

学校代码: 10272

学号:2014210452

上海财经大学

Shanghai University of Finance and Economics

硕士学位论文

Master Dissertation

投资者情绪对我国股市影响的实证研究

—基于东方财富股吧的文本挖掘分析

培养院（系）:	金融学院
学科名称:	金融硕士
论文类别:	硕士学位论文
论文作者:	王科柯
指导教师:	王明涛教授
学 校:	上海财经大学
论文完成时间:	2016年4月15日

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本人完全了解上海财经大学关于收集、保存、使用学位论文的规定，即：按照有关要求提交学位论文的印刷本和电子版本；上海财经大学图书馆有权保留学位论文的印刷本和电子版，并提供目录检索与阅览服务；可以采用影印、缩印、数字化或其它复制手段保存论文；在不以赢利为目的的前提下，可以公布论文的部分或全部内容。（保密论文在解密后遵守此规定）

论文作者签名：

导师签名：

日期： 年 月 日

日期： 年 月 日

摘要

在互联网技术飞速发展的今天,网络传媒已经不再仅仅影响我们的公共生活,更显著地改变了人们的信息获取方式和投资决策行为,并且网络传媒信息正在不断加大对金融市场传递和影响。尤其是在我国股票市场,散户众多,且股民与网民高度重叠,网络传媒已经成了中小投资者获取信息和直接表达情绪的重要平台;并且中小投资者的意见和情绪在这里传播,碰撞,融合,汇聚,并最终转化为实际的投资决策,实现信息向金融市场的传递,这也必然会对金融市场产生重要影响。但是国内外学术界目前从网络传媒信息的角度来研究投资者情绪对股票市场影响的研究还非常匮乏且不够深入。

本文的主要目的是研究中小投资者在股吧论坛发布交流的信息中所蕴含的投资者情绪是否会对股票市场的收益率和波动率产生重要的影响。为此,本文抓取了中国最大的股吧论坛—东方财富网在2014年6月至2015年12月期间近120万条发帖信息进行研究。文章利用网络文本挖掘与分类技术,提取其中带有投资者情感倾向的帖子,并在此基础上构建情绪指标—看涨指数,建立模型研究其对股票市场收益率和波动率的影响和预测作用以及在牛熊市环境下的异化现象。

研究发现:股吧论坛中投资者发布的大量帖子中,确实包含了能够影响甚至预测市场收益率和波动率的重要信息。

本文构建的看涨指数和股票市场收益率呈现49.5%的高度显著正相关,和股票市场波动率呈现-52.2%的高度显著负相关。并且,隔夜开盘前股吧论坛中确实含有能够影响和预测当日股市收益率和波动率的有效信息,但是周末或假期市场上有效信息较少,投资者情绪对市场收益率和对市场收益波动性的影响,都不存在显著的周末效应。

另外,本文发现投资者情绪对市场收益率和波动率的影响在牛熊市环境下存在明显的异化现象。隔夜开盘前股吧论坛中的情绪对当日开盘后市场收益率的影响在牛市环境中大约只能持续约30分钟,但在熊市环境中能持续2个小时以上,并且影响强度更大。在牛市环境下,市场收益的波动和投资者情绪对相互未来期的影响较小,而在熊市环境下,市场收益的波动和投资者情绪对相互未来期的影响较为显著,投资者情绪是引起市场收益波动变化的Granger原因,市场收益波动也是引起投资者情绪变动的Granger原因。

同时,乐观情绪与悲观情绪对股市影响存在显著的不对称性。乐观倾向的投资者情绪对市场收益的波动没有显著影响,对市场收益波动产生显著影响的主要

是悲观倾向的投资者情绪。同样，主要在熊市环境下，隔夜开盘前股吧论坛中的悲观情绪对当日市场收益的波动有非常显著的影响和预测作用，而乐观情绪则没有影响。

最后，本文通过与 3 个常见的传统投资者情绪指标（包括显性情绪指标和隐性情绪指标）实证对比分析，发现基于网络传媒的投资者情绪指标在对市场收益率和波动率的影响作用以及预测能力上优势明显。

本文在总结前人已有研究的宝贵经验和不足之处的基础上，进一步深入研究和分析网络文本信息中蕴含的投资者情绪对股票市场的影响，不仅考虑了数量型指标，更考虑了信息量丰富的文本型数据；并且本文根据股市交易时间对交易日进行了重新界定，使本文构建的情绪指标更加准确地度量该时间段或交易日的投资者情绪。此外，本文对隔夜开盘前的发帖信息单独构建情绪指标并研究其对开盘后股市的影响及其时效性。为了更加准确地度量股市的波动性，体现日内的股市交易活动信息，本文还采用当下最流行的已实现波动率来度量股市波动率。

本文以全新的思路和视角来研究股票市场的投资者情绪，为现代网络传媒和金融市场的结合研究提供了新的思路和方法，不仅为行为金融的发展有较大的理论意义，也为监管层制定相关有效的监管制度以及为投资者制定正确的投资决策，进行风险管理都有较大的现实意义。

关键词：网络传媒 投资者情绪 股票市场 收益率 波动率

Abstract

In the rapid development of Internet technology, online media affects our public life and is even changing the way people have access to information and investment decision-making behavior more significantly. The impact of the online media information on financial markets is increasing.

Network media has become an important platform to get information and directly express their emotion for small investors, especially in the stock market of our country with high proportion of individual investors and high degree of overlap of investors and netizens; and small investors' opinions and sentiments spread, collide, fuse, aggregate here and ultimately covert to the actual investment decision, which will have a significant impact on the financial markets. But now domestic and foreign study about the impact of investor sentiment on the stock market is still very little and not deep enough from the perspective of network media information.

The main purpose of this paper is to study whether small investors' sentiment implied in the information in the stock forum will have an important impact on the yield and volatility of the stock market. And then this article grabs nearly 1.2 million posts between June 1, 2014 and December 31, 2015 from China's largest stock forum-Eastmoney. This article extracts the posts with investors' sentiment, and then constructs a sentiment index- bullish index, establishes models to study the influence of this sentiment index on stock market return and volatility and the alienation phenomenon in bull and bear market environment through the technology of data mining and classification.

This paper finds that the large number of posts in the stock forum does contain important information that can affect or even predict the market rate of return and volatility.

The bullish index is 49.5% highly significant positive correlated with the return of stock market, and the bullish index is -52.2% highly significant negative correlated with realized volatility of stock market. And the stock forum overnight before opening of stock market does contain useful information that can influence and predict the day's return and volatility of stock market, but the influence of the bullish index on the return and volatility has no significant weekend effect for the stock forum has less efficient market information on weekends or holidays.

In addition, this article finds that the influence of investors' sentiment on the return and volatility has significant alienation phenomenon in Bull and Bear environments.

The influence of the sentiment from the information of the stock forum overnight before opening of stock market to the return after the opening can only continue for about 30 minutes in a bull market, which can continue for more than two hours, and greater impact strength in a bear market environment. And in a bull market, the volatility and investor sentiment has little effect in the future of each other, and in a bear market, the volatility and investor sentiment has significant effect in the future of each other, investor sentiment is Granger cause of the changes of volatility, and volatility is also Granger cause of the changes of investor sentiment.

At the same time, the influence of the optimistic and pessimistic sentiment has significant asymmetry. The optimistic sentiment has no significant impact on the volatility of the market but pessimistic sentiment has a significant impact on the volatility of the market. And mainly in the bear market environment, it is the pessimistic sentiment not the optimistic sentiment in the stock forum overnight before opening of stock market that has a very significant predictable effect on that day's volatility of market.

In the end, by contrast with three common traditional investor sentiment indicators (including explicit and implicit sentiment indicators), this article finds that the investor sentiment indicators based on network media has a significant advantage in the influence on and predicting the return as well as the volatility the of the stock market.

This article further studies and analyses the impact of the investor sentiment implied in the online information on the stock market based on previous studies, considering not only the quantitative indicators, but also text-based data which is rich of information; and this article redefines the trading day according to the trading time of stock market which will help measure the investor sentiment of trading day more accurately. In addition, the paper builds the bullish index separately based on posts overnight before the opening of stock market and research its aging impact on the stock market after the opening. In order to more accurately measure stock market volatility, reflecting more trading information in the trading day, this article uses realized volatility which is the currently most popular to measure stock market volatility.

This paper studies the stock market and investor sentiment with new ideas and from new perspective, providing new ideas and methods for binding study of financial markets and the modern web media. It not only has great theoretical significance for

the development of behavioral finance, but also has great practical significance for regulators formulating effective regulations and for investors making the right investment decisions and risk management.

Key Words: online media investor sentiment stock market return volatility

目录

第一章 导论	1
第一节 研究背景与研究意义.....	1
一、研究背景	1
二、研究意义	2
第二节 国内外文献综述.....	3
一、投资者情绪的定义.....	3
二、传统投资者情绪的度量.....	4
三、投资者情绪与股市关系的研究.....	7
四、基于网络传媒文本信息的投资者情绪研究.....	9
五、文献评述	10
第三节 研究思路与文章结构.....	12
一、研究思路	12
二、文章结构框架	13
第四节 本文创新之处与不足之处.....	14
一、本文的创新点	14
二、本文不足点	14
第二章 理论基础和研究技术	16
第一节 本文的理论基础.....	16
一、投资者情绪与金融市场.....	16
二、网络传媒与投资者情绪.....	20
第二节 本文的研究技术.....	21
一、网络文本采集	22
二、文本信息预处理	22
三、中文分词	22
四、特征词选取	23
五、向量空间表示	24

六、文本分类算法	24
第三章 研究设计与模型构建	26
第一节 指标构建与控制变量选取	26
一、基于网络文本的投资者情绪指标构建.....	26
二、交易日时间段界定.....	27
三、模型控制变量的选取.....	28
第二节 投资者情绪指标影响股市收益率的模型构建.....	29
一、投资者情绪对市场指数收益率的影响模型.....	29
二、投资者情绪与市场指数收益率的 VAR 模型.....	30
三、开盘前投资者情绪对上证指数收益率的预测影响模型.....	30
第三节 投资者情绪对股市波动性影响的模型构建.....	31
一、改进已实现波动率的计算.....	31
二、投资者情绪对股市波动性的影响分析.....	32
三、投资者情绪与股市波动性的 VAR 模型.....	33
四、开盘前投资者情绪对市场波动性的预测影响模型.....	33
第四章 研究样本与描述性统计	34
第一节 研究样本获取.....	34
一、样本时间选取	34
二、数据来源	34
第二节 研究样本的描述性统计.....	37
一、论坛发帖量统计描述.....	37
二、变量的描述性统计.....	39
第三节 变量的相关性分析.....	41
第五章 模型实证结果与分析	44
第一节 数据的平稳性检验.....	44
第二节 投资者情绪对市场指数收益率的影响分析.....	45
一、上证指数收益率的自相关性分析.....	45
二、投资者情绪对市场指数收益率的影响分析.....	46

目录

三、投资者情绪和市场收益率的相互影响—VAR 模型与格兰杰检验	48
四、开盘前投资者情绪对上证指数收益率的预测作用和时效性分析	51
第三节 投资者情绪对股市波动性的影响分析	54
一、上证指数波动率的自相关性分析	54
二、投资者情绪对市场指数波动性的影响分析	56
三、投资者情绪和上证指数波动性的 VAR 模型与格兰杰检验分析	58
四、开盘前投资者情绪对上证指数波动性的预测作用	61
第四节 基于网络文本的情绪指标与传统情绪指标对比	62
一、对比传统情绪指标选取与构建	62
二、不同情绪指标对股票市场影响的对比	64
第五节 实证结果分析	64
第六章 研究结论与展望	74
参考文献	76
致谢	80
个人简历及在学期间发表的研究成果	81

第一章 导论

第一节 研究背景与研究意义

一、研究背景

股票市场不仅是我国市场经济的重要组成部分，同时是观察一个国家经济总体运行状况的晴雨表，它能在一定程度上反映宏观经济、中观行业和微观企业的发展变化，从而成为前瞻分析和预测经济的重要参考。我国股票市场在 1990 年一成立便快速发展，到 2015 年底时，沪深两市已经有 2827 家上市公司，股票数量达到了 9353 只，总市值更是突破了 53 万亿元，比 2014 年增加 42.6%，并且 2009 年时就已是除美国以外全世界的第二大股票市场，正对中国甚至世界经济的发展产生越来越重要的影响。但是，我们也看到我国的股票市场还很不成熟，交易制度设计、市场监管等都不完善，内幕交易频繁发生，投资者利益不能得到保护，市场存在严重的非理性和剧烈波动，相比欧美的成熟资本市场，我们还有很长的路要走。

2015 年对中国股市注定将是难忘的一年，在宽松政策加上改革强烈预期的背景下，券商股的集体涨停拉开了此轮牛市的序幕，继而在杠杆资金的推动下股指一路飙升，不到三个月就达到 5178 点高峰，但紧随而来的是千股跌停的股灾模式，让股民一时间经历了天堂和地狱。其实中国股市一直存在着“政策市”、“消息市”和过度投机等问题，这次的过山车式的行情，我们就看到市场中始终充斥着各种信息，并深刻影响了市场中的投资者情绪，并最终促成了这一次暴涨暴跌的闹剧，而在这一过程中，网络传媒扮演了重要的角色。

根据第 37 次 CNNIC 中国互联网信息中心公布的《中国互联网络发展状况统计报告》，截至 2015 年末，中国网民数量达 6.88 亿，互联网普及率为 50.3%，相较 2014 年底上升 2.4%。

在互联网技术飞速发展的今天，网络传媒不仅影响着我们的公共生活，更显著地改变着人们的信息获取方式和投资决策行为。而股票市场也成为目前网络信息技术应用最为广泛的领域之一，如今不仅大部分金融业务都可以借助互联网完成，比如股票的开户和交易，并且网络更是已经成为股市投资信息传播的主要渠道，越来越多的投资者通过网络搜索金融信息，与他人交流投资经验与信息，如各种的股吧论坛、股评微博、金融股票 QQ 群等，信息的发布已经从少数人的特权变成了大量投资者自发的大众行为。过去，传统金融理论认为机构投资者和专

业投资分析师才是信息的提供者，因为他们拥有绝对的信息优势，从而能够主宰股票市场，但是中小投资者由于无法支付昂贵的信息获取与发布费用，他们的观点和意见往往被市场所忽略，根本无法对市场产生有效影响；然而 Web2.0 时代开启之后，投资者拥有众多的网络信息沟通工具（如：论坛、微博、blog、QQ 群等）来获取信息并发布信息，他们不仅是信息的接受者，更是信息的提供者和执行者。特别是在我国这一特殊市场，散户众多，且股民与网民高度重叠，网络传媒已经成了中小投资者获取信息和表达情绪的重要平台；不仅如此，借助以股吧论坛为代表的网络平台，中小投资者的意见和情绪在这里传播，碰撞，融合，汇聚，并最终转化为实际的投资决策，实现信息向金融市场的传递，这也必然会对金融市场产生重要影响。

例如，2015 年底在央行网站发布的一则关于深港通要来的“旧闻”，一时间被各大财经网站转载传播，投资者情绪高涨，使得 A 股和港股市场疯涨，港交所于午间紧急澄清。一场乌龙过后，在券商股带动下 A 股市场却“假戏真做”，继续大涨 4.31%，又创出反弹的新高。信息被非理性解读，并借助网络传媒快速传播扩散，汇聚和放大了非理性情绪，从而加剧了股市的非理性波动。

可见，网络媒体的信息传播正在对金融市场产生重要影响。本文就试图来研究除了因这些特殊事件的网络信息传播会放大投资者的非理性情绪从而导致股市的非理性波动之外，日常的网络信息传播，比如股吧论坛的讨论，是否也会对股票市场产生显著地影响，每天产生的海量发帖中是否蕴含着能够预测未来股市走势的信息等问题。

二、研究意义

（一）理论意义

本文研究了网络传媒对金融市场的影响，为现代网络传媒和金融市场的结合研究提供了新的思路和方法，尤其是在分析文本型数据与我国股票市场的关系方面；另一方面，本文采用文本挖掘技术，以全新的思路与角度来对市场中的投资者情绪进行研究，有较多的借鉴价值。

不仅如此，作为行为金融学的重要研究内容，投资者情绪的研究一方面可以帮助我们认识投资者情绪对金融市场的影响机制，另一方面也可以促进行为金融学理论进一步的发展。

（二）实用价值

本文试图回答来自网络股吧论坛的发帖信息能否显著影响我国的股票市场，影响程度如何，如何产生影响等问题。

对政府监管层而言，本文的研究为他们对网络舆论进行信息管理和信息引导，提前预警和防范网络传媒导致的金融市场风险提供了依据，帮助他们更好地履行监管职能，进一步完善我国股票市场的监管制度，从而促进股票市场的健康稳定发展。

对投资者而言，本文的研究能让他们认识到网络传媒信息对股票市场的作用和影响，或许可以让他们利用网络传媒信息获取短期的超额收益，同时提前对股价波动做好应对风险管理，但更重要的是帮助他们对股票市场的运行规律以及资产价格的特点有更深入的理解和认识，逐渐树立成熟而理性的投资理念。

第二节 国内外文献综述

其实，国内外很早就开始对投资者情绪进行研究和探索，但目前尚未形成一个成熟的、受到大家一致认可的投资者情绪指标，学界和业界还在不断努力探索。同时关于投资者情绪与股票市场的研究结论更是莫衷一是。本文这一部分将就当前已有关于投资者情绪的研究状况做一梳理。

一、投资者情绪的定义

投资者情绪（Investment Sentiment）源于投资者心理，是投资者在金融市场中的投资交易活动中产生的，但目前投资者情绪还没有一个统一标准的定义。

Lee、Shleifer 和 Thaler（1990）最早认为，投资者情绪是投资者对市场一种主观判断，是源于投资者心理因素的非理性产生的认知上的某种偏差；

Barberis（1998）认为投资者行为是在投资者的投资行为模型中，人们如何错误地应用贝叶斯法则或违背主观预期效用理论，形成信念和价值的过程。

Brown 和 Cliff（2004）将投资者情绪定义为对未来股票价格的总体乐观或者悲观预期。

Baker 和 Wurgler（2006）则认为投资者情绪不仅是投资者的某种预期外，还是投资者的某种投机倾向。

国内也有部分学者，如饶育蕾（2010）、王美今（2004）等人将投资者情绪定义为投资者对资产价格的一种有偏差的预期，这与国外对投资者情绪的主流定义在本质上区别不大。

虽然目前学术界对于投资者情绪还未有一个公认的定义,但前人的不同角度的定义,能帮助我们对于投资者情绪的内涵有更加全面深入的认识,并通过建立不同的情绪指标来对投资者情绪进行实证研究。

二、传统投资者情绪的度量

研究投资者情绪,寻找衡量投资者情绪的有效指标是第一步。第一个需要解决的问题是找到能准确度量投资者情绪的情绪指标。目前传统的投资者情绪指标大致分为三类:第一类是利用各种调查得到的直观反映投资者对市场预期的显性情绪指标;第二类是隐性情绪指标,主要由股票市场中的交易变量组成,它们一般被投资者情绪理论认为是受市场投资者情绪的推动而产生相应的变化,所以它们和投资者情绪有某种内在的相关性,从而被选择成为投资者情绪的代理变量;第三类是一些非经济变量,但能够影响投资者对市场的预期,常见有天气因素等。

另外,如今更多的学者是对以上单一的情绪指标构建复合的情绪指标。

(一) 单一投资者情绪指标

1、显性投资者情绪

显性情绪指标是最早的投资者情绪指标。Solt 和 Statman (1988) 先将市场中的情绪划分成看涨与看跌,然后调查投资者对未来市场走势的看法和意见,以此构建看涨和看跌指数作为情绪指标,用 BSI (Bearish Sentiment Index) 来反映投资者情绪。

美国个体投资者协会自 1987 年 1 月开始对其会员的情绪进行调查,主要是会员对未来六个月的股市预测看法--看涨、看跌或看平,并以此构建了美国个体投资者协会指数。Brown 和 Cliff (2005) 就曾经把看涨人数比例与看跌人数比例的差额作为情绪指标。

国外还有一个常见的显性情绪指标是友好指数,它是根据对每周全国的各大报刊、金融机构(如投资机构和基金公司)买卖股票的投资建议的打分,来判断他们对未来市场的乐观态度,并以加权平均的方式构建情绪指数。

情绪指标在国内也很常见,例如“央视看盘”、好淡指数、“三大报”(中国证券报、上海证券报和证券时报)、巨潮投资者信心指数、华鼎多空民意调查数据等。而在众多的显性情绪指标中影响力较大,使用最多的是“央视看盘”数据。它的构建与国外的友好指数很相似,主要根据对业内人士未来的市场观点—乐观还是悲观的调查。

饶育蕾和刘达锋（2003）最早开始将央视看盘数据作为情绪指标来研究其与未来市场收益率的关系。刘超和韩泽县（2006）后来也利用央视看盘数据进行了相关的研究。

2、隐性投资者情绪

显性投资者情绪指标虽然主要是通过对投资者的直接调查获得，但是存在样本量偏小、数据获取难度大、调查方式差异大等等问题；而隐性投资者情绪就可以很好地解决这些问题，因而此后有很多的国内外学者使用隐性投资者情绪指标来进行研究。

目前，国内外学者所使用的隐性投资者情绪指标可以归纳为以下七种：封闭式基金折价率、市场的换手率、消费者信心指数、初次公开募股的数量、新发股票的比率、IPO 首日收益率和股利溢价。

比较典型的例如：

（1）封闭式基金折价率

目前封闭式基金折价率被用得最多，学术界也有大量的相关文献。

Zweig（1973）较早通过发现封闭式基金折价率可以反映投资者情绪，并作为情绪指标。

DeLong 等（1990）也发现封闭式基金折价率可以度量噪声交易者的情绪。

Lee 等（1991）、Neal 和 Wheatley（1998）也认为这一指标可以反映个体投资者的情绪，并发现它和低比例的机构持有股、小盘股的收益变动存在正相关关系。

Elton, Gruber 和 Busse（1998）用封闭式基金折价率作为小投资者的情绪代理指标，并发现其对股票收益率能产生重要的影响。

Baker 和 Wurgler（2006）等人在他们研究中也发现，封闭式基金折价率是反映和度量投资者情绪变化的较好指标。

国内学者如顾娟（2001）、金晓斌等（2001）、张俊喜和张华（2002）、黄少安和刘达（2005）、伍燕然和韩立岩（2007）通过研究发现，国内市场的封闭式基金折价现象，用投资者情绪理论能够进行较好地解释。不仅如此，封闭式基金折价率能够较好的衡量个体投资者的情绪，因为封闭式基金的投资者大多都是个人投资者。

（2）IPO 发行量及首日收益

“热市”和“冷市”是股票市场常见的现象，而 IPO 首日收益较低就是市场

时机选择的结果 (Stigler, 1964; Ritter, 1984, 1991; Ibbotson et al., 1994; Derrien, 2005)。IPO 发行量及上市首日收益都较好地反映了投资者情绪的冷热程度, 且和投资者情绪存在正相关关系, Ljungqvist 和 Wilhelm (2003), Ljungqvist、Nanda 和 Singh (2006), Baker 和 Wurgler (2006) 等的研究都对此进行了验证。国内也有学者得出了相似的结论, 例如王春峰 (2007) 通过研究发现, 投资者情绪与 IPO 新股的发行价、上市首日报价以及抑价存在正相关关系。韩立岩和伍燕然 (2007) 研究表明投资者情绪对资产定价有重要影响, 能对国内新股 IPO 发行的折溢价现象做出较为全面合理的解释。

(3) 消费者信心指数

消费者信心指数可以很好地反映公众对于当前和未来经济的信心程度。一般当经济向好时, 公司的经营业绩也会被看好, 股票价格也会趋于上涨。

Fisher 和 Statman (2003) 研究发现密歇根大学-消费者信心指数能较好地度量 and 反映投资者情绪, 对于小盘股的收益率有一定预测作用, 不过和 S&P500 的收益率并没有显著的相关关系。Qiu 和 Welch (2006) 在研究也发现用消费者信心指数度量的投资者情绪, 能够对小市值股票的超额收益和一些新股 IPO 发行的现象做出解释, 并且其表现要优于封闭式基金折价率。Lemmon 和 Portniaguina (2006) 也发现度量投资者情绪的消费者信息指数能够在一定程度上预测小盘股和机构低持有率的股票的未来收益率。国内学者薛斐 (2005) 在国内股票市场进行类似的研究也发现, 相比封闭式基金折价率, 消费者信心指数能够更好地度量市场投资者情绪。

3、其他情绪指标代理变量

部分非经济变量和投资者情绪存在某种相关关系, 所以也被部分学者用来作为投资者情绪的代理变量, 如仪垂林和王家琪 (2005) 用每天上海天气变量度量投资者情绪, 并研究其与上证综指收益率之间的关系。但是这里情绪指标在研究中很少使用。

