息必须要和公司高度联系，不能是其他非相关内容。

从网络互动平台来看，有不少学者以提问与回复的数量来作为投资者关注的中介变量。岑维等（2016）将投资者互动数量作为投资者关注度的主要解释变量，研究发现，投资者关注度和信息披露质量呈现相互替代的现象，并且高关注度能够显著减少公司和投资者之间的信息沟通问题。丁慧等（2018）以总字数与平均字数作为解释变量，发现基于互动性平台的信息沟通和交流能够有效降低盈余公告期的信息不对称，但是需要监管机构着手于提升投资者的信息解读能力。

## 股价同步性与投资者情绪

研究认为负面情绪有利于信息的传播与沟通，这是因为负面消息相较于正面消息更加受人关注，因而更加容易传播，从而降低股价同步性。

媒体的负面消息能够更好的煽动投资者情绪，作用于资产价格。负面情绪下降，IPO抑价率、超募资金比率和承销商费用都会上升，因此发行公司和承销商都有动力去推动媒体减少负面消息（汪昌云和武佳薇，2015）。倪景铄（2017）通过研究2006年至2014年中国A股上市公司（去除金融类公司）的季度数据，发现投资者的非理性行为会造成整体的羊群效应，也会放大非理性情绪对公司个股与市场的同步性。周世庆（2021）则针对股吧的投资者情绪进行了研究，发现投资者情绪越高涨，股价同步性越高，消极情绪与股价同步性的关系并不显著。方健和任春苗（2021）的研究指出，对于媒体报道数量较多的公司，投资者在互动性平台上表现的负面情绪能够有效降低股价同步性，并且公司回复发挥了中介效应。

表2-1 文献总结

|  |  |  |
| --- | --- | --- |
| 领域 | 研究方向 | 主要内容 |
| 非互动性平台 | 信息披露数量/长度 | 研究报告数量的增加能够提供公司的特有信息，从而降低股价同步性（Liu,2011）；推特链接的转发能够促进信息的传播，从而降低信息不均衡（Blankespoor，2014）。 |
| 投资者情绪 | 股吧中投资者情绪的高涨会造成羊群效应，提升股价同步性，然而消极情绪对市场影响并不显著（周世庆，2021）； |
| 互动性平台 | 信息披露数量与长度 | 互动性平台上的提问-回复数量的增加能够提升投资者关注度，减少公司和投资者之间的信息沟通问题。（岑维，2016） |
| 投资者情绪 | 投资者的非理性情绪会放大公司股价同步性（倪景铄，2017） |

## 本章小结

目前社会责任信息披露与股价同步性的研究目前存在以下不足：（1）结论不一致。可以发现目前所有文献得到的结论之间并不相同，这也暗示着在研究方法等方面仍有改进的空间。（2）对互动式信息信息内容利用不足。大部分期刊文献均未考虑互动式信息内容的影响，部分硕士论文虽然考虑了情绪的影响，却也缺乏对文本质量的衡量。

新时代的社交媒体在投资者与公司交流期间起到了重要的沟通平台作用，传统投资者获取信息的方式主要是通过媒体报纸或者是公告等传统手段，这些传统手段大部分都具有一定的滞后性，并且还容易被相关利益方进行信息管制与操纵，也没有相关投资者交流平台。网络信息互动平台大大改善了传统的B2C交流模型，使其转变为B2C2C模型，并且大大扩展了相关信息节点的沟通方式。大量的企业在社媒传播了巨量的数据，为如今研究社交媒体和股价同步性关系提供了丰富的研究数据。通过文献梳理，微信公众号的研究尚处于起步阶段，大部分研究基于微博等传统研究思路，没有挖掘出属于微信公众号的独特属性，例如朋友圈本质上属于私域流量，微博信息主要属于公域流量，在不同的平台信息传播裂变形式不同，现有研究大多忽略了朋友圈传播的影响。并且，其他研究对微信公众号传导路径持有不同观点，张洁（2018）认为机构投资者是中间渠道，但是缺乏合理的逻辑架构。另外，现有论文大都忽略了股评公众号或者是微博号作用，其可能的困难在于信息噪音的甄别与排除。

对比于国内其他社交媒体如“新浪微博”等，类似于深交所的“互动易”，上交所“ｅ互动”为市场信息效率影响的机制研究提供了一个更有利的互动环境，为个体与董秘交流提供了沟通的桥梁，因而有可能会提升整个市场信息效率水平。很多研究者在有关网络互动平台信息披露的研究开展了相当程度的探索，然而还是存在一些问题：（1）对于互动媒体上信息公开的研究集中于研究其后续相关结果，而缺乏对互动式信息交流机制、行为特征及传导路径的研究。已有研究大都将基于互动的信息交流作为投资者的分析能力、注意力以及相关感性因素的的代理变量，研究其对于资本体系信息不均衡与公司行为决策的影响。（2）对于信息交流的研究主要是关注信息数量等特点，没有考虑到文本质量这一重要因素。对比条数等简单信息，投资者情绪等新颖指标无疑具有更加丰富的研究意义。现关注于信息交流的学者主要都使用简单条数或者数量的作为交流的中介变量，集中于信息供给方视角，没有研究信息交流的特点与其后续影响，缺乏对文本内容的深入分析。

在此基础上，我们可以基于更加发达的信息交流公开平台进行进一步分析与探究。第一，以交流信息内容情绪传染为主要自变量，研究其和股票市场信息不均衡及企业决策的影响，得到更具现实意义的结论；第二，相较于研究互动式信息披露对市场效率的影响，可以对具体后续影响进行进一步分析与深入。

# 基于爬虫技术的投资者负面情绪指标生成

近年来，互联网的兴起以及研究数据量的高速增长，传统的手工收集已经无法满足数据搜集的需要。越来越多的学者开始利用爬虫进行数据的搜集和整理，然而应该选择何种爬虫来进行搜集数据，以及应该如何具体整理数据，还是一个值得探讨的话题。

另外，现有研究指出情绪传染与股价同步性的关系，但是如何选择合适的中介变量是一个难题。本章将会介绍数据库的基本情况，衡量情绪的几种方法以及具体的应用场景。

## 数据库基本介绍

一个典型的上证所网址如下：<https://sns.sseinfo.com/company.do?uid=559。其中后三位为股票代码，加上600>即为股票完整代码600559。我们可以用这个网址来直接访问对应的公司网页。

我们可以从网页中获取到如下几种数据： 提问时间，提问内容，回复时间，回复内容。更加具体的，经过数据处理我们可以获得以下几类数据：提问-回复时间差，月提问/回复数量，月提问/回复字数，平均每条提问/回复字数，提问/回复积极消极程度。

## 爬虫介绍与选择

对于股价同步性与媒体信息的研究需要搜集大量数据，过往研究一般会选择传统数据库，或者手动搜集相关媒体信息（何贤杰等,2016）。但是随着研究的进一步深入，传统的数据库或者是手工搜集无法满足我们对数据的需求，我们都需要在互联网之上收集大部分媒体数据。但是大部分数据并没有被网站进行整理和分析，我们需要利用爬虫技术对相关网页进行数据搜集，更多的研究者也开始使用爬虫进行资料的搜集（张洁,2018；丁慧，2018）。