4、复合情绪指标

找到度量投资者情绪的有效指标是进行投资者情绪相关研究的基础, 目前虽然目前文献中有十几种各种不同的情绪度量指标, 但结果表明不存在一个完全“好”的单一情绪代理变量, 单一代理变量都不能完整地反映投资者情绪。所以, 许多学者现在正尝试通过构建复合指数来更好地度量投资者情绪。

学者们构建复合情绪指数的原理是提取各种代理指标的共同成分并作为投

投资者情绪。这一方面最突出的是 Baker 和 Wurgler (2006), 他们对封闭式基金折价率等六个单一投资者情绪指标进行主成分分析, 而选取第一主成分作为投资者情绪的代理变量。这一方法被国内外的众多学者认可和广泛使用, 如 Yu 和 Yuan (2011)、Rephael、Kandel 和 Wohl (2012) 等。

Glushkov (2006) 对封闭式基金折价率、投资者智能指数、IPO 数量和首日收益、股利收益、交易量、共同基金流入量和牛熊市比等 8 个显性和隐性单一情绪代理指标构建了一个复合指数。用这种复合指标反映投资者情绪比单一指标更加全面和真实, 并且以此得到的实证结论更加科学合理。

国内也有部分学者采用这种方法进行研究, 如黄德龙、文凤华和杨晓光 (2009), 蒋玉梅和王明照 (2010), 宋泽芳和李元 (2012) 等。

国内学者易志高和茅宁 (2008) 采用封闭式基金折价率、IPO 发行量及首日收益、消费者信心指数、A 股新增开户数、交易量构建度量国内股票市场投资者情绪的月度复合指标--CICSI 指数, 并且剔除了宏观经济因素对投资者情绪的影响。

黄德龙, 文凤华和杨晓光 (2009) 通过构建复合情绪指数来度量投资者情绪, 并验证了其与其与股市收益之间的关系, 它的成分指标主要是交易量、封闭式基金折价率和 A 股新开户比率。

三、投资者情绪与股市关系的研究

虽然投资者情绪与股市关系的研究在国内外都很多, 而且基本都是研究投资者情绪与股市收益率、股市波动率之间的关系, 但是尚未形成统一结论。

(一) 投资者情绪与股市收益率的研究

行为金融理论认为, 市场投资者情绪和股票收益率显著相关。学者们主要是对投资者情绪能否解释和预测股票收益率进行研究, 尤其是实证研究投资者情绪对未来股票收益率的预测作用, 因为这对投资者进行实际的投资决策操作有更多的应用价值。

由于不同的学者研究的时间尺度和长度不同, 研究方法也不尽相同, 最后的研究结论存在很大分歧。

国外有学者研究发现投资者情绪和未来的股票收益率不存在显著的相关关系。比如, Solt 和 Statman (1988) 通过对投资者情绪和未来一期的股市收益率进行研究, 发现二者并不显著相关。Clarke 和 Statman (1998) 将看涨指数作为投资者情绪的代理指标, 并分别研究了牛市和熊市环境下投资者情绪对未来股市

收益的影响，结果均发现投资者情绪对未来股市收益没有显著的预测能力。国内也有学者得出相似的结论。余佩琨、钟瑞军（2009）通过研究发现市场收益率对未来的个体投资者情绪有预测能力，但个体投资者情绪尚无法用来预测未来的市场收益率。

然而更多的研究都表明投资者情绪能够预测未来的市场收益。例如，Clarke 和 Statman（1998）在实证研究发现投资者情绪和当期月度的标普指数收益率呈现显著的正相关关系，但与未来的 1、6、12 个月的标普指数收益率却呈现出显著的负相关关系。Brown 和 Cliff（2005）也研究发现，在未来的 1-3 年的时间段内，投资者情绪和未来的股市收益之间依然存在显著的负向关系，他们的情绪指标主要来源于调查问卷数据。Baker 和 Wurgler（2006）用复合情绪指标深入研究市场情绪在影响不同类型股票上是否存在不同，结果发现虽然投资者情绪都与未来的股票收益率负相关，但是这种负相关影响对高波动性股票、小股票、新兴股票等类型的股票会更为显著。

国内的学者用中国市场的历史数据也进行了大量的实证研究，并有相似的发现。例如，刘仁和和陈柳钦（2005）用好淡指数度量投资者情绪进行研究发现投资者情绪确实对未来的股市收益率存在预测作用。伍燕然和韩立岩（2007）认为投资者情绪在对当期的股市收益率产生影响的同时，还能对未来期的股市收益率产生影响作用。张强，杨淑娥和杨红（2007）进一步对投资者情绪进行了细分，划分为机构和个人投资者情绪，研究发现只有机构投资者情绪能对股市收益率产生影响，而个人投资者情绪对股市收益的影响并不显著。

（二）投资者情绪与股市波动性的研究

目前已经有很多的国内外学者研究过投资者情绪对股票市场波动率的影响，因为股市波动率能直接来度量股票市场的风险，而将股市的风险和波动控制在适当的范围内是股市稳定的核心目标之一。当前研究结论都表明投资者情绪能够在一定程度上对股票市场的波动进行解释，高情绪往往会导致股市波动的加剧。

例如，Brauer（1993）的研究发现噪声交易行为可以解释大约 7% 的基金收益波动。Brown（1999）将美国个体投资者协会指数作为情绪指标进行研究，发现异常的情绪波动会导致封闭式基金收益率的大幅波动。Mehra 和 Sah（2002）进一步研究了投资者情绪对股票市场价格的影响机制，并发现当市场投资者情绪发生系统性变化时，将导致投资者风险偏好也产生改变，以致对股票市场的波动产生影响。国内学术界也有很多学者，如许承明和宋海林（2005）、林树和俞乔（2010），通过研究得出了同国外学者相似的结论。其中，林树和俞乔（2010）借助心理学实验进行研究，结果表明，当资产价格处于顶部高位时，主要是宏观

经济因素决定了市场投资者情绪的波动和资产价格的波动，市场情绪和资产价格会相互影响产生一个反馈环，从而最后也许会导致市场的泡沫甚至崩盘的发生。

四、基于网络传媒文本信息的投资者情绪研究

虽然到目前为止已经有很多从多种角度来研究投资者情绪对股票市场的影响，但是基于网络传媒文本信息角度提取投资者情绪并研究其对股票市场的影响的研究还很少。

随着互联网的不断发展，信息的传播方式和结构发生了巨大变化，网络信息中蕴含的投资者情绪对金融市场的影响越来越大，正在引发国内外金融学者的广泛关注和研究。

但是，由于网络文本信息存在难搜集，难量化的问题，所以现有研究大多使用数值型数据，运用文本信息数据的研究还很少见。

从国外研究情况来看，Wysocki（1999）用发帖量作为网络信息中投资者情绪的代理变量，选取了1998年1月到8月中发帖量最大的50家公司的股票进行研究，结果发现发帖量确实对第二天的股票交易量和异常收益率有预测能力。

Da（2011）等人采用Russell 3000指数成分股的股票代码简称在谷歌的搜索强度来量化网络信息中的投资者情绪，研究表明搜索强度能更强地预测小市场的股票收益率，对以机构投资者为主的市场如纳斯达克或者纽约证券交易所的股票收益率的预测能力较弱。

在Da（2011）等人的基础上，Joseph（2011）等人首先将股票按照搜索强度大小划分成5组，并构建了一个投资策略--“每周买入搜索强度最高组的股票”。结果发现，当搜索强度增加时，投资策略的超额收益也会随之而增加，并且搜索强度最高组的平均周收益率要比最低组的高出0.17%。

目前，国外学术界也已经开始运用网络文本型数据进行投资者情绪的研究。

Antweiler和Frank（2004）通过朴素贝叶斯算法对Yahoo! Finance中的150万条帖子分成三类--看多、看空和看平（包括噪音），在此基础上构建了看涨情绪指数，研究表明网络论坛发帖信息对股票价格的波动率有预测作用。

Das, S.R., Chen和M.Y.（2007）也对Yahoo! Finance论坛上的发帖信息进行研究，结果表明投资者情绪呈现出和发帖量很高的相关关系；同时发现投资者在论坛的发帖所反映的情绪能对未来的股票收益率进行预测，而其对股市收益率的预测能力比对个股收益率的预测能力更强。

Bollen（2011）等人利用谷歌的文本挖掘工具，抓取了推特上从2008年2月28日到12月19日之间近1000万条的公共微博，并借助谷歌的OF与GPOMS

工具分析微博文本中的情感倾向,发现该情感倾向对次日道琼斯指数的涨跌有很好的预测作用,其预测准确率高达 87.6%。

国内目前虽然已有学者开始研究网络传媒的数值型变量及其与股票市场的关系,但基于网络文本型数据的研究还非常缺乏。

饶育蕾,彭叠峰,成大超(2010)基于网络上和上市公司股票有关的新闻数量来构建媒体注意力指标,并以此研究媒体注意力与股票月度收益率之间的关系,发现媒体对上市公司二关注度越高,在接下来的一个月中,其股票的平均收益率越低。

饶育蕾,王攀(2010)用新股发行日到上市日期间内容含有新股名称的新闻的数量来构建新股的媒体关注度,研究发现新股的媒体关注度会对投资者情绪产生影响,进而能够正向影响新股的短期累积超额收益,但负向影响新股的长期累积超额收益。

张永杰,张维等人(2011)采用基于搜索引擎的文本语义挖掘算法获得了一组关于个股网络开源信息内容含量的数据,并以此研究了开源信息在资产定价方面的影响。

金雪军,祝宇,杨晓兰(2013)抓取了中国最大的股票论坛上的 580 万条发帖数据,借助文本挖掘技术,构建了看涨指数和意见趋同指数,考察了这两个指数与股票价格、成交量之间的关系。研究发现第一,看涨指数与股票收益率呈正向相关,且看涨指数对第二天的收益率具有预测作用;第二,意见趋同指数影响成交量,意见趋同程度越低,股票的交易量越大;第三,论坛发帖量对股票成交量有着显著的正向影响。

五、文献评述

过去学者关于投资者情绪所做的大量研究,已经得到了许多有价值的研究成果,但也依然存在一些的不足。

(一) 在投资者情绪指标的度量方面

合理有效地构建情绪指标是进行投资者情绪相关研究的基础。虽然学者们已经提出并运用了很多各种不同的情绪指标,但目前尚未形成一个能够较好度量或得到一致认可的投资者情绪的指标(体系),未来还有很长的路要走。

1、单一投资者情绪指标

不管是显性情绪指标或者隐性情绪指标,单一投资者情绪指标都存在代表性

不足的问题,因为其同样会受到指数本身特质的、与投资者情绪无关的因素影响。比如,Boehme et al.(2006)认为换手率不仅反映了投资者情绪,也反映了投资者异质信念。关于封闭式基金折价率,有学者如Chen等(1993)就对封闭式基金折价率和小公司股票收益之间的正向关系提出了质疑;Brown和Cliff(2004)通过研究发现封闭式基金折价率对未来的股票收益不存在预测能力。还有学者认为IPO发行量不能作为投资者情绪指标,如Choe、Masulis和Nanda(1993),Bayless和Chaplinsky(1996)。

2、复合投资者情绪指标

复合投资者情绪指标同样有代表性的问题,并非借助代理变量的选取就可以解决,并且其很容易遗漏除了情绪因素与基本因素的其他重要共同成分,而这些被遗漏的重要共同成分很可能也是投资者情绪。

不仅如此,其原始指标的选取往往具有很大的主观性,而且主要是基于国外已有的研究成果。然而国内外市场环境存在较大的差异性,再加上目前被深入研究的原始情绪指标很少,因此想要合理选取原始情绪指标,构建符合我国市场特征的复合情绪指标难度非常大。

3、基于网络文本信息的投资者情绪指标

互联网的快速发展和普及,信息在网络中传播、流通、碰撞、融合,对投资者行为和决策产生越来越大的影响,更冲击着股票等金融市场;毋庸置疑,基于网络文本信息提取投资者情绪指标是未来的研究方向和趋势。

但目前大多数学者仅使用网络文本的数值型数据进行投资者情绪的相关研究,极少基于网络文本型数据开展研究。以发帖量、搜索量为代表的数值型变量存在诸多缺陷,如维度有限,只能间接反映投资者的情绪和意见;而网络文本型数据,能克服以上缺陷,其数据来源广泛,信息内容丰富完整,因为语言的主观性能够直接反映投资者的情绪和意见,优势十分明显。

因此,这将会是未来研究的发展方向和热点。国外学者已经开始借助网络文本信息数据开展研究,而国内目前在这方面的研究还存在很多空白。

(二) 在投资者情绪与股市关系的研究方面

当前投资者情绪和股票市场关系的研究,主要是基于传统投资者情绪指标,且集中于投资者情绪和股市收益率关系的研究,但是结论多样,无法形成统一定论,尤其是在预测方面。而对于投资者情绪与股票市场波动率关系的研究,尽管大多学者都认为投资者情绪能在一定程度上解释和影响股市波动性,但在对股市

波动性的预测上的研究还很少。而基于网络文本信息的投资者情绪与股市波动性之间的研究更是存在很多空白。

并且过去对股市波动率的研究多基于 GARCH 族波动率模型，还有一部分基于随机波动率 (SV) 模型，存在参数估计困难，难以应用到较高频的数据中等问题，这也就限制了投资者情绪在股市波动性预测能力方面的研究。而网络信息更新速度极快，一般需要采用与之对应的高频数据来进行研究，本文就将采用目前已得到广泛应用的已实现波动率方法，克服了高频数据下其他波动模型的估计困难等等问题，能够较好的研究基于网络文本信息的投资者情绪与股市波动性之间的关系研究。

第三节 研究思路与文章结构

一、研究思路

本文由六章构成：

第一章为导论，其包含了本文的研究背景与意义以及文献综述。这里本文研究了国内外已有的关于投资者情绪研究的大量文献，从中汲取前人有益的研究方法和思路；同时也根据过去研究的不足和空白提出了本文的创新之处，也分析了不足点。

第二章为本文的理论基础和研究技术介绍。本章首先在行为金融理论的框架下分析投资者情绪如何对金融市场产生影响，以及网络传媒信息如何对投资者情绪产生影响；其次介绍了本文在提取投资者情绪指标中所用到的主要的文本挖掘技术。

第三章是本文的研究设计和实证模型构建，包括基于网络文本的情绪指标构建，变量的选取以及研究该情绪指标分别影响上证指数收益率和波动率的模型构建。

第四章主要介绍本文的样本，包括样本的来源和获取，统计性描述以及变量的相关性分析。

第五章为本文的实证部分，实证研究基于股吧论坛的投资者情绪对市场收益率和波动率的影响，并与过去传统的情绪指标进行对比分析。

第六章则是对本文研究成果的总结。

二、文章结构框架

下图 1.1 是本文的框架结构图：

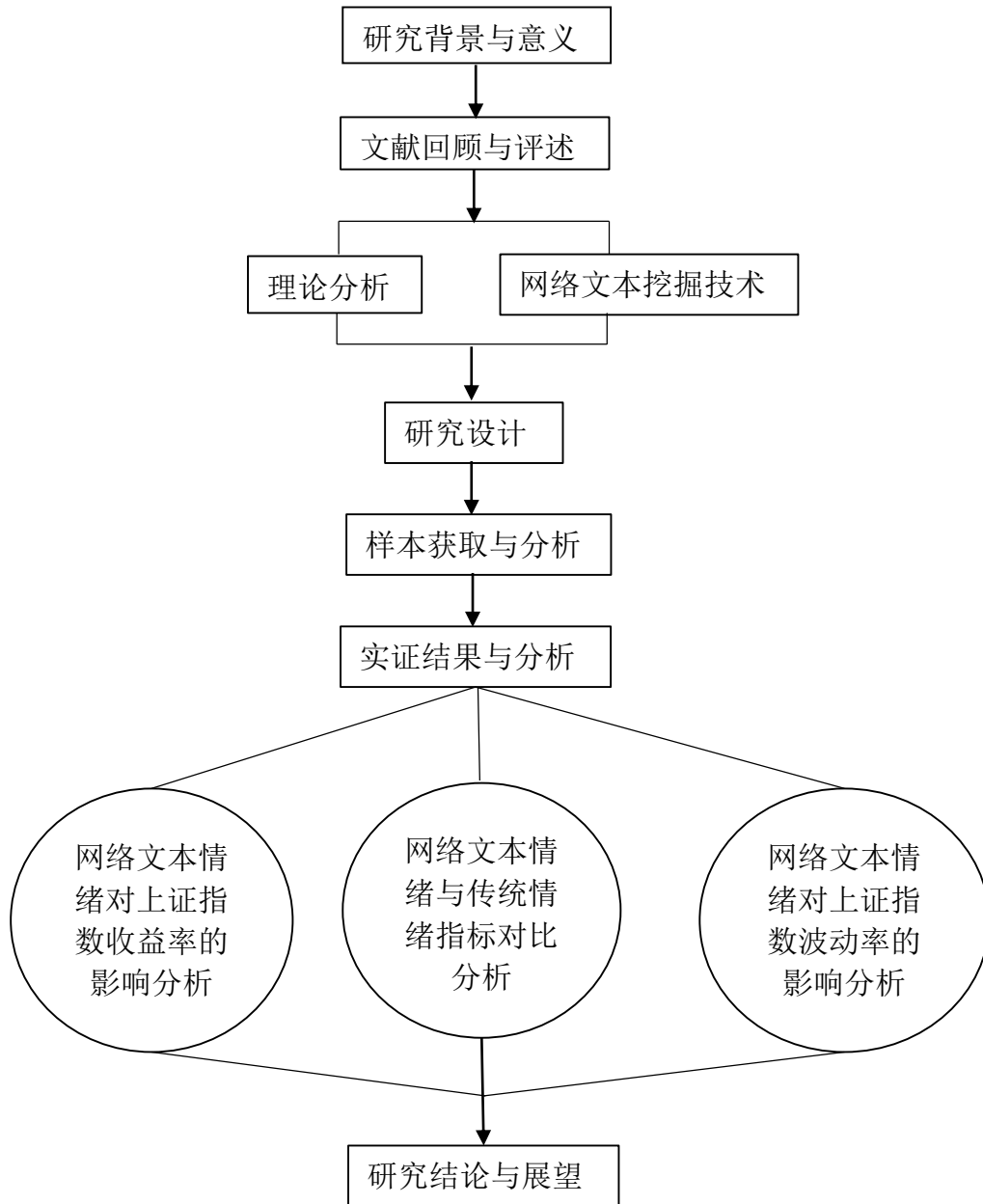


图 1.1 文章框架结构图

第四节 本文创新之处与不足之处

一、本文的创新点

(1) 本文创新性的尝试跳出金融交易市场，从网络媒体的文本信息中去分析市场投资者情绪，不仅是数值型量化指标（如发帖量），更借助文本挖掘技术，从网络文本中提取有效信息并量化，构建情绪指标，以全新的视角和方法深入研究对股票市场收益率和波动率的影响。

(2) 本文对交易日做了重新的界定，并对每个交易日时间段做了更细致的划分，使本文情绪指标更加准确地度量该时间段或交易日的投资者情绪。同时，本文专门研究了每个“交易日”开盘前的非交易时段的股吧论坛发帖信息对股票市场收益和波动的影响。

(3) 在股市波动率的研究上，本文基于股市交易的高频数据采用了已实现波动率这一非参数估计方法来度量股市波动率，更好地反映股市日内波动变化，更好地研究其与基于快速变化的网络文本信息的情绪指标的关系。

(4) 本文也借着本次股市大涨、大跌的行情变化，分别在不同的市场环境下--牛市环境下和熊市环境下，深入分析投资者情绪在其中扮演的角色和作用，研究其中的差异性。

(5) 本文还将基于网络传媒构建的情绪指标与常见的传统投资者情绪指标（包括显性情绪指标和隐性情绪指标）进行对比，更加全面地分析和认识以股吧论坛为代表的网络传媒信息中蕴含的投资者情绪在解释甚至预测市场收益率和波动率方面的优势和实用价值。

二、本文不足点

由于研究水平和技术手段的限制，本文仍然存在不足之处：

(1) 本文选取的样本中的熊市子样本属于股灾暴跌的极端行情，不具有普遍性，所以本文的研究结论的适用性受到限制；未来可以选取更长时间段的样本来研究，分析在各类市场环境--牛市，熊市，震荡，极端行情等，基于网络文本挖掘提取的投资者情绪与证券市场之间的相互影响。

(2) 由于条件的约束，本文只抓取了东方财富股吧论坛上的发帖信息作为研究样本。今后可以进一步扩展样本信息来源，如微博、博客和社交网站等更多的社会化网络传媒，以期更深入，更全面的研究。

(3) 因为当前对数据处理的能力有限，本文的研究仅限于主贴信息，流失

了回复贴等部分网络文本信息，以后可以考虑把回复贴等也纳入研究样本范围，构建更加全面和丰富的投资者情绪指标。

第二章 理论基础和研究技术

第一节 本文的理论基础

一、投资者情绪与金融市场

投资者情绪的研究关系到市场是否有效、资产的价格是否充分且无偏地反映所有信息以及投资者是否理性等问题。投资者情绪对金融市场影响的研究主要是基于行为金融学这一重要理论基础，本文的研究也是如此。

(一) 行为金融学

如果现今的金融市场是有效的，那么市场将一直处在均衡状态，对市场的各种研究都将因不能获取超额收益回报而变得没有价值。但是，自上世纪 80 年代建立股票市场以来，虽然市场在发展和完善，但一直存在很多以有效市场理论为代表的传统金融理论所无法解释的金融异象，如过度反应、羊群效应、过度自信、股票溢价、日历效应等。而行为金融学彻底打破了传统金融理论的研究假设，融合了行为科学以及心理学等相关学科的成果，强调了个体“人”在市场中的突出影响，较好地解释了上述的金融异象。

行为金融学理论认为投资者非完全理性，以此为基础来分析投资者的心理、行为和情绪对投资者的行为和决策的影响，从而对资产价格和整个金融市场产生影响。行为金融学的研究主要分为三个层次：

(1) 个体行为，从微观个体角度研究和分析大量的个体投资者的行为偏差在中长期可能给金融市场带来的影响；

(2) 群体行为，行为金融学认为市场中的“人”都是互相影响，互相依存的，没有“人”是独立存在的，也因而会在一定程度上作用和影响整个金融市场。这方面的研究主要探索在众多投资者互相影响的情况下的群体行为，并且他们又是怎样形成连贯的市场行为的；

(3) 有限套利和非有效性市场，主要是挖掘市场中的异常现象，并对其产生的深层次原因进行分析。大量的研究发现在一个理性交易者和非理性交易者相互影响的市场中，非理性交易者能够长期对资产价格产生实质性的影响。并且在现实中传统金融理论所依赖的套利因存在很多限制而经常难以发生，从而使得金融市场对信息的反应会发生系统性的偏离。

行为金融理论的研究成果，例如市场非有效，投资者非理性，尤其是中小投资者的非理性，为本文研究投资者情绪对股票市场证券价格的影响提供了必要的

理论基础。

(二) 噪声交易理论

噪声交易理论属于行为金融学主要理论中的一个。在金融领域的概念中,和资产价值有关的信号称为信息,而和信息相对的称为噪声,所以噪声和资产的价值没有关系。而那些根据噪声进行投资和交易的投资者就是噪声交易者,他们一般非理性,信息获取和分析的能力比较受限制。市场中的噪声交易者经常会把噪声误以为信息,这些噪声也包括有些人故意制造的虚假信息,由于情绪的存在,噪声交易者对于资产价格的预期往往会有系统性的偏差,

DeLong, Shleifer, Summers, and Waldmann (1990) 在 1990 年提出了噪声交易模型 (DSSW)。DSSW 模型认为,噪声交易者会因对资产价格预期错误而导致系统性的偏差,这样的行为会产生噪声交易风险,这一风险会让理性交易者的套利活动的风险高于系统性风险,因而使得套利行为受到限制。所以,噪声交易理论主要阐述了市场中的噪声交易者和噪声交易风险是如何限制套利活动从而让噪声交易者在市场中得以生存的。噪声交易理论认为,噪声交易者对资产价格预期会产生系统性偏差原因主要是投资者情绪的存在,因此,噪声交易理论为后来的投资者情绪的研究奠定了重要的理论基础,过去大量关于投资者情绪与金融市场关系的研究都是在这一理论基础之上进行的,本文的研究也是如此。该模型主要有以下几个关键假设:

1、金融市场中的交易者可以分为两种,第一种是理性交易者,用 i 来表示,其比例为 $1-\mu$;第二种是噪声交易者,用 n 表示,其比例为 μ 。

2、市场同时存在两类支付同样红利 r 的资产:一类是无风险资产 s ,完全的供给弹性,固定价格为 1;另一类是风险资产 u ,在 t 时期 u 资产的价格为 P_t 。

3、模型假定只有两期,并且投资人的决策只有一种,即对无风险资产与风险资产的组合进行合理选择。

4、投资人都是风险厌恶者,决策目标都为最大化期望效用,关于财富 w 都遵循以下这个效用函数:

$$U = -e^{-(2\gamma)w} \quad (2.1)$$

其中, γ 表示绝对风险厌恶程度的系数。

噪声交易者由于各种因素会错误的认识风险资产的预期价格。其偏离程度是一个独立分布的正态随机变量:

$$\rho_t \sim N(\rho^*, \sigma_\rho^2) \quad (2.2)$$

其中 ρ^* 为错误认识偏离的均值。

式 (2.2) 正好描述了投资者情绪。该式中 ρ_t 可被看为投资者情绪，而 ρ^* 就是前文投资者情绪定义中的投资者预期的系统性偏差程度。所以，该噪声交易模型很好地阐述了投资者情绪和股票市场收益率的关系，以及如何影响股票市场。

在上述模型假设下，当风险资产的收益为正态分布时，期望效用函数等价为：

$$E(U) = \bar{w} - \gamma\sigma_w^2 \quad (2.3)$$

这里 \bar{w} 表示最终的预期财富， σ_w^2 表示一期以前财富的方差。理性投资者选择数量 λ_t^i 的风险资产 u 使得 $E(U)$ 最大化：

$$E(U) = \bar{w} - \gamma\sigma_w^2 = c_0 + \lambda_t^i(r + {}_t p_{t+1} - p_t(1+r)) - \gamma(\lambda_t^i)^2 \{ {}_t \sigma_{p_{t+1}}^2 \} \quad (2.4)$$

这里 c_0 为第一期的劳动收入，变量前的下表表示预期形成的时间，本文定义 t 时期预期 p_{t+1} 的方差为：

$${}_t \sigma_{p_{t+1}}^2 = E_t \{ (p_{t+1} - E_t(p_{t+1}))^2 \} \quad (2.5)$$

类似的，噪声交易者选择数量 λ_t^n 的风险资产 u 使得 $E(U)$ 最大化：

$$E(U) = \bar{w} - \gamma\sigma_w^2 = c_0 + \lambda_t^n(r + {}_t p_{t+1} - p_t(1+r)) - \gamma(\lambda_t^n)^2 \{ {}_t \sigma_{p_{t+1}}^2 \} + \lambda_t^n \{ \rho_t \} \quad (2.6)$$

由 (2.4) 和 (2.6) 可得：

$$\lambda_t^i = \frac{r + {}_t p_{t+1} - p_t(1+r)}{2\gamma \{ {}_t \sigma_{p_{t+1}}^2 \}} \quad (2.7)$$

$$\lambda_t^n = \frac{r + {}_t p_{t+1} - p_t(1+r)}{2\gamma \{ {}_t \sigma_{p_{t+1}}^2 \}} \quad (2.8)$$

假设风险资产供给固定为 1，当供求平衡时，有：

$$(1-\mu)\lambda_t^i + \mu\lambda_t^n = 1 \quad (2.9)$$

结合 (2.7) 和 (2.8) 式，可得：

$$p_t = \frac{r + {}_t p_{t+1} - 2\gamma \{ {}_t \sigma_{p_{t+1}}^2 \} + \mu\rho_t}{1+r} \quad (2.10)$$

上式表明，投资者情绪导致投资者对资产价格的预期偏离，进而影响均衡的资产价格。并且，情绪越高，投资者的价格预期偏离程度就越大，促使价格趋于上升。

为了研究投资者情绪对资产价格的长期影响作用，这里将噪声交易者的相对收益纳入模型中：

$$p_t = 1 + \frac{\mu(\rho_t - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{2\gamma({}_t\sigma_{p_{t+1}}^2)}{r} \quad (2.11)$$

上式中仅第二项为变量，则取其方差，得到：

$${}_t\sigma_{p_{t+1}}^2 = \sigma_{p_{t+1}}^2 = \frac{\mu^2\sigma_\rho^2}{(1+r)^2} \quad (2.12)$$

代入 (2.11) 式得到风险资产 u 的定价公式：

$$p_t = 1 + \frac{\mu(\rho_t - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{2\gamma}{r} \left(\frac{\mu^2\sigma_\rho^2}{(1+r)^2} \right) \quad (2.13)$$

从上式可以看出，投资者情绪的存在使得投资者对资产价格的预期产生偏离，从而对资产均衡价格产生影响。具体来说，当投资者情绪越高涨时，投资者对资产价格预期的错误认识程度也就越大，此时会推动资产价格的上升。

为了在长期进一步分析噪声交易者情绪的存在对资产价格的影响，考虑噪声交易者的相对收益：

$${}_t(\Delta R_{n-i}) = \rho_t - \frac{(1+r)^2(\rho_t)^2}{(2\gamma)\mu\sigma_\rho^2} \quad (2.14)$$

从而

$$E(\Delta R_{n-i}) = \rho^* - \frac{(1+r)^2(\rho^*)^2 + (1+r)\sigma_\rho^2}{(2\gamma)\mu\sigma_\rho^2} \quad (2.15)$$

μ 越大，期望的相对收益就越大。从式 (2.13) 和 (2.15) 可以看出，噪声交易者的情绪导致他们的投资信念不确定，这既会降低风险资产的价格，也会提升自身的相对收益，从而为自己提供了生存的空间。从这里我们会发现，噪声交易者的情绪具有两面性，它一方面使得噪声交易者能在市场中生存下来，另一方面因为对市场造成波动而威胁到了自己的生存。所以，噪声交易者的数量在现实中经常会发生变动。而 DSSW (1990) 又对噪声交易者数量变化的趋势进行研究，发现噪声交易者在市场中要么主导市场，要么完全消失。在现实的市场中，因为套利受限等原因，市场中会存在大量的噪声交易者，因而市场的投资者情绪也会经常发生剧烈波动，严重影响这金融市场的稳定运行。

二、网络传媒与投资者情绪

根据羊群效应理论我们可以发现,现在我国的金融市场尤其是股票市场有严重的信息流羊群效应问题。这也帮助我们理论上在市场信息的传播途径上发现投资者情绪。因为在羊群效应理论成立的前提下,通过网络传媒投资者情绪可以得到快速传播,投资者情绪高涨的股票会吸引更多投资者跟进买入,而情绪低落的股票则会被投资者不断抛售。

在互联网技术被深入普及和应用的今天,网络平台成了信息的主要传播媒介,有越来越多的投资者通过网络进行投资行为,并在各种网络传媒如股票论坛上交换投资的经验与心得。以股吧论坛为代表的网络传媒使得投资者之间的交流越来越密切,其上的帖子在反映投资者的投资心理与行为的同时,也将影响其他市场参与者的投资决策和行为。

股票论坛主要有三个特点:普遍性、专业性和混乱性。有研究显示,在股票市场中访问过股吧论坛的投资者的比例高达 82.1%,如果去掉学生,也就是仅考虑在工作的普通投资者,也有 77.7%的人曾经访问过股吧论坛,同时参与投资的群体比未参与投资的群体有更高的概率访问股吧论坛甚至发帖参与讨论,这说明有越来越多的投资者参与股吧论坛,且该群体的专业性也在不断提高,已经具备一定的影响力。其实,在股吧论坛发帖既是投资者的行为,也会影响投资者或潜在投资者的决策。虽然股吧论坛也存在混乱性的特点,因为没有对股吧论坛的发帖内容进行设定,从而也会存在一些与主题无关的信息,如广告;但是我们也无法否认股吧论坛里面确实存在对股票市场有价值的重要信息。

股票论坛中的信息主要可以概括为以下几类:公开信息,这种信息多是真实的,如上市公司的基本信息、公告等;小道消息,主要是一些自认为知情的人所发布的传闻,但很可能是并不真实;个人观点信息,网民在网上所发表关于股票市场的观点和态度;虚假信息,通过有意发布此类不实信息影响其他投资者的投资决策,从而帮助市场的主力来操控市场;情绪表达信息,当投资者在投资中获利获利或亏损时,经常会发表此类有明显情感倾向的帖子来宣泄心中的情绪。其实以上的任一种信息都可能影响市场中投资者的心理和投资决策,即便是虚假信息,噪声交易者也会因为无法辨识而受到影响,特别是在我国这个散户众多、投机严重、非理性的市场中,股吧论坛信息已经变成对投资者心理和投资决策产生影响的重要因素。

从现实情况来看,网络信息对股票市场的作用在不断增强,由于网络信息的传播导致市场波动的现象越来越多,市场参与者也对其越来越关注和重视了。

第二节 本文的研究技术

本研究的一个难点在于如何将海量的非结构化的文本信息进行量化。本文主要基于网络文本挖掘技术，包括网络信息文本采集、中文分词、词频统计、特征关键词选取、文本的向量空间模型表示和基于算法的文本分类等技术处理，并最终构建投资者情绪指标。

构建投资者情绪指标，所需要解决的一个关键性问题是对于非结构化的海量网络文本进行情感分类——看涨、看跌、中立或无关。

以下是网络文本挖掘的核心框架图：

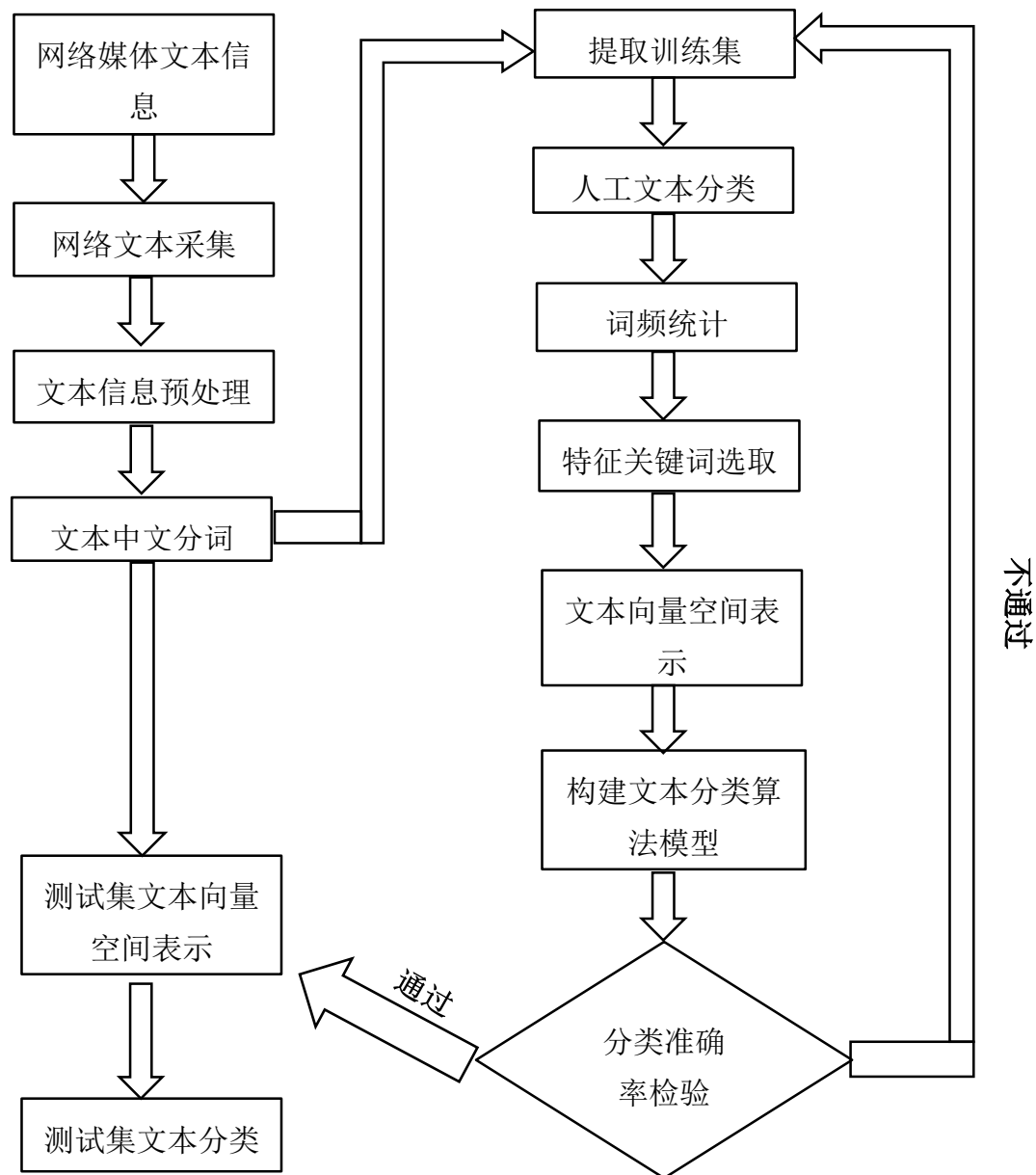


图 2.1 网络文本挖掘技术的核心框架图

一、网络文本采集

网络文本信息主要存在于动态网页中，而本文研究的样本则选取来自股吧论坛的文本信息。本文拟采用火车头采集器抓取股吧论坛中的帖子文本信息。火车头是一个供各大主流文章系统，论坛系统等使用的多线程内容采集发布程序，它可以根据目标网页的信息结构灵活设计信息采集规则和任务，根据用户需求自由批量采集股吧论坛动态网页中的目标字段。采集之后，软件可以自动将采集到的网络文本信息发布到网页、数据库或生成文件（如.csv，.txt 等），保存到本地。

二、文本信息预处理

通过火车头抓取到的文本信息，有很多是与投资者情绪无关的垃圾信息或无效信息，如广告，新闻等；同时，帖子文本中也存在很多无关的字段，如表情、特殊字符和标点等。这些无效帖子或文字虽然不多，但是如果不剔除，将会对之后情感分类的准确率产生很大影响。因此，本文拟利用 excel 软件中的“排序与筛选”、“查找与替换”功能进行手工剔除操作。

表 2.1 无效帖子示例

帖子类别	帖子内容
新闻	新闻:长江经济驱动丝绸之路的核心是武汉不是西安 丝绸之路的前提是修路，那么
广告	盈赫资本金融课堂：给你的财务健康做体检
表情	又踏空了[怒][怒][怒][怒][怒][怒]
特殊字符	\
无关字段	热烈欢迎美军继续巡航南海 12 海里，只有你们来，股市才会暴涨，欢迎你们来。

来源：整理自东方财富网上证指数吧

三、中文分词

中文分词是中文信息处理的基础与关键。分词，就是根据使用时的含义，对句子中的词语进行切分。因为中文句子中，词语和词语之间没有分隔符将其分开，所以对中文信息处理前需要先分词。

中文分词系统很多，但本文拟采用目前世界上最好的汉语词法分析器

ICTCLAS2015。这是一款由中国科学院张华平教授团队开发的一款中文分词系统，主要功能包括中文分词；词性标注；命名实体识别；用户词典功能；支持 GBK 编码、UTF8 编码、BIG5 编码。新增微博分词、新词发现与关键词提取等功能。

表 2.2 中文分词结果示例

原帖	分词结果
预披露这么多 待完的节奏	预 披露 这 么 多 待 完 的 节 奏
老股盘整，以跌为主。新股由于筹码大部分在它们手里，将爆炒！	老 股 盘 整 ， 以 跌 为 主 。 新 股 由 于 筹 码 大 部 分 在 它 们 手 里 ， 将 爆 炒 ！
咔嚓 2000 点断了	咔嚓 2000 点 断 了

四、特征词选取

分完词之后，由于包含的词语非常多，如果将所有词语都纳入特征空间，会造成维数灾难，文本分类的效率和精度将大大降低，所以一般需要先利用特征词选取方法进行降维。在文本分类过程中，特征词选取的难度最大，它主要是基于一定的标准过滤其中信息贡献量较小的词，只保留信息贡献量较大的词，从而达到降维的目的。

它主要是基于一定的算法模型，从中提取出对文本分类比较重要的关键词，由这些关键词来取代整一条文本信息，从而简化信息并为下一步文本的向量空间的数学表达做好准备。目前，业界和学术界有很多特征词的选取方法和算法，如 TF-IDF 方法，词频方法，互信息方法等。本文拟采用现在比较流行的信息增益特征选择方法。信息增益特征选择，主要依据某个特征词给整个分类系统的信息贡献量，信息贡献量越大，这一特征词就越重要。在实践中，信息增益方法有很好的分类表现和效果。

其公式如下：

$$\begin{aligned}
 IG(T) &= H(C) - H(C|T) \\
 &= -\sum_{i=1}^n P(C_i) \log_2 P(C_i) + P(t) \sum_{i=1}^n P(C_i|t) \log_2 P(C_i|t) + P(\bar{t}) \sum_{i=1}^n P(C_i|\bar{t}) \log_2 P(C_i|\bar{t}) \quad (2.16)
 \end{aligned}$$

其中， $P(C_i)$ 表示类别 C_i 出现的概率， $P(t)$ 表示特征 T 出现的概率， $P(C_i|t)$ 表示出现特征 T 时，类别 C_i 出现的概率， $P(C_i|\bar{t})$ 表示不出现特征 T 时，类别 C_i 出现的概率。

当对每个词计算信息增益之后，就可以根据信息增益的大小对词进行排序，然后从中选取信息增益较大的 N 个词作为特征词。

五、向量空间表示

选取特征词之后，就可以借助向量空间模型（VSM: Vector Space Model）对文本进行数学表达，每一个特征词是一个维度，每一条文本信息都可以简化为一条文本向量，继而通过分类算法进行相似度运算。

其表示原理如下：

设 D 为一个包含 m 个文档的文档集合， D_i 为第 i 个文档的特征向量，则有

$$D = \{D_1, D_2, \dots, D_m\}, D_i = (d_{i1}, d_{i2}, \dots, d_{in}), i = 1, 2, \dots, m$$

其中 $d_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ 为文档 D_i 中第 j 个特征词条 t_j 的权值，它一般被定义为 t_j 在 D_i 中出现的频率 t_{ij} 的函数，

关于设置权值，主要有两种方式：

(1) 布尔值， $d_{ij} = 0$ 或 $1, 0$ 表示文档 D_i 中没有第 j 个特征词条， 1 表示文档 D_i 中含有第 j 个特征词条。

(2) 采用 TFIDF 函数，即 $d_{ij} = t_{ij} * \log(N/n_j)$ 其中， N 是文档数据库中文档总数， n_j 是文档数据库含有词条 t_j 的文档数目。

而本文拟采用第二种方式来计算每个特征词的权值，使分类结果更加精确。

向量空间模型把文本信息变成了结构化数据而能够被计算机识别和处理，从而文本的相似性判断就等价于向量的相似性判断。

六、文本分类算法

得到训练集文本的向量空间表达之后，就可以选择一定的算法，来进行相似度的运算，建立分类模型，也即机器学习。

现在对文本信息处理可以选择很多文本分类算法，比如 KNN 算法、决策树、贝叶斯算法等。本文将采用支持向量机算法来处理样本数据，因为该算法被很多相关研究和实践证明能较好的对中文文本信息进行分类，而且准确率较高。

支持向量机（Support Vector Machines, SVM）是 V.Vapnik 等人开发的一种机器学习技术，其基本思想是，对于一个给定的具有有限数量训练样本的学习任务，如何在准确性（对于给定训练集）和机器容量（机器可无错误地学习任意训练集的能力）之间进行折衷，以得到最佳的推广性能。

经过这一系列的步骤之后，即可得到文本的分类模型。通过这一模型，只要输入未分类的测试集文本的空间向量，就可以实现文本的自动分类。本文这一系列过程都主要通过 JAVA 语言编程实现。

第三章 研究设计与模型构建

本文主要通过选取某一具有代表性的股吧论坛，提取股民发表的论坛帖子文本，通过文本挖掘和分类技术，构建投资者情绪指标。基于该投资者情绪指标，研究股吧论坛中所反映的投资者情绪对我国股票市场的影响，分别为投资者情绪对市场收益率的影响以及投资者情绪对市场波动率的影响。

第一节 指标构建与控制变量选取

一、基于网络文本的投资者情绪指标构建

网络股吧论坛为中小投资者提供了相互交流的平台，通过这个平台，他们的观点和意见得以汇聚、融合和扩散。股吧论坛投资者每天海量的发帖信息，纷繁复杂，这些信息当然也包括带有欺骗性的虚假信息，也包括没有实质性内容仅仅表达情绪的信息，因为股吧论坛有匿名、草根、互动等特点。但无论如何，投资者在其中发布的帖子信息一定程度上体现了对某类股票或整个市场未来发展的情绪倾向。而这种情绪倾向会通过网络传媒的信息传播，对众多市场的个体投资者的投资决策行为产生影响，并进而影响未来的股票市场走势。

在这一部分，本文将通过前面介绍的文本挖掘技术，对所提取的样本中的帖子进行情感分类，并构建看涨指数来度量股吧论坛中的投资者情绪。看涨指数反映了股吧论坛中投资者对未来市场走势整体的情感倾向，看涨指数越高说明大部分投资者预期市场指数会上扬。

根据投资者发帖的不同情感倾向和意见类型，我们主要把帖子分成三类：“看涨类”、“中立/噪音类”、“看跌类”。

表 3.1 投资者在股吧发布不同类型帖子

类别	帖子内容
看涨类	“大盘强势上行，个股精彩纷呈，牛市格局再现!”
	“抓紧时间上车啊”
	“中国万点大牛市要来了，赶快准备进场”
中立/噪音类	“听说要停牌。请问真的吗？”
	“主力成功高抛低吸。”
	“这是在震荡上行，还是在等待大跌？”
看跌类	“再见了，我再也没有耐心等你涨了，亏本卖”
	“今天情况不妙，走为上”
	“今日看谁跑的快”

来源：整理自东方财富网上证指数吧

借鉴和参考安特魏莱尔和弗兰克（2004）的方法，构建以下投资者情绪指标——看涨指数 $Sent_t$ ：

$$Sent_t = \frac{M_t^{buy} - M_t^{sell}}{M_t^{buy} + M_t^{sell}} * \ln M_t \quad (3.1)$$

其中， M_t^{buy} 表示某一时段中发表看涨意见的帖子数量， M_t^{sell} 表示某一时段中发表看跌意见的帖子数量， M_t 表示某一时间段中发表看涨和看跌明显意见的帖子数量， $M_t = M_t^{buy} + M_t^{sell}$ 。

该指标综合考虑了带有意见发帖数量的影响以及看涨和看跌相对情绪的变化。从该定义式我们可以大致判断，看涨指数越高，表明越多的人对市场乐观看涨，也即意味着有越多的人愿意买入或持有股票，市场指数或股票价格就越趋向于上涨。

二、交易日时间段界定

由于我国的股票市场的交易时段为上午 9:30-11:30，下午 13:00-15:00，所以本文将交易日时间定义为每日的 15:00 至次日的 15:00，因为一旦当日 15:00 收盘后，则发帖信息将主要作用于下一个交易日的市场交易，直至次日的 15:00 交易结束。每日的时间段划分为 A,B,C,D 四段：



图 3.1 交易日界定与时间段划分

三、模型控制变量的选取

股市作为反映经济运行的指示器，除了受到投资者情绪的影响之外，必然还会受到宏观经济的影响。本文将以 Stephen A. Ross (1976) 的多因素模型为基础建立实证回归模型，参考 Chen, Roll 和 Ross(1986)与 Elton, Gruber 和 Mei(1994) 利用多因素方法进行研究选取的宏观经济控制变量，根据经济因素反映宏观经济的能力，主要拟选取以下 4 个指标作为宏观控制变量。

(1) 规模以上工业增加值环比增速 g (月度值)

工业增加值是工业企业生产过程中新增加的价值，其增速在很大程度上反映了国民经济产出的变化；一般情况下，工业增加值增速越快，经济越向好，有利于股市的上扬。本文拟选取 wind 数据库中的规模以上工业增加值增速数据作为变量 g 的数据来源。(单位：%)

(2) 通货膨胀率 h (月度值)

通货膨胀率往往会加剧市场未来的不确定性，投资者将会要求更高的风险补偿，从而对股票市场产生影响。本文拟选取 wind 数据库中的居民消费价格指数 CPI 的环比数据作为变量 h 的数据来源。(单位：%)

以上两个宏观控制变量均为月度值，为了与其他变量的日度数据匹配，本文将采用学术论文常用的方式，该月内每日的规模以上工业增加值环比增速和通货膨胀率均等于该月的月度值。