然而如何进行具体爬虫种类的选择是很重要的。

爬虫的种类主要有以下几个：

**1）表面爬虫**

对基本网页进行搜索的爬虫就是表面爬虫。基本网站就是能够被自然搜索检索到的页面，还有就是可以通过链接直接访问的页面。

**2）深度爬虫**

对深层网页进行爬取的爬虫称为深度爬虫。深层网页是指必须要通过各种操作来访问你的例如，只有在用户注册后才能看到内容的网页属于深网。

这两种爬虫用到的技术可以分类为静态爬虫与动态爬虫，分别对应表面爬虫与深度爬虫。静态爬虫主要是指没有使用到任何点击操作，而是通过直接访问网页完成的爬取数据。动态爬虫则用处更为广泛，能够允许操作者操作网页，点击按钮等等操作。但是动态爬虫效率更低（因为要模仿人类点击），耗时更长，总体而言并不被人们所接纳。

笔者使用的技术就是动态爬虫技术，因为对应网站需要点击按钮或者下拉界面才能够获取到全部信息。

## 具体数据搜集整理方法

接下来，以上证互动易为例子来讲解一下具体的数据搜集方法。

上证互动易的典型网址是：<https://sns.sseinfo.com/company.do?uid=559>，最后三位数字是股票的公司代码。因此，我们只需要先搜集到所有公司的股票代码，然后进行网址的复制粘贴就可以了。

通过在网址上的搜索，我们可以在上证所网页上获取一个公司名称与股票代码对应的excel，我们通过python中pandas库读取相对代码，然后生成对应的爬取网页，最后得到具体的网页。

接着我们使用python的xpath语法来对网页进行分析，由于提问和回复的结构都类似，我们只需要分析出标签就可以进行数据的搜集和整理。

## 情感分析方法

传统的股价同步性与媒体研究只关注提问与回复数量与字数等信息（丁慧等，2018），但是近期越来越多的研究专注于情感分析，孟庆斌等（2020）发现负面语气对信息传递有很大的作用。

情感分析目前有两种主流方法：词典与机器学习等等。

情感词典即定义出正负情感语句的词语，通过不同词语来进行打分。例如“我喜欢我的午餐”，由于出现了“喜欢”这个词语，所以大部分情况下都会将这句子认定为积极地。

基于机器学习是挑选关键词，讲文本进行分析然后使用矩阵，具体常用方法有利用贝叶斯定理等等来进行分析。

具体来讲，在挑选情感分析方法的时候应该考虑以下几点：

1、情感分析任务时，需要明确项目需求是什么；应该以什么体量的文本作为分析基准，例如章节，段落，句子，短语，单词等；是分析整个文本还是部分文本。

2、在实际的落地应用中，往往需要根据领域特点制定一些情感分析标准，例如如果我们使用词典方法来进行积极消极语境分析，不同的领域有不同的积极/消极词语。例如在电商评价领域就是好评差评。

3、语料库数据处理。词典处理完成前面的工作后，下一步就是对语料库或词典摘要进行处理。在这一步中，不同的方法有不同的任务，基本上都是人的工作。分析标准的制定，因为看到实际数据后，你会发现标准中总会有一些歧义，例如在特定语境中某些词语是积极/消极的。

4、根据数据的特点和规模选择合适的方法，评价方法的优劣。项目中的方法不是单一的方法，不可能用一种方法或模型来解决各种数据源的问题。为了取得好的效果，我们必须采取分而治之的方式。本文主要分析上市公司的提问-回复，但是不同行业有不同的特点，比如在金融行业会有特殊的行业术语，股民们往往关心分红问题等等，具体词语在语境中的含义需要结合具体句子来分析。

传统研究主要采用词典法，但是这种方法主要是针对某个特定类型或者行业的公司研究。原因主要有两个，一个是不同行业的积极消极词汇并不一定相同，有些特定词汇代表着积极/消极意思，需要仔细甄别。一个是某个行业常用高频词数量并不太多，使用词典法能够穷尽大部分可能。

本文采用的是调用第三方情绪库，因为本文处理的数据量大，调用第三方情绪库速度快，问题是无法仔细辨别不同行业的积极/消极词汇。我们在实证阶段采用固定公司行业效应缓解了这个问题。

## 本章小结

本章主要研究数据的搜集与处理问题，我们从上证所搜集了相关的互动性信息，并且根据过往研究进行了对数化等数据处理，形成了研究的数据基础。

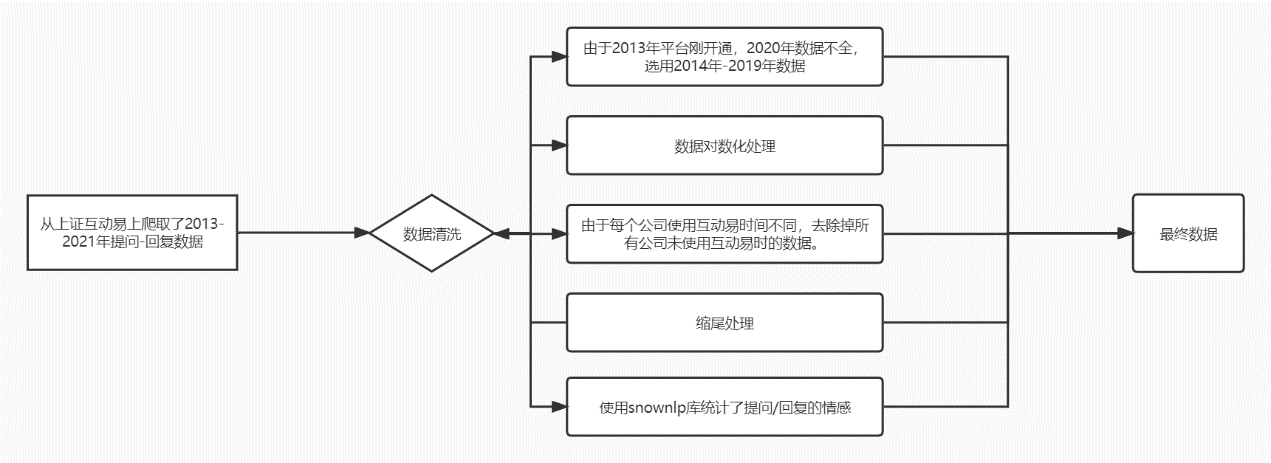


图3-1 爬取总结

# 假设提出与研究设计

本章主要介绍数据来源以及相关处理方法，我们使用爬虫技术搜集了2014年至2019年上证互动易平台上共1200余家企业的提问-回复数据。使用固定效应模型来确定年份和行业的影响，并且根据过往文献控制了提问数量、公司规模和公司成长性等七个变量。

## 假设提出

互动性平台上的信息沟通到底是通过怎样的途径影响股价同步性的？基于传统股价同步性理论，股价同步性即公司个体股价波动受到公司整体股价波动的影响程度，是学者们用来衡量资本配置效率的指标（Morck et al.,2000），同样也代表着市场信息定价效率（王亚平等，2009）。传统研究普遍认为私有信息数量和长度能够作为投资者关注度的中介变量（丁慧等，2018），这一视角主要着眼于投资者-公司这一传统联系，认为投资者针对公司的提问能够增加提问投资者对信息的理解，进一步通过市场机制将其融入到股价中。