(3) 利差 $dshibor$

利率反映了市场所要求的整体收益回报率，而利差则反映了市场收益率的变化；一般，利率上升会导致股市下挫，而利率下降则会促进股市上涨。本文拟选取 wind 数据库中的 7 天银行间同业拆借利率 $Shibor$ 数据作为变量 $dshibor$ 的数据来源。(单位：%)

$$dshibor_t = shibor_t - shibor_{t-1} \quad (3.2)$$

(4) 人民币汇率变化 $dchn$

汇率的变化反映了一国对外经济和国际收支状况,其也会对一国经济和股票市场产生重要影响。本文拟选取 wind 数据库中的美元兑人民币汇率 chn 数据作为变量 $dchn$ 的数据来源。

$$dchn_t = chn_t - chn_{t-1} \quad (3.3)$$

(5) 周末效应虚拟变量 I

因为本文计算周末或假期之后首个交易日的看涨指数时,包括周末和假期期间的发帖,为了分析周末或假期是否含有影响甚至预测市场收益率和波动率的重要信息,所以本文拟引入一个周末效应虚拟变量 I 。

$$I_t = \begin{cases} 1, & \text{含周末或假期} \\ 0, & \text{其他} \end{cases} \quad (3.4)$$

第二节 投资者情绪指标影响股市收益率的模型构建

一、投资者情绪对市场指数收益率的影响模型

这里,本文拟采用同期的时间序列数据进行回归,分析投资者情绪与上证指数收益率的相关程度和对其的解释能力。

以每日上证指数收盘价计算的对数收益率 r_t (其中, $r_t = \ln P_t - \ln P_{t-1}$) 作为被解释变量,以 $g_t, h_t, dr_t, dchn_t$ 四个宏观变量作为控制变量,以 $Sent_t, I_t$ 为研究变量,其中 $Sent_t$ 看涨指数基于每日从上一交易日的 15:00 到当天交易日的 15:00 的帖子信息进行统计和计算而得,也即选取上面时间段中的 A,B,C,D 四段;因为前一日收盘后的帖子信息和表达的情绪将对下一日的交易产生影响。

具体的回归模型构建如下:

$$r_t = \alpha + \sum_{i=1}^d T_i * r_{t-i} + \beta * Sent_t + \theta * I_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (3.5)$$

其中, r_{t-i} 表示 r_t 的滞后项, d 为滞后阶数,通过 AIC 和 BIC 信息准则来确定。

根据回归结果可以分析,如果 R^2 越大,则说明回归模型拟合程度越高,模型的解释能力越强; $Sent_t$ 看涨指数的系数 β 显著性程度越高,值越大,则说明该情绪指标对上证指数收益率的影响越大,符号为正说明看涨情绪有助于推动股市的上涨,反之则说明看涨情绪会抑制股市的上涨;若虚拟变量 I_t 的系数 θ 显著,则说明存在显著的周末效应,反之则不存在。

二、投资者情绪与市场指数收益率的 VAR 模型

在实际的市场交易活动中，投资者情绪不仅会影响市场当期的收益率，也会影响未来跨期的市场收益率；并且，市场收益率的变化亦会对当前及未来的投资者情绪产生影响。所以，本文拟通过建立两元变量的 VAR 自回归模型来分析市场收益率与投资者情绪之间的相互影响关系：

$$\begin{cases} Sent_t = c_0 + \sum_{j=1}^p W_{1j} * r_{t-j} + \sum_{j=1}^p V_{1j} * Sent_{t-j} + \delta_1 * g_t + \omega_1 * h_t + \lambda_1 * dchn_t + \varphi_1 * dshibor_t + \varepsilon_{1t}, \varepsilon_{1t} \sim N(0, \sigma_{1t}^2) \\ r_t = d_0 + \sum_{j=1}^p W_{2j} * r_{t-j} + \sum_{j=1}^p V_{2j} * Sent_{t-j} + \delta_2 * g_t + \omega_2 * h_t + \lambda_2 * dchn_t + \varphi_2 * dshibor_t + \varepsilon_{2t}, \varepsilon_{2t} \sim N(0, \sigma_{2t}^2) \end{cases} \quad (3.6)$$

其中， p 是根据 AIC 和 BIC 信息准则确定的滞后阶数。

根据回归结果可以分析，若系数 W_{1j} 显著，则说明市场收益率的变化可以影响未来的投资者情绪变化，有助于预测投资者情绪；若系数 V_{2j} 显著，则说明投资者情绪指标的变化能够影响甚至预测未来市场的收益率。

三、开盘前投资者情绪对上证指数收益率的预测影响模型

这里，本文拟选取 A 时段这个隔夜休市期间的发帖信息，并以此计算和提取投资者情绪指标；因为自上一日收盘后至当日开盘前（即 A 时段），投资者会发表对当日市场走势的观点和看法，这其中很可能包含了能够预测当日市场走势的信息。

首先，研究开盘前投资者情绪对市场指数收益率的预测作用

具体模型构建如下：

$$r_t = \alpha + \beta * Sent_{A_t} + \sum_{i=1}^d T_i * r_{t-i} + \theta * I_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (3.7)$$

其中， $Sent_{A_t}$ 表示 A 时段休市期间的看涨情绪指数。

从而根据回归结果分析休市期间的发帖信息是否含有能够预测市场未来走势的信息；是否存在显著的周末效应。

其次，研究开盘前投资者情绪对市场指数收益率预测作用的时效性分析

一般而言，随着交易的进行，会有新的帖子不断产生，交易前的帖子里的信息对后续交易的影响价值将会越来越小，所以这里本文拟研究交易前的投资者情

绪对开盘后交易的影响如何随时间变化，在开盘后多少时间内比较有效。

模型同上，但是因变量 r_t 变为隔夜收益率 r_t^O ，开盘后 5min 的收益率 $r_t^{5\min}$ ，开盘后 15min 的收益率 $r_t^{15\min}$ ，开盘后 30min 的收益率 $r_t^{30\min}$ ，开盘后 1 小时的收益率 $r_t^{60\min}$ ，开盘后 2 小时的收益率 $r_t^{120\min}$ 分别进行回归，然后观察和分析回归系数的数值大小和显著性的变化。

第三节 投资者情绪对股市波动性影响的模型构建

投资者情绪包含理性情绪和非理性情绪，而网络股吧论坛基本都是散户宣泄情绪的场所以，更多体现为非理性情绪。一般而言，非理性情绪越强，对股市的波动性影响将越剧烈。

在股票市场中，信息会持续地影响股票市场的价格运动过程，那么数据的离散采集就必然会产生信息不同程度丢失的问题，传统的 GARCH(1, 1)模型就存在这一问题。

一、改进已实现波动率的计算

本文拟采用基于非参数方法估计的已实现波动率 (RV, Realized Volatility) 来进行分析。已实现波动率最早由 Anderson 和 Bollerslev (1998) 提出和使用，目前已得到学术界和业界的广泛认可和使用，它主要是基于日内的高频时间序列计算波动率的全新方法。它没有模型和参数估计，计算简便，并在实践中被证明有更好的估计精度和预测效率。不仅如此，已实现波动率也具有完整的理论基础：当日内收益率达到足够高的采样频率时，已实现波动率就可以无限趋向于瞬时波动率在样本区间内的积分，也就是日波动率的自然度量。

其计算过程如下：

第 t 天的样本区间为 $[t, t+T]$ ，则将样本区间等分为 N 个子区段，每个区段时长为 Δ ，则

$$T = N * \Delta$$

则该 t 日内的已实现波动率为

$$\sigma_{t,\Delta}^2 = \sum_{j=1}^N r^2(t, j\Delta) \quad (3.8)$$

其中， $r(t, j\Delta)$ 为 t 日第 j 个 Δ 时段的对数收益率。

如此，可以计算得 t 日的已实现波动率为

$$\sigma_t^2 = \sigma_{t,\Delta}^2 + (r_t^o)^2 \quad (3.9)$$

其中， r_t^o 为 t 日的隔夜对数收益率。这里，本文特意在传统的已实现波动率的基础上加上隔夜收益方差，构造了一个新的已实现波动率指标，这一改进一方面是为了更全面的描绘当日的市场波动率，另一方面是为了匹配本文的情绪指标，本文的情绪指标不仅考虑了白天交易期间的文本信息，还考虑了隔夜开盘前的文本信息。

在计算已实现波动率时，取样频率过高，虽然可以反映到更多的市场交易信息，但也会夹杂更多的噪音信息，所以取样频率始终为宜，实践中 Δ 一般取 5min~15min，而 Bandi 和 Russell (2006) 基于已实现波动率进行了最优抽样频率的研究，认为中国市场的最佳抽样频率为 13 分钟。所以本文取 $\Delta = 15 \text{min}$ 。

二、投资者情绪对股市波动性的影响分析

同样，本文在这里仍然采用同期的时间序列数据进行回归。过去的大量文献已经证明，金融资产收益波动的时序具有滞后相关性，所以本文拟引入已实现波动率的 s 阶滞后项作为控制变量，具体的模型构建如下：

$$rv_t = \phi + \sum_{j=1}^s \rho_j * rv_{t-j} + \beta_1 * D_t * |Sent_t| + \beta_2 * (1 - D_t) * |Sent_t| + \theta * I_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (3.10)$$

其中， rv_t 为 t 日的已实现波动率（用标准差表示）； D_t 为虚拟变量，

$$D_t = \begin{cases} 1, & \text{if } Sent_t > 0 \\ 0, & \text{others} \end{cases}; \quad s \text{ 滞后阶数通过 AIC 和 BIC 信息准则确定。}$$

如此，通过回归结果可以知道，系数 β_1 和 β_2 显著性程度越高，值越大，则说明该情绪指标 $Sent_t$ 对市场波动性的影响越大，并且若符号为正，则说明会加剧市场的波动，反之则是抑制市场的波动；同时，借助虚拟变量 D_t 可以检验看涨倾向和看跌倾向的投资者情绪是否会对股市收益波动性产生非对称冲击；若虚拟变量 I_t 的系数 θ 显著，则说明存在显著的周末效应，反之则不存在。

三、投资者情绪与股市波动性的 VAR 模型

同市场收益率一样，在实际的市场交易活动中，投资者情绪不仅会影响市场当期的波动率，也会影响未来跨期的市场波动率；并且，市场波动性的变化亦会对当前及未来的投资者情绪产生影响。所以，本文拟通过建立两元变量的 VAR 自回归模型来分析市场波动性与投资者情绪之间的相互影响关系：

$$\begin{cases} |Sent_t| = c_0 + \sum_{j=1}^q W_{1j} * rv_{t-j} + \sum_{j=1}^p V_{1j} * |Sent_{t-j}| + \delta_1 * g_t + \omega_1 * h_t + \lambda_1 * dchn_t + \varphi_1 * dshibor_t + \varepsilon_{1t}, \varepsilon_{1t} \sim N(0, \sigma_{1t}^2) \\ rv_t = d_0 + \sum_{j=1}^q W_{2j} * r_{t-j} + \sum_{j=1}^p V_{2j} * |Sent_{t-j}| + \delta_2 * g_t + \omega_2 * h_t + \lambda_2 * dchn_t + \varphi_2 * dshibor_t + \varepsilon_{2t}, \varepsilon_{2t} \sim N(0, \sigma_{2t}^2) \end{cases} \quad (3.11)$$

其中， q 是根据 AIC 和 BIC 信息准则确定的滞后阶数。

根据回归结果可以分析，若系数 W_{1j} 显著，则说明市场波动性的变化可以影响未来的投资者情绪变化，有助于预测投资者情绪；若系数 V_{2j} 显著，则说明投资者情绪指标的变化能够影响甚至预测未来市场的波动率。

四、开盘前投资者情绪对市场波动性的预测影响模型

模型构建如下：

$$\begin{aligned} rv_t = & \phi + \sum_{j=1}^s \rho_j * rv_{t-j} + \beta_1 * F_t * |Sent_{At}| + \beta_2 * (1 - F_t) * |Sent_{At}| \\ & + \theta * I_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \end{aligned} \quad (3.12)$$

其中， $|Sent_{At}|$ 表示 A 时段休市期间的看涨情绪指数的绝对值。

从而根据回归结果分析休市期间的发帖信息是否含有能够预测市场未来波动的信息；如果有，看涨倾向和看跌倾向的投资者情绪是否会对股市未来收益波动性产生非对称冲击；是否存在显著的周末效应。

第四章 研究样本与描述性统计

第一节 研究样本获取

一、样本时间选取

本文选取的样本时间段为：2014年6月1日至2015年12月31日。这期间，中国股市经历了一轮完整的牛市和熊市，而且在这次暴涨暴跌“过山车”式的行情中，市场情绪经历了极度乐观到极度悲观的过程，为本文对投资者情绪的研究提供了极好的样本。

同时本文为了检验全样本下的结果在牛市环境下和熊市环境下是否稳定存在，根据 Pagn 和 Sossounov（2003）的判定标准将整个样本区间 2014.06.01-2015.12.31（390个交易日）划分为 2014.06.01-2015.06.12（254个交易日）作为牛市子样本，2015.06.13-2015.12.31（136个交易日）作为熊市子样本。

表 4.1 样本区间

样本	时间区间	交易日
全样本	2014.06.01-2015.12.31	390 天
牛市子样本	2014.06.01-2015.06.12	254 天
熊市子样本	2015.06.13-2015.12.31	136 天

二、数据来源

（一）网络文本信息的获取

股吧论坛是一个开放性程度较高的网络交流社区，一般只要注册会员就可以在其中参与讨论和交流，参与群体非常广泛，而且参与者基本都是中国的散户，与本文的研究非常吻合。

本文选择中国目前最大的股吧论坛—东方财富网股吧作为所有的网络发帖信息来源。之所以选择东方财富网，主要基于以下两个原因：

（1）东方财富网的网站流量巨大，在我国金融类网站中流量排名第二¹，能够充分代表和及时反映中小投资者的意见和情绪。

（2）东方财富网上保留了大量的帖子，可追溯时间长，最早可追溯至两年

¹根据互联网协会 <https://www.chinarank.org.cn/>的网站流量排名

前，可以为本文的研究提供所有所需的数据。

东方财富网股吧论坛主要分为两类，一类是针对沪深股市所有股票的个股讨论区，一类是针对某一个主题的主题讨论区，如财经评论吧，一带一路吧，创业板指吧等，投资者可以根据自己需要进入对应的讨论区搜寻相关信息或参与讨论和发表意见。

本文主要研究的是市场的投资者情绪，所以选取上证指数吧作为信息来源，该股吧设立时间早，并且参与者众多，交流活跃，每天都有大量发帖信息，为本文提供了大量的研究样本。

相对于内容丰富的主帖而言，大部分的回帖都很简略，比如最常见的“顶”、“支持楼主”等，很难判断其情感倾向，信息含量很少，所以本文只考虑主贴的内容。

本文在整个样本区间，579 个自然日，390 个交易日，共提取到东方财富网股吧的上上证指数吧 1,162,130 条原始发帖信息。

（二）网络文本挖掘与分类

本文在提取到样本区间的 100 多万条帖子之后，还需运用如前文所述的文本挖掘和分类技术，通过机器学习对提取到的大量的帖子进行意见分类。股吧论坛帖子的意见类型主要分为三类：“看涨类”、“中立/噪音类”、“看跌类”。

利用机器分类之前，需要从样本测试集中选择一定量的帖子作为训练集，进行人工分类；一方面是让计算机借助某一算法学习训练集的分类结果构建自己的分类模型，另一方面是针对训练集应用计算机的分类模型验证，一般是用最严格的 10 折交叉验证，准确率通常达到 70% 以上就说明分类模型有效。

训练集人工分类结果的准确率直接影响计算机分类模型的有效性和准确率。所以，一方面本文从 100 多万条文本中随机抽取若干条帖子（为了具有客观代表性，样本区间每个月份都抽取一定量文本），由作者本人同另外 2 位拥有 2 年以上股票投资经验的金融硕士研究生一起分类，共同商讨，若最后意见不一则归为中立/噪音类。最后，本文从中提取到“看涨类”、“中立/噪音类”、“看跌类”各 1500 条文本，共 4500 条文本作为训练集。

为了保证对待分类测试集分类的准确率，在特征词提取中，需要在保证分类准确率的情况下，从分词中选取尽可能多的特征词，以最大程度涵盖不同类型文本的特征；为此，本文使用目前应用最普遍的数据挖掘开源包 Weka，其中含有许多各种类型的算法可供选择，将多种算法模型准确率对比，并经多次调试和验证后，本文选择改进型的支持向量机算法 SMO（Sequential minimal optimization, SMO，序列最小优化算法），并从 8465 个分词中提取了 5000 个特征词。

表 4.2 是在 Weka 中选择几种不同的分类算法的准确率对比：

表 4.2 不同分类算法结果准确率

算法	10 折交叉检验	训练集再分类检验
J48-决策树	58.96%	82.33%
KNN 算法	40.62%	50.00%
贝叶斯算法	69.84%	80.09%
SMO-支持向量机	71.29%	93.71%

注：本文主要选择了以上四种比较常用的分类算法进行对比，SMO 算法的准确率占有绝对的优势；其中，这里 KNN 算法是在 K=3 最优条件下的准确率；训练集再分类检验是指用构建的分类模型对训练集重新分类的对比结果。

下表 4.3 展示了计算机分类与人工分类的 10 折交叉验证结果，对角线上的比例相加即为总体分类的准确率，达到了 71% 以上，说明该分类算法模型有效；并且，如果用训练集再分类检验，该算法模型的验证准确率甚至达到了 93% 以上。另外，我们也可以看到，模型将一部分看涨类或看跌类的帖子分类到中立/噪音类里，只有不到 10% 的帖子混淆了看涨和看跌的分类，而在训练集再分类检验中，仅有 2.2% 的帖子混淆了看涨和看跌的分类。说明本文的分类模型只是弱化了意见，但在意见方向上的判断错误非常少，总体方向正确。

从分类结果我们也可以看出，东方财富网的股吧论坛帖子有超过 60% 是本文所需要的带有意见倾向性的帖子，说明其中确实含有大量的投资者信息；另外，全样本集分类后，看跌类帖子的比例要高于看涨类的帖子，主要是在这一次的股灾期间，投资者情绪极度悲观，在股吧中尽情宣泄自己的负面情绪所致。

表 4.3 SMO 算法在训练集和全样本测试集中的表现

类别	人工分类结果	SMO 分类结果 (%)		
		看跌	中立/噪音	看涨
看跌	33.33%	24.71%	4.58%	4.04%
中立/噪音	33.33%	5.40%	23.69%	4.24%
看涨	33.33%	5.58%	4.87%	22.89%
训练集	4,500 条	35.69%	33.13%	31.18%
全样本集	1,146,651 条	37.91%	38.02%	24.07%

注：这里的测试集分类结果是 10 折交叉检验结果；训练集共 4,500 条，原帖数量共 1,162,130 条，预处理无关帖子后剩下 1,146,651 条；粗体表示被正确分类的比例。

（三）股市交易数据样本

上证大盘指数是中国投资者（包括机构投资者和个人投资者）参与股票交易最重要的参考指数，因此本文主要以上证综指来度量中国股市的整体收益和波动情况。

本文从 wind 上选取并导出了样本区间上证指数分钟级的收盘数据，以此计算所需要的对数收益率和已实现波动率。

（4）宏观控制变量数据获取

本文主要从 wind 上选取了样本区间内规模以上工业增加值环比增速 g ，通货膨胀率 h （CPI 环比表示），7 天银行间同业拆借利率 $Shibor$ ，美元兑人民币汇率 chn 四个宏观指标数据，并以此计算得本文的控制变量。

第二节 研究样本的描述性统计

一、论坛发帖量统计描述

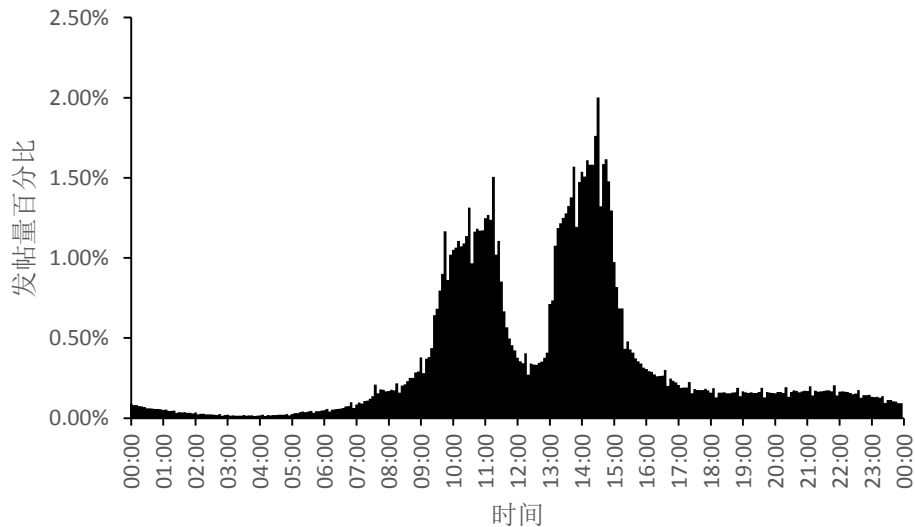


图 4.1 发帖量在一天中的分布

图 4.1 显示了上证指数吧一天 24 小时发帖量分布情况。我们可以发现上证指数吧在每天上午（9:30-11:30）和下午（13:00-15:00）两个交易时段是一天发帖量的高峰，非交易时段发帖相对较少；另外，收盘之后也即傍晚和晚上（17:00-22:00）发帖量会出现一个小高峰，且维持较为稳定的比例，可能是因为股民一般

会在下班之后浏览股市行情，搜寻股票市场相关信息，并在股吧中发表对股市的看法，尤其是对后市的观点，参与到股吧的讨论中。总体来看，交易时段的发帖量占比达到 60.01%，非交易时段的发帖量占比也达到 39.99%，可见股吧中不仅交易时段蕴含了大量的信息，非交易时段的信息同样不可忽视。

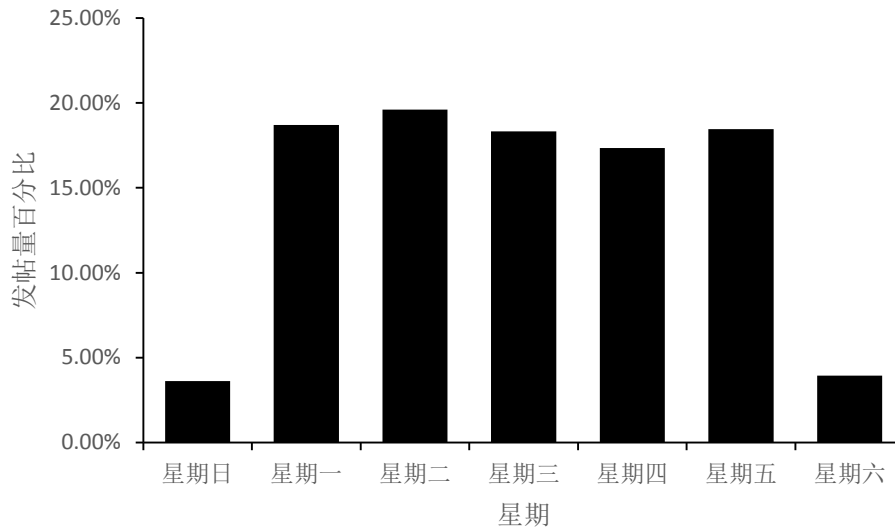


图 4.2 发帖量在一周中的分布

图 4.2 显示了股吧发帖量在一周当中的分布情况，我们发现在周末休市期间，发帖量很少，股吧较为冷清，股民更多是在搜集和了解相关信息；而在周一至周五，发帖量都很大，尤其是在周一，周二和周五，一周交易的开始和结束，发帖量最大，反映出股民在一周的开始和结束时对后市意见分歧较大。

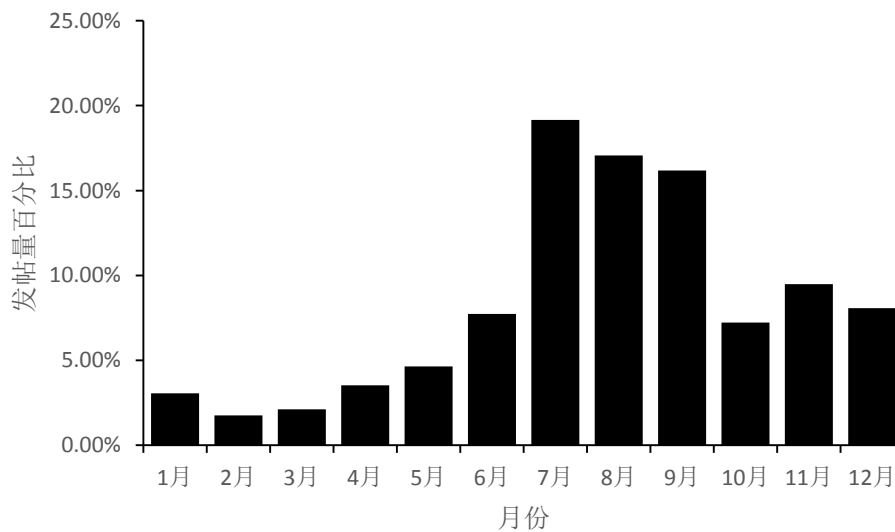


图 4.3 发帖量在一年（2015 年）中的分布

根据图 4.3 显示，我们可以看到发帖量主要集中在三季度和四季度，而一季度和二季度发帖量则较少；主要与中国的春节有关，春节前后股市热度会下降，股市资金在全年处于低位，而过了春节之后，资金慢慢进场，市场的活跃度也会明显上升，股民在股吧的讨论也会越来越多。另外，也与年报的披露时间有关，每年的 3,4 月份上市公司陆续公布年报，而在 8 月份则陆续公布半年报，从而会引发股市的一阵狂热。而需要指出的是，2015 年较为特殊，因为在 6 月底和 7 月份股市经历了股灾，市场弥漫着浓厚的悲观情绪，而股吧则成了股民吐槽和宣泄情绪的场所，所以 2015 年的三季度的发帖量占比特别高。

二、变量的描述性统计

本文主要涉及 17 个变量，

8 个被解释变量： r_t -每日的对数收益率； r_t^o -隔夜对数收益率； $r_t^{5\min}$ -开盘后 5min 的对数收益率； $r_t^{15\min}$ -开盘后 15min 的对数收益率； $r_t^{30\min}$ -开盘后 30min 的对数收益率； $r_t^{60\min}$ -开盘后 60min 的对数收益率； $r_t^{120\min}$ -开盘后 120min 的对数收益率； rv -每日的已实现波动率

2 个解释变量： $Sent_t$ -每日的看涨指数； $Sent_{At}$ -日间非交易时段的看涨指数；

3 个虚拟变量： I_t -研究周末效应的虚拟变量， D_t, F_t -研究看涨看跌情绪不对称效应的虚拟变量

4 个控制变量： g_t -规模以上工业增加值环比增速； h_t -通货膨胀率； $dshibor_t$ -利差； $dchn_t$ -人民币汇率变化。

表 4.4 变量描述性统计

变量	均值	中位数	最大值	最小值	标准差	偏度	峰度
$Sent_t$	-0.9298	-0.8099	1.6771	-4.3207	1.1404	-0.4775	2.8802
$Sent_{At}$	-0.0787	0.1403	2.3449	-3.5114	1.1311	-0.7007	3.0793
r_t	0.0014	0.0026	0.0560	-0.0887	0.0208	-1.1049	6.6565
r_t^O	-0.0018	-0.0004	0.0553	-0.0780	0.0112	-1.6056	13.9262
r_t^{5min}	-0.0011	-0.0002	0.0553	-0.0628	0.0100	-1.6473	15.7608
r_t^{15min}	-0.0009	0.0003	0.0485	-0.0667	0.0103	-1.8329	14.3195
r_t^{30min}	-0.0006	0.0002	0.0348	-0.0711	0.0100	-2.0253	14.3077
r_t^{60min}	0.0000	0.0007	0.0573	-0.0756	0.0115	-1.3669	12.2640
r_t^{120min}	0.0002	0.0012	0.0504	-0.0883	0.0131	-1.2682	10.3139
rv	0.0154	0.0123	0.0662	0.0037	0.0106	1.8895	7.2698
I_t	0.2128	0.0000	1.0000	0.0000	0.4098	1.4033	2.9692
g_t	0.4947	0.4600	0.9000	0.1300	0.1810	0.3093	2.9502
h_t	0.1159	0.0538	1.1946	-0.5460	0.3554	0.7215	4.3538
$dshbor_t$	-0.0022	0.0000	1.3420	-0.5340	0.1170	3.5514	51.4813
$dchn_t$	0.0008	0.0006	0.1136	-0.0341	0.0099	7.1338	77.1648
D_t	0.2051	0.0000	1.0000	0.0000	0.4043	1.4605	3.1331
F_t	0.5462	1.0000	1.0000	0.0000	0.4985	-0.1854	1.0344

从表 4.4 中我们看到，投资者情绪代理变量看涨指数 $Sent_t$ 的平均数-0.9298，中位数-0.8099，都小于 0，说明在本文的样本区间，投资者情绪总体较为悲观，看跌倾向为主，这可能也与 2015 年下半年的股灾和熊市行情有很大关系，使得市场整体陷入悲观情绪之中；而虚拟变量 D_t 的均值为 0.2051，说明样本区间内，390 个交易日，仅 20% 左右的交易日投资者情绪较为积极，80% 左右的时间我国投资者对股市都较为悲观，说明中国投资者对中国股市普遍较为悲观；但虚拟变量 F_t 的均值为 0.5462，相比一整日的投资者情绪，每日开盘前的投资者情绪则有超过一半的时间都较为积极看涨。

市场对数收益率 r_t 的平均数 0.0014，中位数 0.0026，说明样本区间内上证指数的收益率总体为正，但是很小；另外，我们发现开盘后几分钟的收益率的中位数和平均数都为负，显示出中国股市开盘时市场较弱，多为低开。从样本区间的已实现波动率 rv_t 看， rv_t 的平均数为 1.54%，中位数为 1.23%，最大更是达到 6.62%，足以说明在该样本区间中国股市较大的波动性，结果与样本区间经历的“过山车”

式的行情相一致。

第三节 变量的相关性分析

这里本文欲对与研究有关的变量做一相关性分析,尤其是观察本文的情绪指标与市场收益率和市场波动率的相关性,为后面的实证回归做铺垫和基础。

表 4.5 变量间的相关系数表

变量	r_t	$Sent_t$	$Sent_{At}$	rv_t	$dshibor_t$	h_t	$dchn_t$	g_t
r_t	1.00							
$Sent_t$	0.49***	1.00						
$Sent_{At}$	0.21***	0.78***	1.00					
rv_t	-0.20***	-0.52***	-0.44***	1.00				
$dshibor_t$	-0.04	-0.05	-0.05	-0.03	1.00			
h_t	-0.07	-0.16***	-0.17***	0.11**	0.09*	1.00		
$dchn_t$	0.01	0.00	0.04	-0.02	0.01	0.12**	1.00	
g_t	0.05	0.20***	0.12**	-0.06	0.03	0.15***	0.004	1.00

注:表中数据为各变量的相关系数。***, **, *分别表示在 1%, 5%, 10%的显著性水平下显著。

表 4.5 显示了本文几个重要变量的相关关系。从表中发现:

首先,本文的投资者情绪指标—看涨指数 $Sent_t$ 与上证指数的对数收益率 r_t 的相关性达到了 49.5%,看涨指数 $Sent_{At}$ 与上证指数的对数收益率 r_t 的相关性达到了 21%,并且在 1%的显著性水平下显著相关,符合本文的预期。为了更加直观,图 4.4 显示了看涨指数 $Sent_t$ 与上证指数对数收益率的时间序列图,图中我们看到两个变量在变化上有很明显的趋同性。另外,图 4.5 是上证指数与看涨指数的时间序列图,我们发现投资者情绪会随着上证指数的上升而高涨,随着上证指数的下跌而低落;并且在指数出现上升趋势之前投资者情绪相对处于较低点,在指数出现下跌趋势尤其是大跌之前投资者情绪相对处于较高点,另外在指数上涨阶段的情绪的波动要比指数在下跌或者震荡期间小得多。这也正应了股市中常说的“行情总是在绝望中产生,在犹豫中上涨,在疯狂中死亡”。

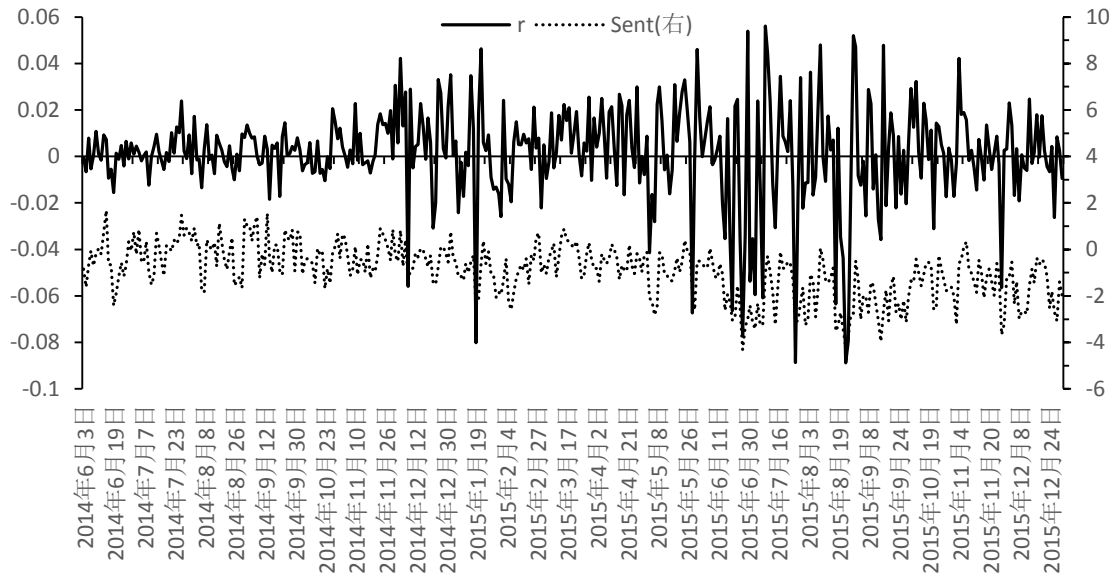


图 4.4 上证指数收益率与情绪指标时间序列图

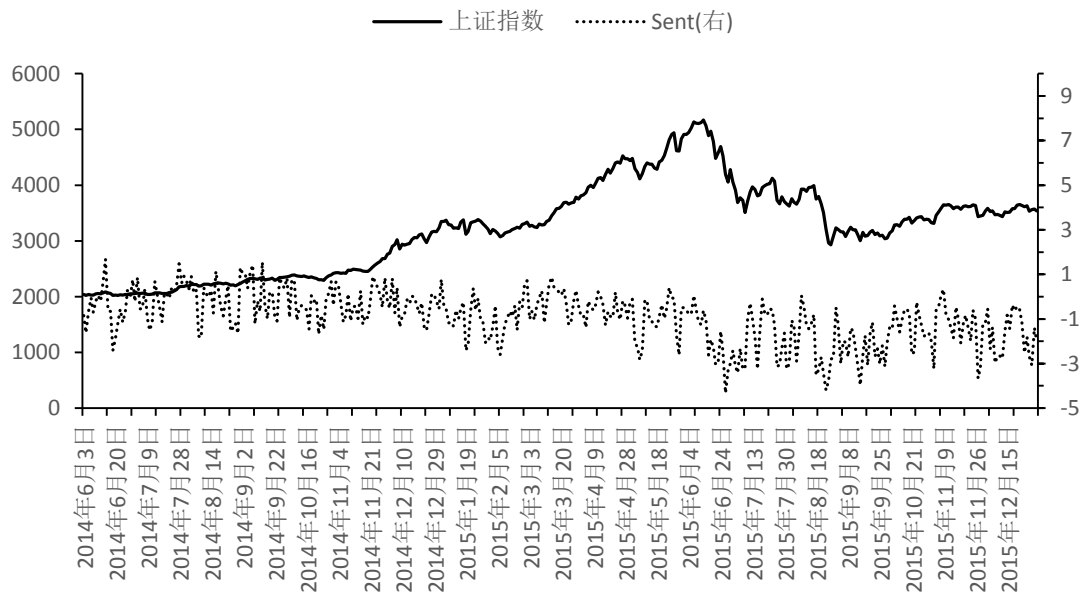


图 4.5 上证指数与情绪指标时间序列图

其次，看涨指数 $Sent_t$ 与上证指数的已实现波动率 rv_t 的相关性达-52.2%，并且在 1% 的显著性水平下显著相关，说明股票市场的波动确实与投资者情绪有密切的关系，符合本文的预期。同时在样本区间内我们也发现看涨指数和市场波动负相关，从图 4.6 我们也可以观察到，看跌的投资者情绪会加剧市场的波动，而看涨的投资者情绪反而会平抑市场波动。

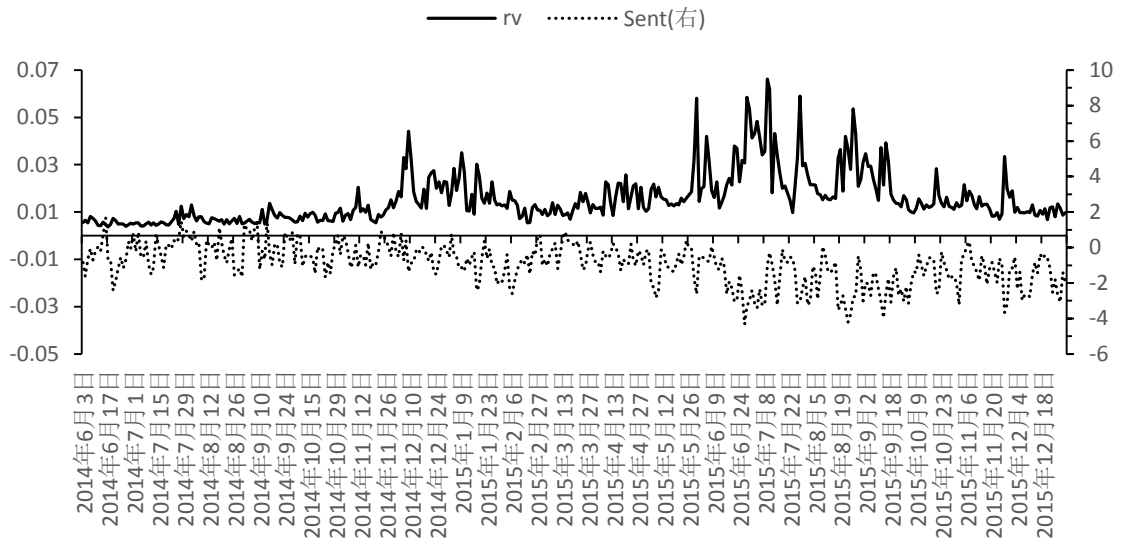


图 4.6 上证指数已实现波动率与情绪指标时间序列图

最后，我们看到上证指数对数收益率 r_t 和控制变量 $g_t, dchn_t, dshibor_t, h_t$ 的相关性都不到 10%，且都不显著，说明中国股市的收益与宏观经济的相关性不强，并不能很好地反映宏观经济的情况，相反却与投资者情绪密切相关，这也说明中国股市的投机和非理性；而上证指数的已实现波动率 rv_t 只与控制变量 h_t 有显著的相关性，相关性达到 10%，说明通货膨胀率能够较明显的影响股市的波动，但相关性程度并不高，而其他宏观变量基本对股市的波动没有影响。

此外，值得注意的是，4 个宏观经济变量 $g_t, dchn_t, dshibor_t, h_t$ 与投资者情绪指标 $Sent_t$ 与 $Sent_{At}$ 的相关性都不超过 20%，相关性程度较低，所以基本不存在多重共线性的问题。

第五章 模型实证结果与分析

第一节 数据的平稳性检验

在对本文的时间序列数据进行回归之前，还需要检验本文的几个重要变量的平稳性，防止出现伪回归现象。

检验时间序列的平稳性，一般采用单位根检验。本文采用 ADF 检验，其原假设 H_0 ：存在单位根。本文在此处将对重要的几个研究变量进行平稳性检验。检验结果如下表所示：

表 5.1 变量平稳性检验表

变量	ADF 值	P-Value 值	1%水平临界值	5%水平临界值	10%水平临界值
r_t	-14.6770	0.0000	-3.4470	-2.8690	-2.5710
rv	-4.3450	0.0004	-3.4470	-2.8690	-2.5710
$Sent_t$	-5.6610	0.0000	-3.4470	-2.8690	-2.5710
$Sent_{At}$	-5.9630	0.0000	-3.4470	-2.8690	-2.5710
r_t^O	-19.7550	0.0000	-3.4470	-2.8690	-2.5710
r_t^{5min}	-9.2094	0.0000	-3.4470	-2.8690	-2.5710
r_t^{15min}	-17.9412	0.0000	-3.4470	-2.8690	-2.5710
r_t^{30min}	-16.4925	0.0000	-3.4470	-2.8690	-2.5710
r_t^{60min}	-16.5363	0.0000	-3.4470	-2.8690	-2.5710
r_t^{120min}	-15.7790	0.0000	-3.4470	-2.8690	-2.5710
$dchn_t$	-11.3714	0.0000	-3.4470	-2.8690	-2.5710
h_t	-3.2916	0.0159	-3.4470	-2.8690	-2.5710
g_t	-3.7719	0.0035	-3.4470	-2.8690	-2.5710
$dshibor_t$	-12.8876	0.0000	-3.4470	-2.8690	-2.5710

从表 5.1 可知，ADF 统计量的 P-Value 值基本都小于 1% 和 5%，拒绝存在单位根的原假设，说明本文的主要变量都不存在单位根，均为平稳时间序列。

第二节 投资者情绪对市场指数收益率的影响分析

一、上证指数收益率的自相关性分析

在进行看涨指数与上证指数收益率的回归之前，本文需要先分析上证指数收益之间的自相关性，并确定其自相关的滞后阶数。

表 5.2 上证指数收益率的自相关性检验

Auto Correlation	Partial Correlation	Lags	AC	PAC	Q-Stat	Prob
. *	. *	1	0.158	0.158	3.4589	0.063
* .	* .	2	-0.146	-0.176	6.4579	0.040
. .	. .	3	-0.018	0.039	6.5012	0.090
. *	. *	4	0.131	0.108	8.9482	0.062
. *	. .	5	0.091	0.052	10.122	0.072
* .	* .	6	-0.184	-0.186	15.034	0.020
. .	. .	7	-0.024	0.071	15.117	0.035
. *	. *	8	0.185	0.125	20.121	0.010
. .	* .	9	-0.027	-0.111	20.231	0.017
** .	* .	10	-0.252	-0.181	29.705	0.001
* .	. .	11	-0.129	-0.036	32.207	0.001
. *	. .	12	0.087	0.014	33.357	0.001
. *	. *	13	0.151	0.112	36.838	0.000
* .	. .	14	-0.08	-0.024	37.821	0.001
* .	. .	15	-0.088	-0.022	39.023	0.001

从表 5.2 可以发现，上证指数的对数收益率 r_t 序列的自相关性检验结果，Q 统计量的 P-Value 值均小于 1%，说明拒绝原假设， r_t 存在显著的自相关性。为此，本文将采用 BIC 和 AIC 信息准则法来确定上证指数收益率的自相关滞后阶数 d ；根据表 5.3 显示，三个信息准则显示的最有滞后阶数不一致，本文优先采用 AIC 准则确定，所以最后确定上证指数收益率的最优滞后阶数为 $d = 4$ 。因此，本文将引入 r_t 的 1~4 阶滞后变量，作为回归模型的控制变量。

表 5.3 上证指数收益率信息准则判别表

Lags	AIC	SC	HQ
0	-4.902	-4.892	-4.898
1	-4.913	-4.893*	-4.905
2	-4.922	-4.892	-4.910*
3	-4.915	-4.875	-4.899
4	-4.928*	-4.877	-4.908
5	-4.920	-4.858	-4.896
6	-4.920	-4.848	-4.891
7	-4.913	-4.831	-4.881
8	-4.915	-4.822	-4.878

注：*表示该信息准则下的最小值

二、投资者情绪对市场指数收益率的影响分析

从下表 5.4 回归结果可以看到，模型全样本的 R² 达到了 30.92%，熊市样本下的 R² 更是达到了 50%，说明模型有较好的拟合效果，模型的 F 统计量的 P-Value=0<1%，模型都整体非常显著。

无论是在熊市样本下还是在牛市样本下， $r_{t-1}, r_{t-2}, r_{t-3}$ 系数都为负，并且 r_{t-1}, r_{t-2} 的系数在 1% 显著性水平下显著。这在一定程度上说明上证指数收益率在中短期存在显著的反转效应。

在全样本下，该模型的研究变量看涨指数 $Sent_t$ 的回归系数为 0.0114，并且在 1% 的显著性水平下显著为正，这说明本文构建的投资者情绪指标——看涨指数 $Sent_t$ 对上证指数收益率有显著的正向影响。当看涨指数 $Sent_t$ 上升，将增加上证指数的收益率，从而推动市场指数的上涨，而当看涨指数 $Sent_t$ 下降甚至为负时，上证指数的收益率将减小，市场指数也趋于下跌。从回归结果看，看涨指数 $Sent_t$ 每上升 1 个单位，上证指数对数收益率将提升 1.14%。

而从牛市子样本和熊市子样本的回归结果看，无论是在牛市环境下还是熊市环境下，投资者情绪 $Sent_t$ 的系数都在 1% 显著性水平下显著为正，说明在不同的市场化境下投资者情绪 $Sent_t$ 都对上证指数收益率 r_t 有非常显著的正向影响；而在熊市环境下，系数为 2.24%，显著大于在牛市环境下的 1.27%，说明在熊市环境下投资者情绪指 $Sent_t$ 对上证指数收益率 r_t 的影响程度要比在牛市环境下大得多。

模型 (3.5) 的回归结果如下:

表 5.4 情绪指标与市场指数收益率回归结果

模型变量	全样本	牛市子样本	熊市子样本
α	0.01593*** (5.1680)	1.8405*** (4.5275)	-0.2926 (-1.4325)
r_{t-1}	-0.1478*** (-2.9104)	-0.2723*** (-4.2910)	-0.3168*** (-3.8659)
r_{t-2}	-0.1916*** (-4.2594)	-0.1830*** (-3.3013)	-0.2148*** (-3.1628)
r_{t-3}	-0.0654 (-1.4577)	-0.0820 (-1.5084)	-0.0708 (-1.0414)
r_{t-4}	0.0309 (0.6854)	0.0482 (0.8756)	-0.0425 (-0.6220)
$Sent_t$	0.0114*** (11.5057)	0.0127*** (9.4897)	0.0224*** (10.1317)
I_t	-0.0021 (-0.9450)	-0.0032 (-1.4905)	-0.0035 (-0.8017)
h_t	0.0003 (0.0959)	-0.0010 (-0.4383)	-0.0027 (-0.3587)
$dshibor_t$	-0.0009 (-0.1116)	-0.0009 (-0.5160)	0.0535 (1.3406)
$dchn_t$	0.0521 (0.5687)	-0.2968*** (-4.4523)	0.02765 (1.6537)
g_t	-0.0060 (-1.1661)	-0.0037 (-0.8631)	-0.0217 (-0.9765)
R-squared	0.3092	0.3152	0.4960
F-statistic	16.7823	11.0025	12.3009

注: 表中***, **, *分别表示在 1%, 5%, 10%的显著性水平下显著; 圆括号中为系数的 t 值统计量。

此外, 模型中的周末效应虚拟变量 I 的系数为负但均不显著。一般在周末或假期, 股民有更多的时间和精力来浏览市场各方面信息, 并在股吧中会有更多的交流、讨论和发帖, 对之后首个交易日的收益率往往会产生消极的影响, 可能是投资者对新一周的投资刚开始会持较为谨慎的态度, 为此模型显示经过周末或假

期的首个交易日市场的收益率会相对较低,影响约-0.2%~-0.3%;但是影响很小,并不显著,说明看涨指数 $Sent_t$ 对上证指数收益率 r_t 的影响基本不存在明显的周末效应。

三、投资者情绪和市场收益率的相互影响—VAR 模型与格兰杰检验

在得到本文的 VAR 模型 (3.6) 之前,需要先确定 VAR 模型的最有滞后阶数。

本文将利用 AIC 和 BIC 信息准则法确定 VAR 模型滞后阶数 p :

表 5.5 VAR 模型滞后阶数信息准则判别表

	Lags	AIC	SC	HQ
全样本	$p = 6$			
	0	-2.091	-1.988	-2.050
	1	-2.838	-2.694	-2.781
	2	-2.923	-2.737*	-2.850
	3	-2.956	-2.729	-2.866*
	4	-2.963	-2.695	-2.857
	5	-2.971	-2.661	-2.848
	6	-2.974*	-2.623	-2.835
	7	-2.960	-2.567	-2.804
	8	-2.956	-2.523	-2.784
牛市子样本	$p = 5$			
	0	-3.148	-3.006	-3.091
	1	-3.568	-3.369*	-3.488*
	2	-3.581	-3.324	-3.478
	3	-3.582	-3.268	-3.455
	4	-3.589	-3.219	-3.440
	5	-3.593*	-3.165	-3.420
	6	-3.581	-3.097	-3.386
	7	-3.555	-3.013	-3.337
	8	-3.537	-2.939	-3.296
熊市子样本	$p = 1$			