诚然，从传统研究的角度，如研究报告等等，信息数量的增加能够直接代表私有信息增加（Liu，2011），但是互动性平台相较于传统渠道有两个新特征：其一是互动性平台不仅有投资者-公司这一联系，而是还有投资者-投资者这一互联网时代更加重要的联系，这意味着在资本市场中情绪传染也十分重要，投资者的负面情绪可能会传染给更多的投资者，从而形成羊群效应，影响到资本市场。其二是信息部分是由投资者来传播的，投资者提问的质量良莠不齐，传统的信息披露渠道多半是有职业专家来披露，信息质量较高，但是不同的投资者可能会带来不同的提问质量，对私有信息的传播力度也不同。

互动平台上内容主要包括个人发出疑问与上市公司答复两个部分，本文主要关注投资者提问。具体而言，本文将分别考察投资者提问的负面语气与文本长度如何影响股价同步性。本文考虑主要考虑投资者提问中的消极提问数量，主要是由于过去学者（Das and Chen, 2007; Tetlock et al.,2008; Davis et al.,2012）研究表明，消极性提问能够更有效地传染给市场上的投资者。

首先，本文对交流平台上个人提问的消极情绪能否传递私有信息，从而影响股价同步性进行理论分析。通过前人的研究可以看到，社交媒体的背景下，网络沟通方式确实能够有效提升信息效率（杨凡、张玉明，2020；周狄， 2020；），并且负面语气对信息传递也有着重要的影响（孟庆斌等，2020）。

因此，由于此类互动平台和社交平台不同，用户大多是专业投资者，对负面信息更加敏感，当投资者在对特定公司的质询包含了可能的消极信息时，就可以吸引其他投资者的兴趣，引导他们通过社交媒体平台对信息进行进一步的分析和传播，从而影响股价同步性。基于以上分析，本文提出假设H1：

**H1：在其他条件不变的情况下，互动平台上投资者提问的负面语气能够显著负向影响公司的股价同步性。**

传统研究认为提问和回复数量是相当重要的内容（张洁，2018；丁慧等，2018），会以其作为投资者关注的中介变量，但是也有研究证实了文本信息含量的重要性，周耿等（2020）在《上证“e互动”对市场信息效率影响及机制研究》中将文本向量化，利用ICTCLAS3.0系统对回复文本进行中文分词处理以测度文本的信息含量，证实了上市公司与投资者的互动性水平能够促进平台提升信息传播质量。

本文对互动平台上平均文本长度能否影响公司异质信息，从而影响股价同步性进行理论分析。传统理论认为文本长度越长可能就能够提供更多的特质信息，但是事实上如果文本虚词占比过高，反而会降低信息的有效性。另一方面，超长文本可能反应投资者内心强烈的情绪波动，反而会含有较少的信息质量。因此，本文认为互动平台上平均文本长度和信息质量有关。基于此，本文提出假设H2：

**H2，在其他条件不变的情况下，互动平台上投资者提问可以根据平均文本长度分为两组，并且两组之间文本质量差异显著。**

## 变量定义与数据描述

鉴于上交所“上证互动易”平台于在2013年启动，且当年数据样本较少，本文选取2014年1月1日至2019年12月31日作为样本区间进行研究，以上交所上市公司作为研究对象。虽然深交所也建立了功能相似的平台，但是数据收集过于复杂，因此本文仅仅针对上交所互动平台进行研究。

在数据处理中，本文剔除了ST上市公司和数据缺失的样本，并对变量做了缩尾与对数化处理。投资者提问的内容均来自于上交所“上证e互动”平台网站（http://sns.sseinfo.com/index.do），机构持股数据来自于Wind数据库，其他数据来自于CSMAR数据库。

## 被解释变量

此研究计划借鉴借鉴朱红军等（2007），以如下模型计算股价同步性，

 （4-1）

模型中和是数据样本期间第t个日期时的个股收益率与总体收益率,其中上交所与深交所的总体收益率分别用两地的综合指数收益率表示。由于我国上市企业的披露年报时间为次年的前四个月,为了使得综合收益率与年度报告总体时间段相对应,我们将每个年度的研究期间定义为该年度5月份第一个交易日至次年度4月份最后一个交易日。模型中的经济含义是个体股价价格的变动能够被总体价格变化所解释的部分。因此, 越大,表明公司股票价格包含了更少的特质性信息,其股价中信息流通更加不顺畅，股价同步性更高。由于 的取值区间为（0,1）, 不符合最小二乘法等的计量准则,为此我们参考MYY（2000）以及李增泉 ( 20 05 ) 的做法 , 对进行如下的对数转换：

 （4-2）

1. **解释变量：投资者提问的负面语气与平均文本长度**

以往研究希望通过建立个别的独立的词库（通常不超过100词）来辨别和分析文本中包含有的负面信息（Das and Chen, 2007; Tetlock et al.,2008;Li,2008;Davis et al.,2012），但由于中文相对更加复杂，仅仅使用词库现对而言准确性较低（汪昌云和武佳薇，2015）。为此，参考夏玉芹和单雪微（2018），本文以 Python常用语言分析库（Snownlp）为工具，对投资者提问逐条进行语义分析，并统计分值低于0.2的语句，从而得到本文的主要解释变量之一：n\_question\_2。

另外，本文还统计了每年的提问总字数，并为了控制文字长度的影响，本文用每年的提问总字数除以总条数得到每年的平均提问长度（len\_question），本文还统计了投资者提问到董秘回复的时间间隔（reply\_time）。

1. **控制变量**

参考过往研究（Cline等，2018），本文选择控制了公司规模。公司成长性，资产收益率，公司年龄等特征变量。另外，根据胡军和王甄（2015）的研究，本文引入了是否四大审计来控制了信息质量。

表4-1 变量定义

|  |  |
| --- | --- |
| 变量 | 定义 |
| **因变量：** |  |
| 股价同步性（SYN） | 公司股票波动与指数价格波动之间的相关性 |
| **自变量：** |  |
| 提问数量（question） | 当年该公司的投资者提问数量 |
| 平均每条提问字数（len\_question） | 当年该公司投资者提问每条平均字数 |
| 消极提问数量：数值低于0.2（n\_question\_2） | 当年该公司的投资者提问中消极提问数量（情绪系数少于0.2） |
| 积极提问数量：数值高于0.8（n\_question\_8） | 当年该公司的投资者提问中积极提问数量（情绪系数高于0.8） |
| 平均每条虚词字数（len\_fword） | 当年该公司的投资者提问每条平均虚词字数 |
| **控制变量：** |  |
| 提问数量（question） | 定义为投资者的提问数量 |
| 公司规模（Size） | 用公司的流通市值作为公司规模的度量指标，在计量模型中，对该指标进行对数化处理。 |
| 公司成长性（TobinQ） | 定义为公司的市场价值与账面价值比。 |
| 财务杠杆（Lev） | 定义为公司的净负债与总资产比值。 |
| 资产收益率（Roa） | 定义为公司的净利润与总资产比值。 |
| 公司年龄（age） | 用公司上市后的月份数作为公司年龄的度量指标，在计量模型中，对该指标作了对数化处理。 |
| 是否四大审计（big4） | 虚拟变量，如果公司是由四大审计的等于1，否则等于0。 |

## 实证模型

参考孟庆斌等（2020）设定了以下模型。

采用模型（3）对假设H1进行检验，即投资者提问的语气越负面股价同步性越低，这里我们会将将每年提问总数（question）也作为控制变量加入。

 （3）

ContralVariables作为控制变量，具体定义见表1，不难看到，如果假设H1成立，则β1应该显著为负。

采用模型（4）对假设H2进行检验，即在其他条件不变的情况下，股价同步性与平均长度应该相关性较弱。

 （4）

## 本章小结

本章研究数据搜集、变量设定、模型建构、数据处理和描述性统计。本章以上交互动易平台上2014年-2019年中的所有提问-回复样本作为研究的基础数据来源。我们借鉴了传统研究（朱红军等，2007；Cline等，2018；胡军和王甄，2015），控制了公司规模，公司成长性和资产收益率等变量，采用了固定效应模型，控制了行业和年份效应（由于可能会有行业或者是年份特定的新闻）。