	0	-1.840	-1.626	-1.753
	1	-2.438*	-2.139*	-2.317*
	2	-2.408	-2.023	-2.252
	3	-2.406	-1.935	-2.215
	4	-2.362	-1.805	-2.135
	5	-2.313	-1.670	-2.052
	6	-2.293	-1.564	-1.997
	7	-2.274	-1.460	-1.943
	8	-2.257	-1.357	-1.891

注：*表示该信息准则下的最小值

根据 BIC 和 AIC 的信息准则法和表 5.5 显示的最优滞后阶数，本文确定在全样本下的 VAR 模型的最优滞后阶数 $p=6$ ，在牛市子样本下的 VAR 模型(3.6)的最优滞后阶数 $p=5$ ，在熊市子样本下的 VAR 模型的最优滞后阶数 $p=1$ 。

表 5.6 情绪指标与市场收益率的 VAR 模型与格兰杰检验结果

自变量	因变量: $Sent_t$			因变量: r_t		
	全样本	牛市子样本	熊市子样本	全样本	牛市子样本	熊市子样本
c_0/d_0	-0.3725*** (-2.6448)	-0.4346*** (-3.1912)	-1.4808*** (-3.4755)	0.0006 (0.1494)	-0.0015 (-0.5132)	-0.0055 (-0.4000)
r_{t-1}	17.9644*** (7.3797)	21.6212*** (6.1327)	16.0631*** (4.9234)	0.1559** (2.3960)	0.1105 (1.4468)	0.2406** (2.2709)
r_{t-2}	-5.3821** (-2.0612)	-3.8941 (-1.0449)		-0.1560** (-2.2351)	-0.0788 (-0.9750)	
r_{t-3}	-5.2258** (-2.0249)	-6.4815* (-1.7336)		-0.0538 (-0.7808)	-0.0258 (-0.3179)	
r_{t-4}	1.6323 (0.6301)	3.2202 (0.8617)		0.1570** (2.2674)	0.2186** (2.6988)	
r_{t-5}	-6.7284*** (-2.5920)	-7.4902** (-2.1485)		-0.0511 (-0.7368)	-0.1422* (-1.8825)	
r_{t-6}	-5.8435** (-2.4433)			-0.1269** (-1.9858)		
$Sent_{t-1}$	0.3082*** (4.7426)	0.2027*** (2.6729)	0.2513*** (2.8435)	-0.0002 (-0.1166)	-0.0021 (-1.2747)	-0.0028 (-0.9733)

$Sent_{t-2}$	0.1756*** (2.6068)	0.2002** (2.5612)		0.0021 (1.1862)	0.0004 (0.2587)	
$Sent_{t-3}$	0.1240* (1.8194)	0.0058 (0.0731)		-0.0005 (-0.2838)	-0.0028 (-1.6378)	
$Sent_{t-4}$	0.0308 (0.4515)	0.0222 (0.2818)		-0.0017 (-0.9744)	-0.0007 (-0.3992)	
$Sent_{t-5}$	0.2233*** (3.3116)	0.1433** (2.1481)		0.0032* (1.7808)	0.0016 (1.1155)	
$Sent_{t-6}$	-0.0282 (-0.5063)			-0.0012 (-0.7833)		
h_t	-0.0878 (-0.7431)	-0.0862 (-0.6998)	-0.2987 (-1.0068)	-0.0041 (-1.2940)	-0.0039 (-1.4714)	-0.0100 (-1.0352)
$dshibor_t$	-0.2325 (-0.6901)	-0.1077 (-0.3402)	-2.1223 (-1.3118)	-0.0042 (-0.4681)	0.0019 (0.2722)	-0.1595*** (-3.0355)
g_t	0.4436** (1.9664)	0.4468** (2.0089)	0.4025 (0.4906)	0.0060 (0.9896)	0.0072 (1.5017)	0.0003 (0.0126)
$dchn_t$	5.2986 (1.3028)	10.9872 (0.9421)	5.9614 (1.2369)	0.0690 (0.6352)	-0.0732 (-0.2898)	0.1103 (0.7045)
R-squared	0.5710	0.3549	0.4012	0.0840	0.0908	0.1054
F-statistic	30.5305	9.1936	14.4059	2.1047	1.6701	2.5336
Granger Causality Test	13.0568***	9.8789***	22.1917***	1.0987	1.3338	0.8650

注：表中***, **, *分别表示在 1%, 5%, 10% 的显著性水平下显著；圆括号中为系数的 t 值统计量。

根据表 5.6 结果，当因变量是投资者情绪 $Sent_t$ 时，不管是牛市还是熊市， $Sent_{t-i}$ 的系数几乎都显著为正，尤其是 $Sent_{t-1}$ 的系数在 1% 显著性水平下显著为正，说明中国股市的投资者情绪存在较长时间的惯性效应，无论是负面情绪还是正面情绪，都会维持一段较长时间；同时，无论在牛市还是熊市环境下， r_{t-1} 的系数都在 1% 的显著性水平下显著为正，说明市场收益率对市场投资者情绪在短期内存在显著的正向影响，收益率越高，投资者情绪会变得越乐观，反之收益率越低，投资者情绪会变得越悲观。Granger 因果检验也表明市场指数的收益率是引发投资者情绪 $Sent_t$ 变动的 Granger 原因，市场收益率对未来的投资者情绪具有良好的预测作用。

当因变量为上证指数收益率 r_t 时，无论是在牛市环境还是熊市环境下，

$Sent_{t-i}$ 的系数都不显著, 说明 $Sent_{t-i}$ 对上证指数收益率均没有显著影响; Granger 因果检验也表明投资者情绪 $Sent_t$ 不是引发市场指数的收益率 r_t 变动的 Granger 原因。说明目前投资者情绪指标 $Sent_t$ 还无法用于预测未来市场收益率的变化。同时, r_{t-1} 的系数在熊市环境下显著为正, 但 r_{t-2} 的系数显著为负, 但在牛市环境下都不显著; 说明熊市环境下市场收益率在短期内存在惯性效应, 而长期来看则会反转。

四、开盘前投资者情绪对上证指数收益率的预测作用和时效性分析

下表 5.7 为根据隔夜网络股吧中发帖信息得到的投资者情绪指标对当日上证指数不同时段收益率的回归结果。

表中结果显示, 利用日间的休市期间(当日 15:00-次日 9:30)也即隔夜的非交易时段的发帖信息提取的看涨指数 $Sent_{At}$ 对开盘后一段时间内的市场收益率有非常显著的预测作用和正向影响, $Sent_{At}$ 系数在 1% 的显著性水平下显著为正; 开盘前投资者情绪越高涨, 情绪指标 $Sent_t$ 越大, 开盘后一段时间内的市场收益率越大。而且对开盘后 120 分钟甚至整日的市场收益率均有较显著的预测作用, 系数在 5% 的显著性水平下显著, 并且系数都显著为正。说明开盘前隔夜的非交易时段的股吧发帖信息中确实含有能够预测和影响次日市场收益率的重要信息, 因为投资者在每日收盘后都会在股吧中宣泄情绪, 发表看法, 讨论下一日的市场走势, 从而对次日股市的交易产生重要影响。

表 5.7 开盘前情绪指标与市场收益率的回归结果

	变量	r_t^O	r_t^{5min}	r_t^{15min}	r_t^{30min}	r_t^{60min}	r_t^{120min}	r_t
全样本	$Sent_{At}$	0.005*** (6.9202)	0.003*** (4.8759)	0.003*** (4.1679)	0.003*** (4.9677)	0.003*** (4.1582)	0.002*** (2.8943)	0.004*** (2.7429)
	r_{t-1}	-0.132*** (-3.5045)	-0.010 (-0.2878)	0.067** (1.9579)	0.044 (1.3643)	0.071* (1.8702)	0.085* (1.9035)	-0.001 (-0.0156)
	r_{t-2}	-0.069*** (-2.6130)	-0.067*** (-2.8408)	-0.034 (-1.4070)	-0.066*** (-2.8976)	-0.045* (-1.7047)	-0.047 (-1.5008)	-0.144*** (-2.7701)
	r_{t-3}	-0.065** (-2.4231)	-0.064*** (-2.6994)	-0.048** (-2.0057)	-0.042* (-1.8414)	-0.065*** (-2.4531)	-0.057* (-1.8190)	-0.035 (-0.6685)

第五章 模型实证结果与分析

	r_{t-4}	0.037 (1.3863)	0.066*** (2.7738)	0.067*** (2.7629)	0.074*** (3.2391)	0.091*** (3.4186)	0.101*** (3.1793)	0.097* (1.8503)
	g_t	0.002 (0.6211)	0.002 (0.5990)	0.001 (0.3055)	0.003 (1.1016)	0.001 (0.3198)	0.000 (0.1256)	0.005 (0.7854)
	$dshibor_t$	0.000 (-0.0394)	-0.001 (-0.2028)	0.000 (0.0926)	0.002 (0.4788)	0.003 (0.5936)	-0.001 (-0.1944)	-0.003 (-0.3296)
	h_t	-0.002 (-0.9693)	-0.001 (-0.9182)	-0.001 (-0.7291)	-0.001 (-0.7178)	-0.001 (-0.8033)	-0.001 (-0.6661)	-0.003 (-0.9274)
	$dchn_t$	0.003 (0.0540)	0.026 (0.5454)	0.025 (0.5161)	0.068 (1.4560)	0.045 (0.8242)	-0.038 (-0.5922)	0.039 (0.3729)
	I_t	0.001 (0.7017)	0.001 (0.5925)	0.001 (0.5786)	0.001 (0.7960)	0.002 (1.1455)	0.001 (0.5083)	0.002 (0.6890)
	α	-0.002 (-1.2569)	-0.001 (-1.0528)	-0.001 (-0.8456)	-0.002 (-1.4052)	0.000 (-0.2982)	0.000 (0.0790)	0.000 (-0.1488)
	R-squared	0.1733	0.1777	0.2059	0.2394	0.2188	0.1534	0.0837
	F-statistic	7.8586	8.1044	9.7214	11.8026	10.5047	6.7930	3.4244
牛市子样本	$Sent_{At}$	0.002*** (3.2726)	0.001** (2.3810)	0.001** (2.1170)	0.002** (2.4466)	0.001* (1.8866)	0.001 (0.7435)	0.002 (1.1595)
	r_{t-1}	-0.088** (-2.0415)	-0.041 (-1.2047)	0.067 (1.7652)	0.032 (0.8732)	0.018 (0.4186)	0.004 (0.0774)	-0.029 (-0.3292)
	r_{t-2}	-0.073** (-2.2889)	-0.054** (-2.0990)	-0.034 (-1.1956)	-0.070** (-2.5890)	-0.049 (-1.5407)	-0.079* (-1.8489)	-0.104 (-1.5820)
	r_{t-3}	-0.030 (-0.9596)	-0.029 (-1.1375)	-0.004 (-0.1361)	-0.003 (-0.1064)	-0.046 (-1.4898)	-0.036 (-0.8566)	-0.084 (-1.3125)
	r_{t-4}	-0.006 (-0.1951)	0.016 (0.6251)	-0.017 (-0.6284)	0.038 (1.4203)	0.037 (1.2040)	0.049 (1.1745)	0.118* (1.8353)
	g_t	0.000 (0.1591)	0.001 (0.4160)	0.001 (0.3887)	0.002 (1.1735)	0.003 (1.2918)	0.003 (0.8465)	0.006 (1.2484)

	$dshibor_t$	0.000 (0.0957)	0.000 (0.0187)	0.000 (-0.1422)	0.002 (0.7901)	0.004 (1.0531)	-0.001 (-0.2299)	0.003 (0.4886)
	h_t	-0.001 (-0.5668)	0.000 (-0.2702)	0.000 (-0.4308)	0.000 (-0.2210)	-0.002 (-1.1886)	-0.001 (-0.7130)	-0.003 (-0.9614)
	$dchn_t$	0.082 (0.6846)	0.087 (0.9052)	0.078 (0.7441)	0.045 (0.4472)	-0.014 (-0.1175)	0.049 (0.3067)	-0.021 (-0.0848)
	I_t	-0.001 (-0.6954)	-0.001 (-0.8647)	0.000 (-0.0979)	0.000 (0.0450)	0.001 (0.6829)	0.002 (1.4534)	0.003 (1.3133)
	α	0.000 (-0.0489)	0.000 (0.0649)	-0.001 (-0.4425)	-0.001 (-0.8529)	-0.001 (-0.3970)	0.000 (0.0694)	0.000 (0.0208)
	R-squared	0.0613	0.0484	0.1065	0.1101	0.0798	0.0467	0.0596
	F-statistic	1.5613	1.2153	2.8494	2.9583	2.0734	1.1704	1.5142
熊市子样本	$Sent_{At}$	0.008*** (4.9300)	0.005*** (3.4470)	0.005*** (2.9950)	0.005*** (3.7000)	0.006*** (3.6294)	0.005*** (2.8461)	0.006* (1.6996)
	r_{t-1}	-0.209*** (-2.8525)	-0.028 (-0.4110)	0.020 (0.2923)	0.004 (0.0550)	0.025 (0.3409)	0.053 (0.6504)	0.007 (0.0463)
	r_{t-2}	-0.064 (-1.3835)	-0.079* (-1.8557)	-0.039 (-0.9140)	-0.065 (-1.6105)	-0.053 (-1.1280)	-0.040 (-0.7796)	-0.153* (-1.6913)
	r_{t-3}	-0.098** (-2.1247)	-0.094** (-2.2160)	-0.093** (-2.1689)	-0.078* (-1.9187)	-0.093** (-2.0031)	-0.091* (-1.7835)	0.003 (0.0329)
	r_{t-4}	0.029 (0.6288)	0.064 (1.4879)	0.097** (2.2489)	0.070* (1.7126)	0.085* (1.8266)	0.089* (1.7396)	0.026 (0.2830)
	g_t	0.001 (0.0670)	-0.005 (-0.3426)	-0.01 (-0.7505)	-0.004 (-0.3081)	-0.023 (-1.5667)	-0.027* (-1.7390)	-0.018 (-0.6403)
	$dshibor_t$	-0.031 (-1.1065)	-0.034 (-1.3233)	0.004 (0.1543)	-0.019 (-0.7732)	-0.024 (-0.8592)	-0.023 (-0.7278)	-0.134** (-2.4236)
	h_t	-0.006	-0.008	-0.006	-0.006	-0.007	-0.009	-0.010

		(-1.2611)	(-1.6526)	(-1.3691)	(-1.4274)	(-1.3462)	(-1.5973)	(-1.0049)
	$dchn_t$	0.011 (0.1406)	0.053 (0.7130)	0.049 (0.6431)	0.092 (1.2982)	0.079 (0.9688)	-0.011 (-0.1267)	0.084 (0.5333)
	I_t	0.005 (1.6549)	0.004 (1.6043)	0.003 (1.1457)	0.003 (1.2019)	0.004 (1.2606)	-0.001 (-0.2622)	-0.001 (-0.1421)
	α	0.000 (-0.0314)	0.002 (0.3590)	0.005 (0.7598)	0.003 (0.4238)	0.012 (1.7130)	0.015 (1.9515)	0.011 (0.7832)
	R-squared	0.2842	0.3016	0.2879	0.3351	0.344	0.2865	0.1451
	F-statistic	4.9636	5.3976	5.0536	6.3009	6.564	5.0200	2.1218

注：表中***, **, *分别表示在 1%, 5%, 10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

从影响的时效性上看，在牛市环境下， $Sent_{At}$ 的系数仅在开盘后前 30 分钟非常显著，说明开盘前的投资者情绪仅对开盘后前 30 分钟的市场收益率有显著的预测作用和影响，而对开盘后 60 分钟及更长时间的市场收益率的影响则不再显著；在熊市环境下， $Sent_{At}$ 的系数在开盘后前 120 分钟都非常显著，开盘前的投资者情绪对开盘后上午的市场收益率都有很显著的预测和影响作用，但对下午的影响则明显减弱；并且在熊市环境下 $Sent_{At}$ 的系数要比在牛市环境下的 $Sent_{At}$ 系数大得多。说明开盘前的投资者情绪在熊市环境下对开盘后的市场收益率有更强影响作用和更长的影响时效。

另外，虽然经过假期和周末，情绪指标中包含有更多的投资者发帖信息，但是从实证结果看，虚拟变量 I 的系数都不显著，这说明开盘前的投资者情绪对上证指数收益率的预测影响作用不存在周末效应。

第三节 投资者情绪对股市波动性的影响分析

一、上证指数波动率的自相关性分析

已有大量的研究已经表明，金融市场的波动性往往表现出波动的丛集性，也即当期的波动率往往与其最近几期滞后波动率有一定的正相关关系，所以本文在进行上证指数波动率的回归研究之前，需要先对波动率进行自相关性分析，并确定滞后阶数。

表 5.8 上证指数已实现波动率的自相关性检验

Auto Correlation	Partial Correlation	Lags	AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.745	0.745	217.99	0.000
. *****	. *	2	0.646	0.205	382.45	0.000
. *****	. **	3	0.632	0.227	540.29	0.000
. ****	. *	4	0.607	0.117	686.07	0.000
. ****	. .	5	0.559	0.030	810.09	0.000
. ****	. .	6	0.541	0.071	926.74	0.000
. ****	. .	7	0.520	0.032	1034.5	0.000
. ****	. *	8	0.524	0.098	1144.4	0.000
. ***	* .	9	0.465	-0.078	1231.0	0.000
. ***	. .	10	0.449	0.035	1312.0	0.000
. ***	. .	11	0.439	0.012	1389.6	0.000
. ***	. *	12	0.466	0.128	1477.4	0.000
. ***	. *	13	0.481	0.101	1571.2	0.000
. ***	* .	14	0.415	-0.119	1641.1	0.000
. ***	. .	15	0.419	0.066	1712.7	0.000

表 5.8 上证综指已实现波动率 rv 序列的自相关性检验结果，Q 统计量的 P-Value 值均小于 1%，说明拒绝原假设， rv 存在非常显著的自相关性。

为此，本文仍将采用 BIC 和 AIC 信息准则法来确定上证指数收益波动率的自相关滞后阶数 s 。

表 5.9 上证指数波动率信息准则判别表

Lags	AIC	SC	HQ
0	-6.248	-6.238	-6.244
1	-7.055	-7.035	-7.047
2	-7.091	-7.061	-7.079
3	-7.138	-7.097*	-7.122
4	-7.144*	-7.093	-7.124*
5	-7.138	-7.076	-7.113
6	-7.135	-7.063	-7.107
7	-7.129	-7.047	-7.097
8	-7.133	-7.040	-7.096

注：*表示该信息准则下的最小值

根据表 5.9 显示，三个信息准则显示的最优滞后阶数几乎一致，本文优先采用 AIC 准则确定，所以最后确定上证指数收益波动率的最优滞后阶数为 $s = 4$ 。

因此，本文将引入 rv 的 1~4 阶滞后变量，作为回归模型的控制变量。

二、投资者情绪对市场指数波动性的影响分析

从下表 5.10 可以看到，全样本下模型的 R^2 达到了 64.1%，模型有很好的拟合效果，模型的 F 统计量的 $P\text{-Value}=0.000<1\%$ ，模型整体非常显著。

下表的回归结果显示，不同市场环境下， rv_{t-1} 的系数都在 1% 显著性水平下显著为正，充分显示了市场收益波动的惯性效应或丛集性。 $D_t^*|Sent_t|$ 的系数为正但不显著， $(1-D_t)^*|Sent_t|$ 为正且在 1% 显著性水平下显著，说明投资者情绪对市场的波动性存在正向影响作用，无论是积极的情绪还是消极的情绪，情绪越强，市场波动越剧烈，与我们的一般预期相符；同时，也说明乐观倾向的投资者情绪与悲观倾向的投资者对市场收益的波动性影响存在不对称性，乐观倾向的投资者情绪对市场波动性的影响并不显著，造成市场剧烈波动的主要是悲观倾向的投资者情绪，悲观的情绪指标大小每增加一个单位，市场已实现波动率将上升约 0.002~0.003。

此外，代表周末效应的虚拟变量 I_t 的系数不显著，说明投资者周末或假期在股吧的发帖信息并不能显著地影响市场的波动性，不存在显著的周末效应。

模型 (3.10) 的回归结果如下：

表 5.10 情绪指标与市场指数波动率的回归结果

模型变量	全样本	牛市子样本	熊市子样本
ϕ	-0.0001 (-0.1188)	0.0002 (0.1804)	-0.0045 (-0.9496)
rv_{t-1}	0.4393*** (8.6639)	0.4479*** (7.0842)	0.4296*** (4.9125)
rv_{t-2}	0.0813 (1.4889)	0.0928 (1.3449)	0.0828 (0.8761)
rv_{t-3}	0.1489*** (2.7292)	0.0455 (0.6575)	0.2379** (2.5584)
rv_{t-4}	0.0944* (1.9160)	0.2172*** (3.4968)	0.0160 (0.1883)
$D_t^* Sent_t $	0.0010	0.0011	0.0207

第五章 模型实证结果与分析

	(0.6950)	(0.8170)	(0.8163)
$(1-D_t)^* Sent_t $	0.0025*** (6.1569)	0.0024*** (3.9720)	0.0030*** (3.8733)
h_t	-0.0014 (-1.4403)	-0.0017 (-1.4166)	-0.0010 (-0.3460)
$dshibor_t$	0.0004 (0.1419)	0.0014 (0.5907)	-0.0194 (-1.1657)
g_t	0.0028 (1.4615)	0.0017 (1.0203)	0.0092 (1.0504)
$dchn_t$	-0.0050 (-0.1498)	0.0203 (0.2414)	-0.0053 (-0.1094)
I_t	-0.0003 (-0.3457)	-0.0005 (-0.6161)	0.0004 (-0.2077)
R-squared	0.6407	0.5615	0.6014
F-statistic	60.4536	27.5852	17.0057

注：表中***，**，*分别表示在 1%，5%，10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

三、投资者情绪和上证指数波动性的 VAR 模型与格兰杰检验分析

表 5.11 VAR 模型滞后阶数信息准则判别表

	Lags	AIC	SC	HQ
全样本	$q = 3$			
	0	-3.818	-3.797	-3.809
	1	-4.962	-4.900	-4.938
	2	-4.992	-4.889	-4.951
	3	-5.088*	-4.944*	-5.031*
	4	-5.081	-4.895	-5.007
	5	-5.072	-4.845	-4.982
	6	-5.064	-4.796	-4.958
	7	-5.063	-4.754	-4.940
牛市子样本	$q = 4$			
	0	-5.207	-5.179	-5.196
	1	-6.008	-5.923*	-5.974
	2	-6.054	-5.912	-5.997*
	3	-6.060	-5.861	-5.980
	4	-6.069*	-5.813	-5.966
	5	-6.041	-5.728	-5.915
	6	-6.027	-5.656	-5.878
	7	-6.021	-5.593	-5.849
熊市子样本	$q = 3$			
	0	-3.171	-3.128	-3.154
	1	-4.020	-3.892*	-3.968
	2	-4.035	-3.821	-3.948
	3	-4.144*	-3.844	-4.022*
	4	-4.099	-3.713	-3.942
	5	-4.070	-3.599	-3.879
	6	-4.026	-3.469	-3.800
	7	-3.977	-3.335	-3.716
8	-3.957	-3.229	-3.661	

注：*表示该信息准则下的最小值

根据 BIC 和 AIC 的信息准则法和表 5.11 显示的最优滞后阶数，本文确定在全样本下的 VAR 模型 (3.11) 的最优滞后阶数 $q=3$ ，在牛市子样本下的 VAR 模型 (3.11) 的最优滞后阶数 $q=4$ ，在熊市子样本下的 VAR 模型 (3.11) 的最优

滞后阶数 $q = 3$ 。

表 5.12 情绪指标与市场波动率的 VAR 模型与格兰杰检验结果

自变量	因变量: $ Sent_t $			因变量: rv_t		
	全样本	牛市样本	熊市样本	全样本	牛市样本	熊市样本
c_0 / d_0	0.368*** (2.8352)	0.647*** (4.9381)	0.398 (0.8410)	0.001 (1.1579)	0.004*** (2.7687)	-0.003 (-0.7295)
rv_{t-1}	9.874* (1.8260)	4.025 (0.6052)	12.359 (1.3971)	0.496*** (9.3959)	0.476*** (7.1248)	0.440*** (4.9580)
rv_{t-2}	-10.174* (-1.7021)	5.131 (0.6963)	-21.953** (-2.2612)	0.074 (1.2740)	0.137* (1.8470)	0.026 (0.2691)
rv_{t-3}	9.424* (1.7601)	-10.226 (-1.3873)	19.021** (2.1972)	0.248*** (4.7540)	0.04 (0.5381)	0.335*** (3.8492)
rv_{t-4}		1.767 (0.2652)			0.198*** (2.9620)	
$ Sent_{t-1} $	0.598*** (11.3568)	0.391*** (5.7219)	0.584 (6.5531)	0.001*** (2.8090)	-0.001 (-0.9080)	0.002*** (2.7891)
$ Sent_{t-2} $	-0.159*** (-2.6581)	-0.174** (-2.3787)	-0.25 (-2.5245)	-0.001 (-0.9192)	-0.001* (-1.9373)	-0.0004 (-0.4040)
$ Sent_{t-3} $	0.207*** (3.8621)	0.065 (0.8822)	0.194 (2.1391)	-0.0001 (-0.2820)	-0.0003 (-0.4348)	-0.0005 (-0.5219)
$ Sent_{t-4} $		-0.053 (-0.7810)			-0.0001 (-0.1942)	
h_t	0.117 (1.0912)	0.211** (2.1803)	0.256 (0.8312)	-0.001 (-0.5962)	0.0003 (-0.2961)	-0.001 (-0.2049)
$dshibor_t$	0.177 (0.5817)	0.007 (0.0312)	1.648 (0.9783)	0.001 (0.4361)	0.002 (0.7342)	-0.014 (-0.8422)
g_t	-0.229	-0.171	0.553	0.001	0.000	0.011

	(-1.1402)	(-1.0254)	(0.6322)	(0.5842)	(0.1322)	(1.2190)
$dchn_t$	-0.657	-1.393	-5.459	-0.002	-0.016	-0.008
	(-0.1820)	(-0.1621)	(-1.0974)	(-0.0632)	(-0.1851)	(-0.1571)
R-squared	0.4591	0.1802	0.3759	0.6049	0.5445	0.5812
F-statistic	31.9181	4.3442	7.5181	57.6790	23.6280	17.3491
Granger Causality Test	2.8671**	0.6310	2.8552**	2.6331**	2.2950*	3.1592**

注：表中***, **, *分别表示在 1%, 5%, 10% 的显著性水平下显著；圆括号中为系数的 t 值统计量。

当因变量是投资者情绪 $|Sent_t|$ 时，不管是牛市还是熊市， $|Sent_{t-1}|$ 的系数均为正，但是仅在牛市环境下 $|Sent_{t-1}|$ 的系数在 1% 的显著性水平下显著，而在熊市环境下系数并不显著；这再一次说明了投资者情绪的短期惯性效应，但主要是在牛市环境中，短期内乐观情绪能够被积聚并维持一定时间。同时，在牛市环境下， rv_{t-i} 的系数都不显著，说明市场的收益波动对未来投资者情绪变化没有显著影响；但在熊市环境下， rv_{t-1} 的系数不显著，但 rv_{t-2} 的系数在 5% 的显著性水平下显著为负， rv_{t-3} 的系数在 5% 的显著性水平下显著为正，说明在短期（滞后一期）市场收益波动对投资者情绪没有显著影响，但在中期（滞后两期）市场收益波动会减弱投资者情绪强度，而从长期来看（滞后三期）市场收益波动的增加会加剧熊市环境下的投资者悲观情绪。Granger 因果检验也表明，在牛市环境下，市场收益波动不是引起投资者情绪变动的 Granger 原因；但在熊市环境下，市场收益波动确是引起投资者情绪变动的 Granger 原因。

当因变量是市场收益波动 rv_t 时，在牛市环境下， $|Sent_{t-i}|$ 的系数都不显著，说明投资者情绪对市场收益波动的变化没有显著影响；但在熊市环境下， $|Sent_{t-1}|$ 的系数在 1% 的显著性水平下显著为正，说明投资者情绪强度对未来短期的市场收益波动存在显著的正向影响，投资者情绪强度越大，未来短期市场收益波动越剧烈。Granger 因果检验也表明，在牛市环境下，投资者情绪不是引起市场收益波动变化的 Granger 原因；但在熊市环境下，投资者情绪确是引起市场收益波动变化的 Granger 原因。

四、开盘前投资者情绪对上证指数波动性的预测作用

下表 5.13 为根据隔夜网络股吧中发帖信息得到的投资者情绪指标对当日上证指数波动率的模型 (3.12) 回归结果:

表 5.13 开盘前情绪指标与市场波动率的回归结果

模型变量	全样本	牛市子样本	熊市子样本
ϕ	0.001 (0.5850)	0.001 (0.8001)	-0.002 (-0.3350)
rv_{t-1}	0.419*** (7.8822)	0.474*** (7.1961)	0.368*** (4.0431)
rv_{t-2}	0.100* (1.7950)	0.112 (1.5700)	0.086 (0.9201)
rv_{t-3}	0.162*** (2.9441)	0.021 (0.2921)	0.254*** (2.7610)
rv_{t-4}	0.108** (2.1791)	0.230*** (3.6222)	0.004 (0.0431)
$F_t^* Sent_{At} $	0.001 (1.5332)	0.001 (1.5431)	0.001 (0.4851)
$(1 - F_t)^* Sent_{At} $	0.003*** (5.3360)	0.001* (1.7530)	0.004*** (3.9471)
$dchn_t$	-0.008 (-0.2391)	-0.005 (-0.0630)	-0.022 (-0.4461)
h_t	-0.001 (-0.9441)	-0.001 (-0.7682)	-0.0001 (-0.0352)
g_t	0.002 (0.9391)	0.001 (0.4960)	0.010 (1.1010)
$dshibor_t$	0.001	0.002	-0.016

	(0.2331)	(0.6871)	(-0.9852)
I_t	-0.0005	-0.001	-0.0002
	(-0.5671)	(-1.2551)	(-0.1333)
R-squared	0.6321	0.5391	0.6079
F-statistic	58.3203	25.2598	17.4645

注：表中***, **, *分别表示在 1%, 5%, 10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

上表的回归结果显示，模型的 F 统计量的 P-Value=0.000<1%，模型都整体非常显著，全样本下模型的 R^2 达到了 63.2%，模型有很好的拟合效果和解释能力。

在牛市环境下， $F_t^*|Sent_{A_t}|$ 和 $(1-F_t)^*|Sent_{A_t}|$ 的系数都为正，但都不显著，说明开盘前的投资者情绪强度对于当日的市场收益波动没有显著的影响和预测作用。但在熊市环境下， $F_t^*|Sent_{A_t}|$ 和 $(1-F_t)^*|Sent_{A_t}|$ 的系数都为正，且 $(1-F_t)^*|Sent_{A_t}|$ 的系数在 1%的显著性水平下显著为正，但 $F_t^*|Sent_{A_t}|$ 并不显著；说明在熊市中，开盘前的投资者情绪对当日的市场收益波动有显著的影响和预测作用，并且开盘前乐观倾向的投资者情绪与悲观倾向的投资者情绪对当日市场收益波动的影响同样具有不对称性，只有悲观倾向的投资者情绪对当日的市场收益波动有非常显著的预测能力，并会显著加剧市场收益的波动性，乐观倾向的投资者情绪对当日的市场收益波动没有显著影响。另外，从开盘前看涨指数 $|Sent_{A_t}|$ 的系数大小来看，在熊市中 $(1-F_t)^*|Sent_{A_t}|$ 的系数为 0.004,也即悲观倾向的投资者情绪每下降一个单位，将促使市场波动率增加 0.004。

同时，结果显示代表周末效应的虚拟变量 I_t 的系数不显著，所以不存在显著的周末效应，多一个假期或周末的帖子信息，并不会对于之后首日的市场波动性产生显著影响。

第四节 基于网络文本的情绪指标与传统情绪指标对比

为了更加全面的分析网络传媒文本中的投资者情绪，本文这一部分将与传统的投资者情绪指标尤其是已经在实际投资中运用的情绪指标作一比较。

一、对比传统情绪指标选取与构建

考虑到本文构建的网络文本情绪指标为日度数据以及数据的可得性，所以本文选取了以下 3 个常见的日度频率的传统情绪指标进行对比，包括显性投资者情

绪指标—新浪多空指数 $SinaBSI$ ，隐性投资者情绪指标—换手率 $Turn$ 和 ADR 指数。

（一）新浪多空指数 $SinaBSI$

本文主要参照黄德龙等人（2009）的方法构建显性情绪指标--新浪多空指数 $SinaBSI^2$ ，

$$SinaBSI_t = \frac{B_t + \frac{H_t}{3}}{B_t + H_t + S_t} \quad (5.1)$$

其中， B_t, H_t, S_t 分别为新浪财经每日多空调查的看多、看平、看空的投资者数量。考虑到看平投资者的数量可能会使该情绪指数失真，所以黄德龙等人（2009）等人把部分看平投资者数量加入到了分子中。

（二）换手率 $Turn$

换手率是指一段时期内证券被交易买卖的比例，是最传统的投资者情绪指标之一，被很多学者用于投资者情绪的研究。不仅如此，业界也早已将换手率作为市场情绪的观测指标之一，如申万宏源证券研究所就已将换手率纳入其市场投资者情绪的观测指标体系³。

本文使用的换手率是上证综指成分股的换手率，来源于 wind 数据库。

一般而言，当市场情绪高涨时，投资者尤其是短线的噪声交易者就会频繁追逐那些看起来比较容易很快获利的公司股票，导致市场交投趋于活跃，换手率也会相对较高；反之，当市场情绪低落时，投资者的投机意愿就会大大降低，买卖股票会比较谨慎，从而导致换手率相对较低。

（三）ADR 指数

ADR (Advanced-Decline Ratio) 指数是指一定时间内市场中累积上涨家数和累积下跌家数的比值，根据该指标我们可以分析市场中的多空力量对比情况，从而帮助我们判断未来的市场发展走势。该指数在构建时一般会选择几天之上的上涨家数和下跌家数，而非仅选择一天，这样做可以规避由于某一天的异常妨碍我们做出准确判断。但是具体天数这个参数的选择没有定规，本文采用大家常用的

² 新浪多空指数，主要是根据每日的新浪财经-中证报联合多空调查数据构建，<http://survey.finance.sina.com.cn/static/20205/20140602.html?pid=20205&dpc=1>，每天都会有大量的投资者在其中参与调查，对下一交易日的走势判断进行预测，所以该指数主要是用于对第二日市场走势的预测。

³ 详见于蒋俊阳分析师，申万宏源证券研究所，2015年12月发布的研报《从市场情绪中寻找涨跌信号》，其中包括四大情绪指数相关指标—普通投资者情绪，杠杆投资者情绪，特殊投资者情绪和市场技术面情绪，而换手率就是普通投资者情绪中的其中一个重要观测指标。

10 天，即选取上证综指成分股的 10 日 ADR 指数，数据来源于通达信金融终端的行情数据，具体计算为：

$$ADR = \frac{\sum_{n=1}^{10} \text{上涨家数}}{\sum_{n=1}^{10} \text{下跌家数}} \quad (5.2)$$

二、不同情绪指标对股票市场影响的对比

这里，本文同样选取 2014.06.01 至 2014.12.31 作为样本区间进行不同情绪指标的对比，主要从对市场收益率和对波动性的影响两个方面进行对比分析。

（一）不同情绪指标与市场收益率和波动性的相关性对比

从下表 5.14 可以看出，本文构建的基于网络文本的情绪指标 $Sent_t$ 与市场收益率 r_t 的相关系数 49.46%，远高于三个传统情绪指标，最高的 ADR 指数与市场收益率的相关性也仅为 25.13%，另外两个还不到 10%。

此外，本文构建的基于网络文本的情绪指标 $Sent_t$ 与市场已实现波动率 rv_t 的相关系数 -52.19%，其相关性程度远高于新浪多空指数 $SinaBSI_t$ 和 ADR_t 指数，但低于换手率 $Turn_t$ ，换手率 $Turn_t$ 与市场已实现波动率 rv_t 的相关性程度高达 62.21%。

表 5.14 不同情绪指标与市场收益率和波动率的相关系数表

情绪指标	$Sent_t$	$SinaBSI_t$	$Turn_t$	ADR_t
r_t	0.4946***	0.0921*	0.0191	0.2513***
rv_t	-0.5219***	-0.2206***	-0.6221***	-0.2915***

注：表中***，**，*分别表示在 1%，5%，10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

（二）不同情绪指标对市场收益率影响的对比

首先，本文分析不同情绪指标对同期市场收益率的影响，回归建模如下：

$$r_t = \alpha + \sum_{i=1}^d T_i * r_{t-i} + \beta * E_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (5.3)$$

其中， E_t 表示情绪指标，同前文滞后阶数 $d = 4$ 。下表是模型的回归结果。

表 5.15 不同情绪指标与同期市场收益率的回归结果

模型变量	全样本	牛市子样本	熊市子样本
$Sent_t$	0.011*** (11.5061)	0.013*** (9.4900)	0.022*** (10.1321)
R-squared	0.3091	0.3152	0.4959
F-statistic	16.7821	11.0029	12.3011
$Turn_t$	-0.001 (-0.3490)	0.003** (1.9831)	-0.005 (-1.2582)
R-squared	0.0622	0.0592	0.1358
F-statistic	2.7609	1.6618	2.2071
ADR_t	0.011*** (4.3721)	0.009*** (3.7658)	0.014** (2.5171)
R-squared	0.1071	0.0968	0.1667
F-statistic	5.0102	2.8521	2.8110

注：表中***，**，*分别表示在 1%，5%，10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

由于新浪多空指数 $SinaBSI_t$ 是在当日开盘前对当期市场走势的预测，所以这里不对新浪多空指数进行分析。从上表 5.15 结果可以看到，本文构建的情绪指标 $Sent_t$ 的系数无论是在牛市环境下还是熊市环境下都在 1%的显著性水平下显著，说明情绪指标对同期的市场收益率有很显著的影响；换手率 $Turn_t$ 的系数仅在牛市环境下在 5%的显著性水平下显著，在熊市环境下不显著，说明换手率仅在牛市情况下对市场收益率有影响，但在熊市环境下对市场收益率几乎没有作用； ADR_t 指数的系数在牛市环境下和熊市环境下分别在 1%和 5%的显著性水平下显著，说明该情绪指标对同期的市场收益率有较显著的影响和作用。

而从模型的拟合优度 R2 来看， $Sent_t$ 与 r_t 回归的 R2 在牛市和熊市环境下分别达到了 31.50%和 49.6%，而 $Turn_t$ 和 ADR_t 与 r_t 回归的 R2 均不到 20%，说明 $Sent_t$ 对同期市场收益率的影响程度较高，解释能力较强，而 $Turn_t$ 和 ADR_t 则很弱。

其次，本文分析不同情绪指标对市场收益率的预测作用。因为本文研究的情绪指标均为短期指标，时效性较短，所以这里仅分析滞后一期的情绪指标对当期市场收益率的预测作用，回归建模如下：

$$r_t = \alpha + \sum_{i=1}^d T_i * r_{t-i} + \beta * E_{t-1} + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (5.4)$$

其中, E_{t-1} 表示情绪指标, 同前文滞后阶数 $d=4$, 但是新浪多空指数 $SinaBSI_t$ 取当期值, 因为其在开盘前做出的对当期市场走势的预测。下表是模型的回归结果。

表 5.16 不同情绪指标对市场收益率预测作用的回归结果

模型变量	全样本	牛市子样本	熊市子样本
$Sent_{t-1}$	0.0005 (0.3821)	-0.003* (-1.9673)	-0.001 (-0.1721)
R-squared	0.0620	0.0581	0.1258
F-statistic	2.7639	1.6541	2.0101
$Turn_{t-1}$	-0.002 (-0.9981)	0.002 (1.4289)	-0.008* (-1.9242)
R-squared	0.0641	0.0512	0.1501
F-statistic	2.8643	1.4420	2.4769
ADR_{t-1}	-0.004 (-1.5691)	-0.001 (-0.2388)	-0.010* (-1.7427)
R-squared	0.0681	0.0429	0.1461
F-statistic	3.0382	1.2123	2.3920
$Sin aBSI_t$	0.001 (0.0571)	0.023 (1.4920)	-0.034 (-1.2701)
R-squared	0.0619	0.0516	0.1362
F-statistic	2.7470	1.4637	2.2109

注: 表中***, **, *分别表示在 1%, 5%, 10% 的显著性水平下显著; 圆括号中为系数的 t 值统计量。

从上表 5.16 结果可以看到, 不论在熊市环境还是牛市环境下, 本文构建的情绪指标以及三个传统的情绪指标的回归系数基本都不显著, 说明目前无论是传统的显性情绪指标, 隐性情绪指标还是基于网络传媒的情绪指标, 基于当期的情绪指标都还无法对下一期的未来市场收益率做出有效预测。

最后, 同前文一样, 这里本文分析和比较不同情绪指标对下一期的日内收益率的预测作用及其时效性。回归结果如下:

表 5.17 不同情绪指标对市场收益率预测时效性结果

	变量	r_t^0	$r_t^{5\min}$	$r_t^{15\min}$	$r_t^{30\min}$	$r_t^{60\min}$	$r_t^{120\min}$	r_t
$Sent_{At}$	全样本	0.005*** (6.9201)	0.003*** (4.8761)	0.003*** (4.1681)	0.003*** (4.9681)	0.003*** (4.1580)	0.002*** (2.8938)	0.004*** (2.7432)
	R-squared	0.1730	0.1782	0.2059	0.2391	0.2187	0.1533	0.0841
	F-statistic	7.8588	8.1037	9.7211	11.8030	10.5045	6.7931	3.4242
	牛市子样本	0.002*** (3.2731)	0.001** (2.3811)	0.001** (2.1172)	0.002** (2.4471)	0.001* (1.8870)	0.001 (0.7442)	0.002 (1.1600)
	R-squared	0.0611	0.0482	0.1068	0.1109	0.0802	0.0473	0.0607
	F-statistic	1.5611	1.2153	2.8491	2.9580	2.0727	1.1712	1.5142
	熊市子样本	0.008*** (4.9300)	0.005*** (3.4471)	0.005*** (2.9948)	0.005*** (3.7001)	0.006*** (3.6291)	0.005*** (2.8456)	0.006* (1.7001)
	R-squared	0.2841	0.3020	0.2873	0.3348	0.3441	0.2870	0.1445
	F-statistic	4.9639	5.3976	5.0540	6.3011	6.5642	5.0203	2.1220
$Turn_{t-1}$	全样本	-0.003*** (-3.2871)	-0.002** (-2.2292)	-0.002*** (-2.9035)	-0.001* (-1.9368)	-0.001 (-1.4901)	-0.001 (-1.5010)	-0.002 (-0.9981)
	R-squared	0.0851	0.1320	0.1832	0.1912	0.1802	0.1368	0.0642
	F-statistic	3.8826	6.3532	9.3851	9.8800	9.1556	6.6369	2.8636
	牛市子样本	0.0004 (0.5421)	0.001 (1.1391)	-0.001 (-0.8039)	0.001 (1.4858)	0.002** (2.2589)	0.001 (1.2881)	0.002 (1.4290)
	R-squared	0.0201	0.0310	0.0911	0.0942	0.0801	0.0393	0.0512
	F-statistic	0.5542	0.8482	2.6823	2.7771	2.3172	1.0961	1.4420
	熊市子样本	-0.008*** (-3.7131)	-0.005** (3.4470)	-0.005** (-2.2959)	-0.005*** (-2.6917)	-0.006*** (-2.6712)	-0.005** (-1.9831)	-0.008* (-1.9240)
	R-squared	0.2051	0.2472	0.2561	0.2901	0.3012	0.2630	0.1501
	F-statistic	3.6051	4.6022	4.8092	5.7067	6.0145	5.0031	2.4770
ADR_{t-1}	全样本	0.002 (1.7641)	-0.00004 (-0.0311)	0.0002 (0.1982)	0.001 (0.4810)	-0.0004 (-0.3223)	-0.002 (-1.2272)	-0.004 (-1.5693)
	R-squared	0.0661	0.1212	0.1650	0.1838	0.1748	0.1346	0.0678
	F-statistic	2.9764	5.7262	8.2673	9.4014	8.8721	6.5412	3.4243
	牛市子样本	0.001 (0.9651)	-0.0003 (-0.3353)	-0.0002 (-0.2001)	0.0002 (0.1918)	-0.0001 (-0.0493)	-0.001 (-0.6573)	-0.001 (-0.2390)
	R-squared	0.0231	0.0263	0.0892	0.0856	0.0602	0.0351	0.0433
	F-statistic	0.6262	0.7133	2.6081	2.5132	1.7141	0.9547	1.2116
	熊市子样本	0.004 (1.1772)	-0.0001 (-0.0378)	0.0003 (0.1020)	0.001 (0.2364)	-0.001 (-0.2741)	-0.003 (-1.0712)	-0.010* (-1.7430)
	R-squared	0.1271	0.2132	0.2251	0.2488	0.2606	0.2472	0.1461
	F-statistic	2.0434	3.7972	4.0551	4.6433	4.9541	4.5962	2.3920
	全样本	0.040***	0.027***	0.017***	0.018***	0.016**	0.004	0.001

Sin aBSI _t		(5.8040)	(4.4301)	(2.7632)	(3.0781)	(2.3262)	(0.5331)	(0.0571)
	R-squared	0.1360	0.1642	0.1821	0.2030	0.1872	0.1332	0.0621
	F-statistic	6.5842	8.2051	9.2778	10.6577	9.5856	6.3846	2.7470
	牛市子样本	0.041*** (5.8422)	0.027*** (4.7171)	0.014** (2.2020)	0.021*** (3.4721)	0.023*** (3.1620)	0.026*** (2.7281)	0.023 (1.4922)
	R-squared	0.1412	0.1080	0.1071	0.1302	0.0981	0.0623	0.0524
	F-statistic	4.3871	3.2368	3.1941	3.9742	2.8960	1.7601	1.4642
	熊市子样本	0.034** (2.3151)	0.026** (1.9852)	0.018 (1.3401)	0.014 (1.1074)	0.014 (0.9572)	-0.015 (-0.9951)	-0.034 (-1.2702)
	R-squared	0.1537	0.2369	0.2346	0.2563	0.2660	0.2461	0.1362
	F-statistic	2.5432	4.3540	4.3114	4.8162	5.0801	4.5730	2.2110

从上表我们发现，换手率指标 $Turn_{t-1}$ 在牛市环境下对下一期日内收益率都没有显著影响，但在熊市环境下对下一期开盘后前 2 个小时内的收益率均有显著的影响和预测作用；而 ADR_{t-1} 指数无论在牛市或者是熊市环境中，对下一期的日内收益率无任何影响；显性情绪指数 $Sin aBSI_t$ 在牛市环境下对下一期开盘后前 2 个小时内的收益率均有显著的预测能力，而在牛市环境下对下一期日内收益率则基本没有影响。

所以，相比之下本文构建的基于网络文本的情绪指标 $Sent_{At}$ 在对下一期日内收益率的预测能力方面很有优势，无论是在牛市环境下还是在熊市环境下，都能对开盘后一定时间内收益率进行预测；只有显性情绪指标 $Sin aBSI_t$ 在牛市环境下有更强的预测能力，本文的 $Sent_{At}$ 在牛市中预测时效为 30min，而 $Sin aBSI_t$ 为 120min 以上。

（三）不同情绪指标对市场波动性影响的对比

同样，先分析不同情绪指标对同期市场波动率的影响，回归建模如下：

$$rv_t = \phi + \sum_{j=1}^s \rho_j * rv_{t-j} + \beta * E_t + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (5.5)$$

其中， E_t 表示情绪指标，同前文滞后阶数 $s=4$ 。

回归结果如下：

表 5.17 不同情绪指标与同期市场波动率的回归结果

模型变量	全样本	牛市子样本	熊市子样本
$Sent_t$	-0.002*** (-5.9181)	-0.002*** (-3.6650)	-0.003*** (-3.8091)
R-squared	0.6357	0.5532	0.5992
F-statistic	73.0292	32.9791	20.8763
$Turn_t$	0.004*** (7.1138)	0.005*** (8.1618)	0.008*** (4.8803)
R-squared	0.6492	0.6301	0.6243
F-statistic	77.3834	45.4968	23.1858
ADR_t	-0.002* (-1.9312)	0.0005 (0.5761)	-0.005*** (-2.6973)
R-squared	0.6064	0.5290	0.5771
F-statistic	64.2874	29.8952	19.0790

注：表中***, **, *分别表示在 1%, 5%, 10% 的显著性水平下显著；圆括号中为系数的 t 值统计量。

根据上表 5.17 回归结果，我们发现 $Sent_t$ ， $Turn_t$ 的回归系数在牛熊市环境下都在 1% 的显著性水平下显著，说明这两个情绪指标都对同期的市场波动率有显著的影响；而 ADR_t 指标的回归系数仅在熊市环境下在 1% 的显著性水平下显著，在牛市环境下并不显著，说明其只在熊市环境下对市场波动率有显著的影响。

其次，本文分析不同情绪指标对市场波动率的预测作用，回归建模如下：

$$rv_t = \phi + \sum_{j=1}^s \rho_j * rv_{t-j} + \beta * E_{t-1} + \delta * g_t + \omega * h_t + \lambda * dchn_t + \varphi * dshibor_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2) \quad (5.6)$$