# 主要结果

研究发现，投资者提问中含有负面情绪的提问数量越多，所关注公司的股价同步性越低；考虑到文本质量的问题，平均投资者提问长度与股价同步性相关性不显著。以上主要发现在变更主要变量模型、变换样本年份区间、更改控制变量后依然稳健。

## 描述性统计

表2为各变量的描述型统计，SYN1的均值为-0.923，方差为1.305；提问平均长度len\_question的均值和方差分别为3.997和0.327，提问数量question的均值和方差分别为3.26和1.120，回复时间的均值和方差分别为2.925和1.467，投资者负面情绪指标n\_question\_2的均值和方差分别为2.525和1.065。说明样本期内股价同步性及各解释变量存在充分变异，保证了本文的可行性。

表 5-1 描述性统计

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | N | mean | sd | min | p25 | p50 | p75 | max |
| SYN1 | 5617 | -0.923 | 1.305 | -14.03 | -1.497 | -0.717 | -0.0716 | 2.071 |
| len question | 5607 | 3.997 | 0.327 | 2.398 | 3.821 | 4.022 | 4.209 | 5.303 |
| question | 5617 | 3.206 | 1.120 | 0 | 2.485 | 3.258 | 4.007 | 7.129 |
| reply time | 5617 | 2.925 | 1.467 | 0 | 1.782 | 2.717 | 3.845 | 7.729 |
| n question 2 | 5617 | 2.525 | 1.065 | 0 | 1.792 | 2.565 | 3.258 | 6.165 |
| size | 5614 | 22.68 | 1.436 | 19.95 | 21.65 | 22.48 | 23.57 | 26.96 |
| age | 5614 | 12.42 | 7.971 | 0 | 4 | 14 | 19 | 29 |
| tobin | 5525 | 1.962 | 2.576 | 0.694 | 1.161 | 1.474 | 2.085 | 102.4 |
| roa | 5614 | 0.0412 | 0.0490 | -0.147 | 0.0152 | 0.0356 | 0.0665 | 0.187 |
| lev | 5223 | 1.438 | 1.091 | 0.553 | 0.994 | 1.100 | 1.382 | 8.553 |
| anaattention | 3990 | 9.915 | 10.19 | 1 | 3 | 6 | 14 | 75 |
| reportatte~n | 3993 | 21.08 | 26.76 | 1 | 4 | 11 | 28 | 253 |

从相关系数表来看（见表3）：投资者提问负面语气与股价同步性收益率的相关系数为-0.031，在10%的水平下显著，说明投资者提问的语气越负面，未来股价同步性越低，初步证明了假设H1。投资者提问平均长度与股价同步性之间的Pearson相关系数较为不显著，我们在后续会进行更深一步的讨论。

表 5-2 主要变量的Pearson和Spearman相关系数

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | SYN1 | len\_question | question | reply\_time | n\_question\_2 | size | age | tobin | roa | lev |
| SYN1 | 1 | 0.036\*\* | -0.041\*\* | 0.037\*\* | -0.046\*\*\* | 0.179\*\*\* | 0.158\*\*\* | -0.204\*\*\* | -0.146\*\*\* | 0.043\*\*\* |
| len\_question | 0.022 | 1 | 0.293\*\*\* | -0.180\*\*\* | 0.253\*\*\* | 0.089\*\*\* | 0.097\*\*\* | 0.012 | 0.022 | -0.021 |
| question | -0.030\* | 0.333\*\*\* | 1 | -0.291\*\*\* | 0.965\*\*\* | 0.062\*\*\* | 0.009 | -0.066\*\*\* | 0.121\*\*\* | -0.068\*\*\* |
| reply\_time | 0.014 | -0.195\*\*\* | -0.308\*\*\* | 1 | -0.276\*\*\* | 0.126\*\*\* | 0.095\*\*\* | -0.046\*\*\* | -0.106\*\*\* | 0.094\*\*\* |
| n\_question\_2 | -0.031\* | 0.288\*\*\* | 0.960\*\*\* | -0.294\*\*\* | 1 | 0.077\*\*\* | 0.005 | -0.080\*\*\* | 0.119\*\*\* | -0.059\*\*\* |
| size | 0.155\*\*\* | 0.088\*\*\* | 0.063\*\*\* | 0.113\*\*\* | 0.073\*\*\* | 1 | 0.419\*\*\* | -0.614\*\*\* | -0.316\*\*\* | 0.377\*\*\* |
| age | 0.147\*\*\* | 0.080\*\*\* | 0.001 | 0.092\*\*\* | -0.001 | 0.372\*\*\* | 1 | -0.149\*\*\* | -0.255\*\*\* | 0.231\*\*\* |
| tobin | -0.080\*\*\* | 0.002 | -0.018 | 0.006 | -0.026 | -0.282\*\*\* | -0.015 | 1 | 0.395\*\*\* | -0.366\*\*\* |
| roa | -0.143\*\*\* | 0.039\*\* | 0.117\*\*\* | -0.086\*\*\* | 0.112\*\*\* | -0.282\*\*\* | -0.234\*\*\* | 0.244\*\*\* | 1 | -0.625\*\*\* |
| lev | 0.038\*\* | -0.024 | -0.072\*\*\* | 0.049\*\*\* | -0.065\*\*\* | 0.174\*\*\* | 0.144\*\*\* | -0.096\*\*\* | -0.382\*\*\* | 1 |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | | | | | | | | | |

## 对传统结论的检验

我们利用模型对假设进行检验，结果如下，提问数量系数并不显著，平均提问长度系数分别为0.173和0.148并且都在1%的程度上显著。

表 5-3投资者提问数量、平均提问长度与股价同步性

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | SYN1 | SYN1 | SYN1 | SYN1 |
| question | 0.013 | -0.004 |  |  |
|  | (0.015) | (0.016) |  |  |
| len\_question |  |  | 0.160\*\*\* | 0.146\*\* |
|  |  |  | (0.047) | (0.059) |
| Controls\_1 | No | Yes | No | Yes |
| Controls\_2 | Yes | Yes | Yes | Yes |
| Year | 13\_19 | 13\_19 | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.255 | 0.273 | 0.258 | 0.300 |
| F | 0.816 | 23.898 | 11.462 | 17.334 |
| Observations | 5609 | 5139 | 5599 | 3738 |
| Year\_FE | Yes | Yes | Yes | Yes |
| Industry\_FE | Yes | Yes | Yes | Yes |
| Ownership\_FE | Yes | Yes | Yes | Yes |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

为了追求结果的稳健性，我们尝试了更改因变量计算模型与增加公司聚类的方法，结果均验证了结论的稳健性。这与过往研究相悖，针对媒体平台的研究表示信息数量的增加能够促进信息传播（Blankespoor，2014），而岑维（2016）针对互动性平台的研究也表明，提问数量的提升能够贡献私有信息，增加投资者关注度，减少公司和个人投资者之间的沟通问题，因此能够降低股价同步性。

本文的基础研究与过往文献相悖，我们猜测投资者情绪和文本质量起到了一定影响。倪景铄（2017）指出投资者的非理性情绪会放大公司的股价同步性，而周耿等（2020）则通过文本向量化测量了文本质量，并且发现文本信息含量对信息传递也有影响。因此我们需要进一步探讨情绪与文本质量的影响。

表 5-4 投资者提问数量、平均提问长度与更换模型的股价同步性

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | SYN2 | SYN2 | SYN3 | SYN3 |
| question | 0.009 |  | 0.019\* |  |
|  | (0.012) |  | (0.011) |  |
| len\_question |  | 0.102\*\* |  | 0.095\*\* |
|  |  | (0.047) |  | (0.040) |
| Controls1 | Yes | Yes | Yes | Yes |
| Controls2 | Yes | Yes | Yes | Yes |
| Year | 13\_19 | 13\_19 | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.272 | 0.298 | 0.285 | 0.309 |
| F | 35.407 | 22.153 | 31.643 | 20.192 |
| Observations | 5137 | 3736 | 5137 | 3736 |
| Year\_FE | Yes | Yes | Yes | Yes |
| Industry\_FE | Yes | Yes | Yes | Yes |
| Ownership\_FE | Yes | Yes | Yes | Yes |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

表 5-5 控制公司固定效应：投资者提问数量、平均提问长度与股价同步性

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | SYN1 | SYN1 |
| question | -0.004 |  |
|  | (0.016) |  |
| len\_question |  | 0.146\*\* |
|  |  | (0.058) |
| Controls\_1 | Yes | Yes |
| Controls\_2 | Yes | Yes |
| Year | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.273 | 0.300 |
| F | 17.369 | 16.405 |
| Observations | 5139 | 3738 |
| Firm\_FE | Yes | Yes |
| Year\_FE | Yes | Yes |
| Industry\_FE | Yes | Yes |
| Ownership\_FE | Yes | Yes |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

## 投资者提问的负面语气与股价同步性

我们猜想是否是由于投资者提问时的负面语气导致的，于是我们利用模型（1）对假设H1进行检验，并且我们更换了解释变量为投资者积极提问次数，控制提问总次数。结果如表7所示。投资者消极提问和积极提问的系数分别为-0.169和0.119，且均在1%的水平下显著。说明投资者负面语气提问越多，股价同步性越低，这就证明了本文的H1假设。并且，积极提问的系数也解释了上文的疑问：为什么提问越多，股价同步性反而越低。经济意义上，对于一家公司，投资者每以消极的态度提问一次，公司股价同步性就会下降0.169.反之，投资者每做出一次积极提问，股价同步性就会升高0.119.统计意义和经济意义上均说明，在控制公司，行业等效应后，投资者提问仍然具有信息含量，其传递的负面情绪能够影响其他的投资者，从而降低公司股价同步性。

表5-6 投资者负面提问语气、积极提问语气与股价同步性

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | SYN1 | SYN1 |
| n\_question\_2 | -0.169\*\*\* |  |
|  | (0.052) |  |
| p\_question\_8 |  | 0.119\*\*\* |
|  |  | (0.034) |
| Controls\_1 | Yes | Yes |
| Controls\_2 | Yes | Yes |
| Year | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.274 | 0.274 |
| F | 22.268 | 22.454 |
| Observations | 5139 | 5139 |
| Year\_FE | Yes | Yes |
| Industry\_FE | Yes | Yes |
| Ownership\_FE | Yes | Yes |  |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

方健和任春苗（2021）的研究表明，投资者在互动性平台上的负面情绪表达能够显著降低股价同步性。行为金融理论认为，投资者的理性是有限的，投资者做出买入卖出的行为会受到环境或者个人影响。在网络互动平台中，个人对企业的消极情绪会导致他们传播消极言论，并且影响投资者的情绪波动。这些会影响到其他投资者（Hoffman ML，2001），并且由于我国投资者大部分为个人投资者，换手率高，投资交易频繁，投机性强（段江娇等，2017），市场难以达到均衡，所以负面信息会高速传播。

## 投资者提问平均长度与股价同步性

主回归显示的另一个问题是投资者提问平均长度越长，股价同步性越高，并且投资者提问正显著，这不符合前人的研究。因此我们猜想投资者平均提问长度中另一个关键变量是信息质量，本文假设提问可以按照长度分为两组，并且两组长度有显著的差距，长度较低的那组信息质量随着提问平均长度而提升，长度较高的那组信息质量提升不明显。本文用虚词数量来衡量信息质量，控制长度后，虚词越多的文本信息质量越低，会降低股价同步性。通过平均虚词数量与平均提问长度的相关性分析（见表8），相关系数为0.937并且在1%上显著，这部分说明我们猜想的正确性，文本质量和文本长度之间的相对关系需要进一步的考量。

表 5-7 平均提问长度与平均虚词数量的相关性分析

|  |  |  |
| --- | --- | --- |
|  | len\_question | len\_fword |
| len\_question | 1 | . |
| len\_fword | 0.937\*\*\* | 1 |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | | |
|  | | |

表9和表10中的T检验证明了提问内部的确可以根据50%分位点或者是大于90%，亦或是小于10%分位点分为两组，并且均值差异显著。

表5-8以平均提问长度50%分组的平均虚词数量T检验表

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | obs1 | obs2 | Mean1 | Mean2 | dif | St Err | t value | p value |
| len fword by len 5~1 | 2803 | 2804 | 2.728 | 3.15 | -.421 | .005 | -76.75 | 0 |

表5-9以平均提问长度小于10%和高于90%分组的平均虚词数量T检验表

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | obs1 | obs2 | Mean1 | Mean2 | dif | St Err | t value | p value |
| len fword by x: 0 1 | 552 | 561 | 2.621 | 3.144 | -.523 | .022 | -24.25 | 0 |

根据孟庆斌（2020）的研究，文本信息质量能够显著影响信息传递，有的信息大多是敷衍了事，多使用虚词如感谢，您好等等，没有传递任何有效信息，因此平均文本质量并不能完全代表文本的信息含量，其与股价同步性的关系也并不显著。这意味着传统观点中由于过度关注传统媒体渠道如研究报告和监管报告等文本质量相对较高的渠道，忽略了文本质量低下的问题，这警示监管机构要加强投资者教育，提升投资者提问质量。

## 稳健性检验

上文实证结果表明，投资者情绪确实影响了股价同步性，主要是通过投资者情绪传播实现的。进一步研究发现，投资者提问质量也会影响信息含量。然而考虑到其他因素的影响，以及样本年份包括了A股股市波动较大的年份，为了进一步保证结果的稳健性，本文还进行了以下检验：

1. **采用其他方法计算股价同步性**

考虑到传统计算股价同步性的方法或许不够完善，参考孟庆斌等（2020）的方法，我们选取了另外两种常用的计算股价同步性的方法，做了主回归，发现结果和上文一致，证明了结论的稳健性。

负面情绪提问能够显著降低股价同步性，这是因为负面情绪的传播能够激发投资者关注，从而促进私有信息的传播，降低信息不均衡

表5-10 选择另外两个主变量进行回归

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
|  | SYN2 | SYN2 | SYN3 | SYN3 |
| n\_question\_2 | -0.147\*\*\* |  | -0.137\*\*\* |  |
|  | (0.041) |  | (0.036) |  |
| p\_question\_8 |  | 0.086\*\*\* |  | 0.088\*\*\* |
|  |  | (0.027) |  | (0.023) |
| Controls\_1 | Yes | Yes | Yes | Yes |
| Controls\_2 | Yes | Yes | Yes | Yes |
| Year | 13\_19 | 13\_19 | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.273 | 0.273 | 0.287 | 0.287 |
| F | 32.628 | 32.266 | 29.629 | 29.520 |
| Observations | 5137 | 5137 | 5137 | 5137 |
| Year\_FE | Yes | Yes | Yes | Yes |
| Industry\_FE | Yes | Yes | Yes | Yes |
| Ownership\_FE | Yes | Yes | Yes | Yes |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

1. **公司聚类的方法**

参考胡军和王甄（2015），为了缓解可能存在的残差项方差和相关性问题，采用了公司聚类的方法重新验证显著性，发现结果不变，证明了其稳健性。

表5-11增加公司聚类进行主回归

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | SYN1 | SYN1 |
| n\_question\_2 | -0.169\*\*\* |  |
|  | (0.052) |  |
| p\_question\_8 |  | 0.119\*\*\* |
|  |  | (0.034) |
| Controls\_1 | Yes | Yes |
| Controls\_2 | Yes | Yes |
| Year | 13\_19 | 13\_19 |
| Adjusted R-squared | 0.274 | 0.274 |
| F | 16.499 | 16.504 |
| Observations | 5139 | 5139 |
| Year\_FE | Yes | Yes |
| Industry\_FE | Yes | Yes |
| Ownership\_FE | Yes | Yes |

Standard errors in parentheses

\* *p* < 0.10, \*\* *p* < 0.05, \*\*\* *p* < 0.01

注：controls\_1 为控制变量表中的公司层面的控制变量，controls\_2为是否控制公司提问数量和平均提问长度。

1. **小结**

通过以上的两种方法进行了稳健性检验，使用的方法包括更换因变量，增加公司聚类等，这些方法在一定程度上加强了结果的稳健性。

## 本章小结

本章研究投资者情绪与文本质量对股价同步性的影响。传统观点认为，股价同步性代表着私有信息融入股价的程度，因此互动性平台上提问数量/长度会显著影响股价同步性，但是这与我们的实证结果相悖，并且在经历了更换变量等稳健性检验后结果依然稳健。我们认为这是因为投资者情绪与提问文本质量的影响。

描述性统计表明样本期内股价同步性及各解释变量存在充分变异，保证了本文的可行性，并且投资者提问负面语气与股价同步性呈现负相关的情况，但是投资和平均提问长度与股价同步性之间的关系仍不明显。

进一步研究，本文发现，互动性平台上的负面提问越多，股价同步性越低。这主要是因为负面情绪传播性较强，能够更好在网络空间中传播私有信息。另外，平均提问长度对股价同步性影响不显著，是因为提问质量不高，虚词数量过多。

本文在以下方面丰富了现有文献，具有较强的理论意义。本文首先对传统研究中认为信息数量能够显著影响资本市场中私有信息的理论（岑维，2016）给出了反例，并且给出了投资者情绪和文本质量这两个影响因素。过往研究主要从投资者关注度的角度研究信息的传播效果（Blanespoor，2014；Liu，2011），本文侧重网络空间中的情绪传播，因为投资者天生对负面消息更为敏感，网络空间会放大投资者对负面情绪的关注。本文进一步研究提问质量对股价同步性的影响，不同于传统研究（张洁，2018），本文证实了互动性平台上平均信息长度并不能显著影响股价同步性，这是因为提问质量无法和平均信息长度挂钩，必须要用更好的方法来衡量股价同步性。

本文的主要研究启示监管机构需要进一步重视投资者负面提问对信息的传播作用，推动企业反应负面消息。另外也是要加强对投资者的教育，提升提问质量。

# 结束语

本文以2014年1月1日至2019年12月31日上交所上市为研究对象，考察了“互动易”平台上投资者提问负面语气与投资者平均提问长度的信息含量。

本文研究发现，在控制行业和年份影响的情况下，投资者提问的负面语气能够显著降低股价同步性，另外，投资者平均提问长度的增加并不意味着信息质量的增加，只有当投资者平均提问长度较低时平均提问长度的提升才能降低股价同步性。

尽管本文从虚词数量验证了文本质量与股价同步性没有太高的相关性，但是终究还是缺乏进一步的实证研究，因为衡量文本质量不完全依赖于虚词。股价同步性是衡量市场上是否有充分的公司私有信息的（Morck et al.,2000）。但是事实上就提问而言，看起来完全不同的提问可能并没有提供新的私有信息，由于语言的暧昧性，我们需要使用更好的方法来衡量不同的提问是否提供了增量信息。

我们知道虚词没有办法提供增量信息，于是我们处理数据时首先会剔除掉大部分虚词。另外，我们来考虑两种词语，举个例子，比如说“红利政策”可能是经常出现的词语，在过去的大部分时间段，已经有很多关于这方面的提问，如果一个人询问红利政策，那么这个词语很可能并没有给出一个增量的私有信息。但是，如果在一个时间段内，大量的人询问“某某机场”的建立是否对公司有影响。这实际上意味着这个词语传播力度广，而且由于这个词语总体上出现次数很少，我们可以认为它携带有更多的私有信息。

衡量这种词语的算法叫做：TF-IDF值算法，该算法是用来衡量某一个特定文本段内高频率出现，但是在整个内容中的低频率的情况。假设我们有10段文本，如果一个词语在某一段中出现频率越高，但是在整篇文章中出现次数较少，它的TF-IDF值就越高。该算法也常用于关键词衡量等等。

我们试图使用该算法来衡量文本质量，但是结论并不显著，这很大概率是因为TF-IDF值无法处理一词多义的情况。例如偿债能力和更好的负担负债其实是一个意思，但是TF-IDF值无法分辨。

我们进行了相关研究，但是由于能力原因并没有成功完成最后的测算。

本文的研究在理论上对公司管理层与投资者信息交流方面的研究起到了有益的补充。同时，从本文的结论总还可以得到如下的政治启示：其一，我国资本市场的建设不仅仅要强调对投资者投资者的保护，监管部门也还需要增加其积极性并发挥其有益的信息发现能力。其二。本文发现个人投资者虽然是有信息的，但是能够合理通过提问传递信息的是少数人。总体而言，大部分投资者无法通过互动平台提升资本市场信息效率。这就说明，继续加强个人投资者教育，提高个人投资者交易水平仍然任重道远。

# 致谢

我在喻家山下完成了四年的大学生涯，即将结束自己并不成熟的少年时光，投入到研究生的工作与学习中。时间匆匆流逝，不觉已经站在离开的门口。着实怀念在校园生活的日子，怀念与同学们老师们相处的时光。

感谢自己大创小组，我们一起参与了很多项目和课后作业，大家合作很愉快，并且也成为了很好的朋友，祝愿大家万事如意。

感谢李安泰老师，我当时发送第一封邮件想要和老师沟通时，老师很亲切地回复了我，给了我很多帮助，祝愿您科研上万事如意。

感谢父母，在我心情低落时给了我很多支持，也在人生道路的十字路口给了我很多关键的建议，帮助我走了不会后悔的道路。

这四年里我学习到了很多知识，也成长为和四年前完全不同的人，人生的路还很漫长，我必然会继续前进。

何卓扬

2022年5月9日

# 参考文献

1. Blankespoor E, Miller G S, White H D. The role of dissemination in market liquidity: Evidence from firms' use of Twitter™[J]. The Accounting Review, 2014, 89(1): 79-112.
2. Bushee B J, Jung M J, Miller G S. Do investors benefit from selective access to management?[J]. Journal of Financial Reporting, 2017, 2(1): 31-61.
3. Bushman R M, Piotroski J D, Smith A J. What determines corporate transparency?[J]. Journal of accounting research, 2004, 42(2): 207-252.
4. Crawford，S．S．，D．T．Roulstone，and E．C．So． Analyst Initiations of Coverage and Stock Return Synehronicity [J]．Accounting Review，2012，87(5) : 1527-1553．
5. Cline B N, Walkling R A,Yore A S. The Consequences of Managerial Indiscretions:Sex,Lies,and Firm Value[J].Journal of Financial Economics,2018,127(2): 389-415.
6. Das S R, Chen M Y. Yahoo! for Amazon: Sentiment extraction from small talk on the web[J]. Management science, 2007, 53(9): 1375-1388.
7. Davis A K, Piger J M, Sedor L M. Beyond the numbers: Measuring the information content of earnings press release language[J]. Contemporary Accounting Research, 2012, 29(3): 845-868.
8. Durnev, Art, Randall Morck, and Bernard Yeung. Value－Enhancing Capital Budgeting and Firm － Specific Stock Return Variation[J], The Journal of Finance, 2004, 59(1) : 65-105.
9. Eun C S， Wang L， Xiao S C．Culture and R2 [J]．Journal of Financial Economics，2015，115（2）：283-303．
10. Gul F A, Kim J B, Qiu A A. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China[J]. Journal of financial economics, 2010, 95(3): 425-442.
11. Hoffman ML.How Automatic and Representational is Empathy,and Why[J].Behavioral&Brain Ences,2001,25(1):38~39.
12. Kim E H, Youm Y N. How do social media affect analyst stock recommendations? Evidence from S&P 500 electric power companies' Twitter accounts[J]. Strategic Management Journal, 2017, 38(13): 2599-2622.
13. Kimbrough M D. The effect of conference calls on analyst and market underreaction to earnings announcements[J]. The Accounting Review, 2005, 80(1): 189-219.
14. Kirk M P, Markov S. Come on over: Analyst/investor days as a disclosure medium[J]. The Accounting Review, 2016, 91(6): 1725-1750.
15. Li F. Annual report readability, current earnings, and earnings persistence[J]. Journal of Accounting and economics, 2008, 45(2-3): 221-247.
16. Liu, M.H.C.and Z.Zhuang．Management Earnings Forecasts and the Quality of Analysts' Forecasts：The Moderating Effect ofAudit Committees[J]．Journal of Contemporary Accounting&Economics, 2011, 7 (1) :31-45.
17. Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks[J]. The Journal of finance, 2011, 66(1): 35-65.
18. Morck R, Yeung B, Yu W. 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.
19. Teoh S H, Yang Y G, Zhang Y. R-square and market efficiency[J]. Available at SSRN 926948, 2009.
20. Tetlock P C, Saar‐Tsechansky M, Macskassy S. More than words: Quantifying language to measure firms' fundamentals[J]. The Journal of Finance, 2008, 63(3): 1437-1467.
21. West K D. Dividend innovations and stock price volatility[J]. Econometrica: Journal of the Econometric Society, 1988: 37-61.
22. 岑维,李士好,童娜琼.投资者关注度对股票收益与风险的影响——基于深市“互动易”平台数据的实证研究[J].证券市场导报,2014(07):40-47.
23. 岑维,童娜琼,何潇悦.投资者关注度与中小股东利益保护——基于深交所“互动易”平台数据的实证研究[J].证券市场导报,2016(02):54-62.
24. 陈晓娟. 投资者情绪对股票价格的影响研究[D].南京邮电大学,2020.
25. 丁慧,吕长江,黄海杰.社交媒体、投资者信息获取和解读能力与盈余预期——来自“上证e互动”平台的证据[J].经济研究,2018,53(01):153-168.
26. 段江娇,刘红忠,曾剑平.中国股票网络论坛的信息含量分析[J].金融研究,2017(10):178-192.
27. 方健,任春苗.社交媒体情绪表达对股价波动的影响研究——基于“上证e互动”的实验证据[J].金融与经济,2021,(07):12-21.
28. 何贤杰,王孝钰,赵海龙,陈信元.上市公司网络新媒体信息披露研究:基于微博的实证分析[J].财经研究,2016,42(03):16-27.
29. 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.
30. 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(05):121-130.
31. 蒋艳辉,郑佳尔.新媒体信息披露对股票市场的影响研究——来自“互动易”平台的证据[J].会计之友,2017(08):48-55.
32. 李新丽,万寿义,张雪梅.企业社会责任能降低股价同步性吗?——基于强制披露企业社会责任报告的准自然实验[J].科学决策,2018(09):1-27.
33. 李增泉. 所有权结构与股票价格的同步性——来自中国股票市场的证据[J]. 中国会计与财务研究, 2005, 7(3): 57-100.
34. 刘姗. 公司特质性信息对股价同步性的影响[D].华东政法大学,2017.
35. 马连福,赵颖.上市公司社会责任信息披露影响因素研究[J].证券市场导报,2007(03):4-9.
36. 孟庆斌,黄清华,张劲帆,王松.上市公司与投资者的互联网沟通具有信息含量吗?——基于深交所“互动易”的研究[J].经济学(季刊),2020,19(02):637-662.
37. 倪景铄. 个人投资者和机构投资者情绪与股价同步性研究[D].湖南大学,2015.
38. 沈洪涛,万拓,杨思琴.我国企业社会责任报告鉴证的现状及评价[J].审计与经济研究,2010,25(06):68-74.
39. 谭松涛,阚铄,崔小勇.互联网沟通能够改善市场信息效率吗?——基于深交所“互动易”网络平台的研究[J].金融研究,2016(03):174-188.
40. 汪昌云,武佳薇.媒体语气、投资者情绪与IPO定价[J].金融研究,2015(09):174-189.
41. 王亚平,刘慧龙,吴联生.信息透明度、机构投资者与股价同步性[J].金融研究,2009(12):162-174.
42. 王艳艳,于李胜,安然.非财务信息披露是否能够改善资本市场信息环境?——基于社会责任报告披露的研究[J].金融研究,2014(08):178-191.
43. 夏玉芹,单雪微.基于Python的简单文本情感分析[J].阴山学刊(自然科学版),2018,32(04):58-62.
44. 许年行,洪涛,吴世农,徐信忠.信息传递模式、投资者心理偏差与股价“同涨同跌”现象[J].经济研究,2011,46(04):135-146.
45. 杨凡,张玉明.网络媒介、互动式信息披露与分析师行为——来自“上证e互动”的证据[J].山西财经大学学报,2020,42(11):113-126.
46. 杨扬. 企业社会责任信息披露对股价同步性影响的研究[D].北京交通大学,2019.
47. 张洁. 微信公众号信息披露与股价同步性关系研究[D].哈尔滨工业大学,2018.
48. 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(08):109-120.
49. 周耿,周舟,范从来.上证“e互动”对市场信息效率影响及机制研究[J].上海经济研究,2020(02):110-121.
50. 周世庆. 投资者情绪与网络新闻对股价同步性的影响研究[D].哈尔滨工业大学,2020.DOI:10.27061/d.cnki.ghgdu.2020.000552.
51. 朱红军,何贤杰,陶林.中国的证券分析师能够提高资本市场的效率吗——基于股价同步性和股价信息含量的经验证据[J].金融研究,2007(02):110-121.



**本科毕业设计（论文）任务书**

题 目 投资者情绪、文本质量和股价同步性

——基于互动易的研究

（任务起止日期：2021年11月2日～2022年6月5日）

院 系 管理学院财务金融系

专业班级 财务管理1801

姓 名 何卓扬

学 号 U201815934

指导教师 李安泰

教研室（系、所）负责人 2021年 11月 2日审查

院（系）负责人 2022年 2月 21日批准

|  |
| --- |
| 课题内容：  股价同步性是衡量资本市场定价效率与资源配置效率的重要指标，党的十九大报告中指出，中国金融市场的改革方向是“深化金融体制改革，增强金融服务实体经济能力”。探索如发现在变更主要变量模型、变换样本年份区间、更改控制变量后依然稳健。 |
| 课题任务要求：  本课题的研究，至少要达到如下方面的预期目标：   * 1. 系统介绍股价同步性的一般理论   2. 实证分析我国上市公司投资者关系管理的工作的现状与作用   3. 分析网络互动平台对投资者关系的意义   4. 研究投资者提问的负面语气与股价同步性的相关关系   5. 研究投资者提问的平均文本长度与股价同步性的相关关系。   6. 力争在期刊上发表几篇相应研究论文。 |
| 主要参考文献（由指导教师选定）：  Blankespoor E, Miller G S, White H D. The role of dissemination in market liquidity: Evidence from firms' use of Twitter™[J]. The Accounting Review, 2014, 89(1): 79-112.  Bushee B J, Jung M J, Miller G S. Do investors benefit from selective access to management?[J]. Journal of Financial Reporting, 2017, 2(1): 31-61.  Bushman R M, Piotroski J D, Smith A J. What determines corporate transparency?[J]. Journal of accounting research, 2004, 42(2): 207-252.  Das S R, Chen M Y. Yahoo! for Amazon: Sentiment extraction from small talk on the web[J]. Management science, 2007, 53(9): 1375-1388.  Davis A K, Piger J M, Sedor L M. Beyond the numbers: Measuring the information content of earnings press release language[J]. Contemporary Accounting Research, 2012, 29(3): 845-868.  Durnev, Art, Randall Morck, and Bernard Yeung. Value－Enhancing Capital Budgeting and Firm － Specific Stock Return Variation[J], The Journal of Finance, 2004, 59(1) : 65-105.  Gul F A, Kim J B, Qiu A A. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China[J]. Journal of financial economics, 2010, 95(3): 425-442.  Kim E H, Youm Y N. How do social media affect analyst stock recommendations? Evidence from S&P 500 electric power companies' Twitter accounts[J]. Strategic Management Journal, 2017, 38(13): 2599-2622.  Kimbrough M D. The effect of conference calls on analyst and market underreaction to earnings announcements[J]. The Accounting Review, 2005, 80(1): 189-219.  Kirk M P, Markov S. Come on over: Analyst/investor days as a disclosure medium[J]. The Accounting Review, 2016, 91(6): 1725-1750.  Li F. Annual report readability, current earnings, and earnings persistence[J]. Journal of Accounting and economics, 2008, 45(2-3): 221-247.  Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks[J]. The Journal of finance, 2011, 66(1): 35-65.  Morck R, Yeung B, Yu W. 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.  Teoh S H, Yang Y G, Zhang Y. R-square and market efficiency[J]. Available at SSRN 926948, 2009.  Tetlock P C, Saar‐Tsechansky M, Macskassy S. More than words: Quantifying language to measure firms' fundamentals[J]. The Journal of Finance, 2008, 63(3): 1437-1467.  West K D. Dividend innovations and stock price volatility[J]. Econometrica: Journal of the Econometric Society, 1988: 37-61.  岑维,李士好,童娜琼.投资者关注度对股票收益与风险的影响——基于深市“互动易”平台数据的实证研究[J].证券市场导报,2014(07):40-47.  岑维,童娜琼,何潇悦.投资者关注度与中小股东利益保护——基于深交所“互动易”平台数据的实证研究[J].证券市场导报,2016(02):54-62.  陈晓娟. 投资者情绪对股票价格的影响研究[D].南京邮电大学,2020.  丁慧,吕长江,黄海杰.社交媒体、投资者信息获取和解读能力与盈余预期——来自“上证e互动”平台的证据[J].经济研究,2018,53(01):153-168.  何贤杰,王孝钰,赵海龙,陈信元.上市公司网络新媒体信息披露研究:基于微博的实证分析[J].财经研究,2016,42(03):16-27.  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.  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(05):121-130.  蒋艳辉,郑佳尔.新媒体信息披露对股票市场的影响研究——来自“互动易”平台的证据[J].会计之友,2017(08):48-55.  李新丽,万寿义,张雪梅.企业社会责任能降低股价同步性吗?——基于强制披露企业社会责任报告的准自然实验[J].科学决策,2018(09):1-27.  李增泉. 所有权结构与股票价格的同步性——来自中国股票市场的证据[J]. 中国会计与财务研究, 2005, 7(3): 57-100.  刘姗. 公司特质性信息对股价同步性的影响[D].华东政法大学,2017.  马连福,赵颖.上市公司社会责任信息披露影响因素研究[J].证券市场导报,2007(03):4-9.  孟庆斌,黄清华,张劲帆,王松.上市公司与投资者的互联网沟通具有信息含量吗?——基于深交所“互动易”的研究[J].经济学(季刊),2020,19(02):637-662.  沈洪涛,万拓,杨思琴.我国企业社会责任报告鉴证的现状及评价[J].审计与经济研究,2010,25(06):68-74.  谭松涛,阚铄,崔小勇.互联网沟通能够改善市场信息效率吗?——基于深交所“互动易”网络平台的研究[J].金融研究,2016(03):174-188.  汪昌云,武佳薇.媒体语气、投资者情绪与IPO定价[J].金融研究,2015(09):174-189.  王亚平,刘慧龙,吴联生.信息透明度、机构投资者与股价同步性[J].金融研究,2009(12):162-174.  王艳艳,于李胜,安然.非财务信息披露是否能够改善资本市场信息环境?——基于社会责任报告披露的研究[J].金融研究,2014(08):178-191.  夏玉芹,单雪微.基于Python的简单文本情感分析[J].阴山学刊(自然科学版),2018,32(04):58-62.  许年行,洪涛,吴世农,徐信忠.信息传递模式、投资者心理偏差与股价“同涨同跌”现象[J].经济研究,2011,46(04):135-146.  杨凡,张玉明.网络媒介、互动式信息披露与分析师行为——来自“上证e互动”的证据[J].山西财经大学学报,2020,42(11):113-126.  杨扬. 企业社会责任信息披露对股价同步性影响的研究[D].北京交通大学,2019.  张洁. 微信公众号信息披露与股价同步性关系研究[D].哈尔滨工业大学,2018.  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(08):109-120.  周耿,周舟,范从来.上证“e互动”对市场信息效率影响及机制研究[J].上海经济研究,2020(02):110-121.  朱红军,何贤杰,陶林.中国的证券分析师能够提高资本市场的效率吗——基于股价同步性和股价信息含量的经验证据[J].金融研究,2007(02):110-121. |
| 同组设计者：  无 |
| 指导教师签名：  2021 年 10 月 28 日 |