回归结果如下：

表 5.18 不同情绪指标对市场波动率预测作用的回归结果

模型变量	全样本	牛市子样本	熊市子样本
$Sent_{t-1}$	-0.001** (-2.0231)	0.001** (1.9810)	-0.002*** (-2.8111)
R-squared	0.6062	0.5350	0.5791
F-statistic	64.3892	30.7412	19.2361
$Sent_{At}$	-0.001*** (-3.9572)	-0.0002 (-0.436)	-0.003*** (-3.790)
R-squared	0.6181	0.5282	0.5979
F-statistic	67.6189	(29.8621)	20.8402
$Turn_{t-1}$	0.003*** (4.8762)	0.005*** (6.4281)	0.007*** (3.5519)
R-squared	0.6256	0.5972	0.5933
F-statistic	69.8865	39.5420	20.4074
ADR_{t-1}	0.001 (1.0171)	0.002*** (2.777)	-0.001 (-0.001)
R-squared	0.6032	0.5432	0.5527
F-statistic	63.5352	31.6324	17.3272
$Sin aBSI_t$	-0.007** (-2.1101)	0.002 (0.6402)	-0.016** (-2.3878)
R-squared	0.6073	0.5291	0.5716
F-statistic	64.4901	29.9132	18.6903

注：表中***，**，*分别表示在 1%，5%，10%的显著性水平下显著；圆括号中为系数的 t 值统计量。

以上回归结果表 5.18 显示，在对下一期的市场波动率的预测方面，本文构建的情绪指标 $Sent_{t-1}$ 的回归系数在牛市环境下和熊市环境下分别在 5%和 1%的显著性水平下显著，而 $Sent_{At}$ 也在熊市环境下在 1%的显著性水平下显著；而 $Turn_{t-1}$ 的回归系数无论在牛市还是在熊市环境下都是在 1%的显著性水平下强显著；至于另外两个情绪指标， ADR_{t-1} 的回归系数仅在牛市环境下 1%的水平下显著， $Sin aBSI_t$ 的回归系数仅在熊市环境下 5%的水平下显著。

综合来看， $Turn_{t-1}$ 对未来市场波动率的预测能力更强，本文构建的情绪指标也都能在牛市和熊市环境下对未来的市场波动率进行预测，而 ADR_{t-1} 和 $Sin aBSI_t$ 对市场波动率的预测作用则不理想。

第五节 实证结果分析

根据市场收益率和投资者情绪的同期回归结果,投资者情绪对市场收益率有非常显著的解释作用和正向影响,高涨的投资者情绪意味着较高的市场收益率,低落的投资者情绪意味着较低的市场收益率,尤其是在熊市环境下,这与我们的预期相符。

根据 VAR 模型结果和 Granger 因果检验,无论在熊市还是牛市环境中,市场收益率都是引起投资者情绪变化的原因,在短期内,市场收益率越高,将增强投资者的市场信心,在赚钱效应与羊群效应的影响下促进投资者情绪积聚和高涨,而市场收益率下降,则又会使得投资者情绪迅速回落。中国股市,投资者情绪受当前市场收益率的影响很大。另外,本文也发现我国股市投资者情绪存在较长时间的惯性,无论是乐观情绪还是悲观情绪,都会持续一定时间;但持续不会太长时间,因为研究结果显示,在熊市环境下市场收益率短期内存在惯性,从中长期来看具有反转效应,当收益为正时,投资者会谨慎和担心上涨不会持续,从而情绪又会变得悲观,导致股价又开始下跌。

但是,目前投资者情绪还不是引发市场指数收益率变动的 Granger 原因,尚不能用投资者情绪对未来的市场收益率进行预测和分析,这可能是因为中国的股票市场还是由相对理性的机构投资者主导,中小投资者的力量相对还比较分散,中小投资者的情绪还不足以对未来的市场收益率产生显著的影响。

然而,隔夜开盘前的投资者情绪对当日开盘后短时间内的市场收益率却有显著的正向影响和预测作用。在牛市环境下,开盘前投资者情绪大约对当日开盘后前 30 分钟的收益有显著的影响;在熊市环境下,这种影响的程度更大,时间更长,大约持续一整个上午的交易时间。可能是因为前一日收盘后,股民在晚上有充分的时间搜集、观察、分析市场中的相关信息,并在股吧中宣泄情绪,发表看法,相互交流市场未来的走势,意见比较容易统一,所以在当日开盘后短时间内能对股市产生显著的正向影响;当大家开盘前一致看涨时,就会集体大量买入使得市场指数在短时间内趋于上涨;当大家开盘前一致看跌时,就会纷纷抛售股票导致市场指数在短时间内趋于下跌;但持续时间有限,主要是因为开盘后又会有大量新的信息并对股市产生影响,而开盘前的信息则会“旧信息”而逐渐丧失影响力。

根据市场每日已实现波动率和投资者情绪的同期回归结果,无论是在牛市阶段还是熊市阶段,投资者情绪都对市场收益的波动较很强的解释作用和影响,并且乐观倾向的投资者情绪与悲观倾向的投资者对市场收益的波动性影响存在不对称性,乐观倾向的投资者情绪对市场收益的波动没有显著影响,可能是在市场

收益较好，有赚钱效应，情绪较为乐观时，投资者会选择持有手中的股票，期待进一步的上涨空间，但也会因为担心市场的变化，对于进一步买入会比较谨慎，所以不会产生太大波动；而对市场收益波动产生显著影响的主要是悲观倾向的投资者情绪，无论是在熊市还是牛市，当投资者情绪开始悲观看跌，大家就会快速抛售手里的股票，尤其是在 2015 年这一波行情中，由于加了杠杆，投资者因担心“爆仓”纷纷不计成本盲目抛售股票，一度造成市场的流动性危机，从而大大地加剧了市场的波动性。

根据 VAR 模型结果和 Granger 因果检验，在牛市环境下，由于市场投资者整体情绪较为乐观，市场收益的波动和投资者情绪的相互影响较小，投资者情绪不是引起市场收益波动变化的 Granger 原因，市场收益波动也不是引起投资者情绪变动的 Granger 原因。但在熊市环境下，市场悲观情绪较多，市场收益的波动和投资者情绪的相互影响较为显著，投资者情绪是引起市场收益波动变化的 Granger 原因，市场收益波动也是引起投资者情绪变动的 Granger 原因，当市场收益波动增加时，市场风险也趋于上升，投资者担心被“套牢”，悲观情绪随之上升，纷纷抛售股票避险，从而进一步加剧市场的波动性；反之，当市场产生甚至弥漫悲观情绪时，中小投资者就会非理性的大量卖出，可以预期未来市场的波动性亦会大幅增加。

另外，隔夜开盘前股吧论坛中含有一定量能够预测当日市场波动率的有效信息。根据结果分析，虽然牛市环境下，由于情绪普遍较为乐观，开盘前的投资者情绪指标对当日的市场收益波动没有显著的影响和预测作用，但在熊市环境下，开盘前的悲观投资者情绪对当日市场收益的波动有非常显著的影响和预测作用，其会明显加剧市场的波动性，而乐观倾向的投资者情绪则对市场收益波动没有影响，同样存在不对称性。

此外，本文研究发现，无论是投资者情绪对市场收益率的影响还是投资者情绪对市场收益波动性的影响，都不存在显著的周末效应；说明虽然经过周末或假期，其第二天的投资者情绪指标中比平常的交易日会含有更多的股吧论坛信息，但也并没有对市场收益和波动产生额外的影响。可能一方面，周末或假期市场上有效信息较少，投资者在股吧论坛的交流和发帖也相对较少，所以周末论坛中基本没有影响市场的实质性信息。

最后，根据与传统的不同情绪指标对比发现，本文构建的情绪指标在对同期市场收益率的影响和解释能力方面，无论是在牛市还是熊市环境下，本文从网络文本中提取的情绪指标要明显优于其他传统的显性指标和隐性指标；虽然目前当期的情绪指标，包括传统的情绪指标，都还无法对市场未来的收益率做出预测，但是本文根据开盘前的发帖信息构建的情绪指标能对当日开盘后相对更长时间

内的收益率进行预测。

在对市场波动率的解释作用和预测能力方面,传统情绪指标和基于网络文本的情绪指标影响虽然都很显著,但是只有换手率和本文构建的情绪指标对其的影响较强,也较为有效。

综合来看,相比传统的投资者情绪指标,本文构建的基于网络传媒的投资者情绪指标在对市场收益率和波动率的影响作用以及预测能力上都有很好地表现;只有市场换手率能比本文的情绪指标相对更好地对市场波动率进行解释和预测。

第六章 研究结论与展望

本文通过提取东方财富网从 2014 年 6 月至 2015 年 12 月近 120 万条的帖子信息,构建投资者情绪指标——看涨指数,以此研究网络传媒信息传递的投资者情绪信息与股票市场(以上证指数为例)的相互影响作用。并且本文还借鉴 Pagn 和 Sossounov (2003) 的判定标准将样本区间划分成牛市子样本和熊市子样本,通过线性回归模型,VAR 模型以及格兰杰因果检验,研究不同市态下投资者情绪与股票市场收益及其波动性之间的异化现象。

本文得出的结论是,股吧论坛中投资者发布的大量帖子中,确实包含了能够影响甚至预测市场收益率和波动率的重要信息。

首先,本文构建的看涨指数和股票市场收益率呈现 49.5% 的高度显著正相关,投资者情绪的代理指标看涨指数对市场收益率有非常显著的解释作用和正向影响,尤其是在熊市环境中,影响更为明显;目前市场收益率能够较好地正向影响和预测未来几天的投资者情绪,而当期网络文本中的情绪对未来市场收益率(比如第二天的市场收益率)的影响并不显著,尚无法用来预测和分析未来的市场收益率。但是,隔夜开盘前的股吧论坛中含有大量的信息,能够显著影响和预测当日开盘后短时间内的市场收益率,在牛市环境中大约只能持续约 30 分钟,但在熊市环境中能持续 2 个小时以上,并且影响强度更大。

其次,本文构建的看涨指数和市场收益的已实现波动率也呈现-52.2% 的高度显著负相关,投资者情绪强度都对市场收益的波动有很强的解释作用和影响,并且乐观倾向的投资者情绪与悲观倾向的投资者对市场收益的波动性影响存在不对称性,乐观倾向的投资者情绪对市场收益的波动没有显著影响,对市场收益波动产生显著影响的主要是悲观倾向的投资者情绪。同时,在牛市环境下,市场收益的波动和投资者情绪对相互未来期的影响较小,投资者情绪不是引起市场收益波动变化的 Granger 原因,市场收益波动也不是引起投资者情绪变动的 Granger 原因。而在熊市环境下,市场收益的波动和投资者情绪对相互未来期的影响较为显著,投资者情绪是引起市场收益波动变化的 Granger 原因,市场收益波动也是引起投资者情绪变动的 Granger 原因;悲观的投资者情绪预期将加剧市场的波动性,市场波动也将导致市场悲观情绪上升。此外,隔夜开盘前股吧论坛中含有一定量能够预测当日市场波动率的有效信息,主要在熊市环境下,开盘前的悲观投资者情绪对当日市场收益的波动有非常显著的影响和预测作用,其会明显加剧市场的波动性,而乐观倾向的投资者情绪则对市场收益波动没有影响,同样存在不对称性。

同时,本文研究发现,无论是投资者情绪对市场收益率的影响还是投资者情

绪对市场收益波动性的影响，都不存在显著的周末效应，周末或假期论坛中基本没有影响市场的实质性信息。

最后，本文通过与3个常见的传统投资者情绪指标（包括显性情绪指标和隐性情绪指标）实证对比分析，发现无论在牛市环境还是熊市环境下，基于网络传媒的投资者情绪指标在对市场收益率和波动率的影响作用以及预测能力上优势明显；只有市场换手率能相对更好地对市场波动率进行解释和预测。

本文的实证发现为相关监管部门和投资者判断我国股票市场的一些基本特征提供了一定的参考和基础。在可预见的未来，随着网络技术的快速更新和发展，股民网络参与度的不断提升，以股吧论坛等为代表的互联网信息平台必将成为信息传播的最主要途径，其对股市的影响也将变得越来越不可小觑。我国当前的信息披露制度还很不健全，网络中充斥着各种信息，其中含有大量的虚假信息，再加上我国投资者群体的非理性情绪较高，对股市稳定性将产生重大的威胁，2015年的持续股灾就是最好的例证。未来金融相关监管机构需要充分认识到网络传媒的重要性，并根据网络信息传播的特点，完善金融市场的信息披露制度，对网络信息进行合理的引导与管理。

参考文献

- [1] 董大勇, 2011: 《投资者参与股票论坛的影响因素》, 《系统工程》, 第 01 期, 第 51-56 页。
- [2] 顾娟, 2001: 《中国封闭式基金贴水问题研究》, 《金融研究》, 第 11 期, 第 62-71 页。
- [3] 郭亚维, 刘晓霞, 2012: 《文本分类中信息增益特征选择方法的研究》, 《计算机工程与应用》, 第 27 期, 第 119-122 页。
- [4] 高大良, 2013: 《投资者情绪及其对股票市场收益的影响研究》, 《湖南大学博士学位论文》, 第 17-42 页。
- [5] 黄德龙, 文凤华, 杨晓光, 2009: 《投资者情绪指数及中国股市的实证》, 《系统科学与数学》, 第 01 期, 第 01-13 页。
- [6] 金雪军, 祝宇, 杨晓兰, 2013: 《网络媒体对股票市场的影响——以东方财富网股吧为例的实证研究》, 《新闻与传播研究》, 第 12 期, 第 36-51 页。
- [7] 金晓斌, 高道德, 石建民, 刘红忠, 2001: 《中国封闭式基金折价问题实证研究》, 《中国社会科学》, 第 11 期, 第 55-65 页。
- [8] 蒋玉梅, 王明照, 2010: 《投资者情绪、盈余公告与市场反应》, 《管理科学》, 第 23 期, 第 70-78 页。
- [9] 贾春新, 赵宇, 孙萌, 汪博, 2010: 《投资者有限关注与限售股解禁》, 《金融研究》, 第 11 期, 第 108-122 页。
- [10] 刘超, 韩泽县, 2006: 《投资者情绪和上证综指关系的实证研究》, 《北京理工大学学报(社会科学版)》, 第 08 期, 第 57-60 页。
- [11] 刘仁和, 陈柳钦, 2005: 《股市投资者情绪及其预测》, 《学术交流》, 第 08 期, 第 110-113 页。
- [12] 林树, 俞乔, 2010: 《有限理性、动物精神及市场崩溃: 对情绪波动与交易行为的实验研究》, 《经济研究》, 第 08 期, 第 115-127 页。
- [13] 兰秋军, 马超群, 2011: 《文本感知——金融研究新动态》, 《长沙理工大学学报(社会科学版)》, 第 05 期, 第 37-40 页。
- [14] 饶育蕾, 刘达锋, 2003: 《行为金融学》, 上海, 上海财经大学出版社。
- [15] 饶育蕾, 王攀, 2010: 《媒体关注度对新股表现的影响——来自中国股票市场的证据》, 《财务与金融》, 第 03 期, 第 1-7 页。
- [16] 饶育蕾, 杨琦, 2003: 《我国封闭式基金折价交易的行为金融学实证分析》, 《中南大学学报(社会科学版)》, 第 03 期, 第 346-350 页。
- [17] 饶育蕾, 彭叠峰, 成大超, 2010: 《媒体注意力会引起股票的异常收益吗? ——来自中国股票市场的经验证据》, 《系统工程理论与实践》, 第 03 期, 第 287-297 页。
- [18] 宋泽芳, 李元, 2012: 《投资者情绪与股票特征关系》, 《系统工程理论与实践》, 第 32 期, 第 27-33 页。
- [19] 王美今, 孙建军, 2004: 《中国股市收益、收益波动与投资者情绪》, 《经济研究》, 第 32 期, 第 27-33 页。
- [20] 王春峰, 赵威, 房振明, 2007: 《新股投资者情绪度量及其与新股价格行为关系》, 《系统工程》, 第 07 期, 第 01-06 页。
- [21] 伍燕然, 韩立岩, 2007: 《不完全理性、投资者情绪与封闭式基金之谜》, 《经济研究》, 第 03 期, 第 117-129 页。

- [22] 薛斐, 2005: 《基于情绪的投资者行为研究》, 《复旦大学博士学位论文》, 第 01-203 页。
- [23] 许承明, 宋海林, 2005: 《个人投资者情绪能预测市场收益率吗》, 《经济研究》, 第 03 期, 第 108-118 页。
- [23] 余佩琨, 钟瑞军, 2009: 《中国封闭式基金价格报酬过度波动的经验分析》, 《南开管理评论》, 第 01 期, 第 96-101 页。
- [24] 仪垂林, 王家琪, 2005: 《天气、季节性情绪混乱与股票收益--基于上证综合指数的研究》, 《统计与决策》, 第 06 期, 第 79-82 页。
- [25] 易志高、茅宁, 2009: 《投资者情绪测量实证研究: CICSI 的构建》, 《金融研究》, 第 11 期, 第 174-184 页。
- [26] 张俊喜、张华, 2002: 《解析我国封闭式基金折价之谜》, 《金融研究》, 第 12 期, 第 49-60 页。
- [27] 张永杰, 张维, 金曦, 熊熊, 2011: 《互联网知道的更多么? ——网络开源信息对资产定价的影响》, 《系统工程理论与实践》, 第 04 期, 第 577-586 页。
- [28] 张强, 杨淑娥, 杨红, 2007: 《中国股市投资者情绪与股票收益的实证研究》, 《系统工程》, 第 07 期, 第 13-17 页。
- [29] Antweiler, Werner and Murray Z. Frank, 2004, "Is all that talk just noise? The information content of internet stock message boards", *The Journal of Finance*, Vol.59, PP1259-1294.
- [30] Andersen T G, Bollerslev T., 1998, "Answering the critics: Yes, ARCH models do provide good volatility forecasts", *International Economic Review*, Vol.4, PP885-905.
- [31] Bandi, J.R.Russell and J.Zhu, 2008, "Using High-Frequency Data in Dynamic Portfolio Choice", *Econometric Reviews*, Vol.27, PP163-198.
- [32] Brown, G. and M. Cliff, 2004, "Investor Sentiment and the Near-Term Stock Market", *Journal of Finance*, Vol.11, PP1-27.
- [33] Barberis N., A. Shleifer, R.Vishny, 1998, "A Model of Investor Sentiment", *Journal of Financial Economics*, Vol.49, PP307-343.
- [34] Baker, M. and J. Wurgler, 2006, "Investor Sentiment and the Cross-Section of Stock Returns", *Journal of Finance*, Vol.61, PP1645-1680.
- [35] Brown, G. W. and M. T. Cliff, 2005, "Investor Sentiment and Asset Valuation", *Journal of Business*, Vol.78, PP405-440.
- [36] Brown, G. W., 2006, "Volatility, Sentiment, and Noise Traders", *Financial Analysts Journal*, Vol.55, PP82-90.
- [37] Ben-Rephael A, Kandel S, Wohl A., 2012, "Measuring investor sentiment with mutual fund flows", *Journal of Financial Economics*, Vol.104, PP363-382.
- [38] Brauer G A., 1993, "Investor sentiment" and the closed-end fund puzzle: A 7 percent solution", *Journal of Financial Services Research*, Vol.7, PP199-216.
- [39] Boehme, R., B. Danielsen and S. Sorescu, 2006, "Short-Sale Costs, Differences of Opinion and Over-valuation", *Journal of Financial and Quantitative Analysis*, Vol.41, PP455-487.
- [40] Bollen, Johan, Huina Mao and Xiaojun Zeng, 2011, "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol.2, PP1-8.
- [41] Chen,F., R. Kan and M. Miller, 1993, "Are the Discount on Close-end Funds a Sentiment Index", *Journal of Finance*, Vol.48, PP795-800.

- [42] Clarke R G. Statman M., 1998, "Bullish or bearish?", *Financial Analysts Journal*, Vol.54, PP63-72.
- [43] Delong, J. B., Shleifer, A., Summers, L.H. and R.J. Waldmann, 1990, "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, Vol.98, PP703-738.
- [44] Derrien, F., 2005, "IPO Pricing in Hot Market Conditions: Who Leaves Money on the Table?", *Journal of Finance*, Vol.60, PP487-521.
- [45] Da Zhi, Joseph Engelberg & Pengjie Gao, 2011, "In search of attention", *Journal of Finance*, Vol.66, PP1461-1499.
- [46] Elton, E. J., Gruber, M. J. and J. A. Busse, 1998, "Do Investors Care about Sentiment", *Journal of Business*, Vol.71, PP477-500.
- [47] Edwin J. Elton, Martin J. Gruber, and Jiaping Mei, 1994, "Cost of Capital Using Arbitrage Pricing Theory: A Case Study of Nine New York Utilities", *Financial Markets, Institutions and Instruments*, Vol.3, PP46-68.
- [48] Fisher, K. L. and M. Statman, 2003, "Consumer Confidence and Stock Returns", *Journal of Portfolio Management*, Vol.30, PP115-128.
- [49] Glushkov, D., 2006, "Sentiment Beta", Working Paper, University of Texas.
- [50] Ibbotson, R., J. Sindelar and J. Ritter, 1994, "The Market's Problems with the Pricing of Initial Public Offerings", *Journal of Applied Corporate Finance*, Vol.7, PP66-74.
- [51] Joseph, Kissan, M. Babajide Wintoki & Zelin Zhang, 2011, "Forecasting abnormal stock returns and trading volume using investor sentiment : Evidence from online search", *International Journal of Forecasting*, Vol.27, PP1116-1127.
- [52] Karpoff, Jonathan M., 1987, "The relation between price changes and trading volume: A survey", *Journal of Financial and Quantitative Analysis*, Vol.22, PP109-126.
- [53] Lee, Shleifer and Thaler, 1991, "Investor sentiment and the close-end fund puzzle", *Journal of Finance*, Vol.46, PP75-110.
- [54] Ljungqvist, A. and W. Wilhelm, 2003, "IPO Pricing in the Dot-com Bubble", *Journal of Finance*, Vol.58, PP723-752.
- [55] Ljungqvist, A., V. Nanda and R., Singh, 2006, "Hot Markets, Investor Sentiment, and IPO Pricing", *Journal of Business*, Vol.79, PP1667-1702.
- [56] Lemmon, M. and E. Portniaguina, 2006, "Consumer Confidence and Asset Prices: Some Empirical Evidence", *Review of Financial Studies*, Vol.19, PP1499-1529.
- [57] Laura, Xiaolei Liu, Ann E. Sherman, Yong Zhang, 2008, "The Role of the Media in Initial Public Offerings", Working paper.
- [58] Mehra R., Sah R., 2002, "Mood fluctuations, projection bias, and volatility of equity prices", *Journal of Economic Dynamics and Control*, Vol.26, PP869-887.
- [59] Neal, R. and S. M. Wheatley, 1998, "Do Measures of Investor Sentiment Predict Returns?", *Journal of Financial and Quantitative Analysis*, Vol.33, PP523-547.
- [60] N. Chen, R. Roll, and S. Roll, 1986, "Economic Forces and the Stock Market", *Journal of Business*, Vol.59, PP383-403.
- [61] Pagan A R, Sossonov K A., 2003, "A simple frame work for analyzing bull and bear markets", *Journal of Applied Econometrics*, Vol.18, PP23-46.
- [62] Qiu, L. and I. Welch, 2006, "Investor Sentiment Measures", NBER Working Paper.

- [63] Ritter, J. R., 1984, "The Hot Issue Market of 1980", *Journal of Business*, Vol.57, PP215-241.
- [64] Ritter, J. R., 1991, "The Long-run Performance of Initial Public Offerings", *Journal of Finance*, Vol.42, PP365-394.
- [65] Solt, M E, and M.Statman, 1988, "How Useful is the Sentiment Index", *Financial Analysts Journal*, Vol.44, PP45-55.
- [66] Stephen A. Ross, 1976, "Return, Risk, and Arbitrage," *Risk and Return in Finance*, Cambridge, MA: Ballinger.
- [67] Stigler, G.J., 1964, "Public Regulation of the Securities Markets", *Journal of Business*, Vol.37, PP117-142.
- [68] Tetlock P.C., 2007, "Giving content to investor sentiment: Therole of media in the stock market", *Journal of Finance*, Vol.62, PP1139-1168.
- [69] Wysocki, Peter D., 1999, "Cheap talk on the web; The determinants of postings on stock messageboards", Working paper, University of Michigan.
- [70] Yu J, Yu Y., 2011, "Investor sentiment and the mean-variance relation", *Journal of Financial Economics*, Vol.100, PP367-381.
- [71] Zweig, M., 1973, "An Investor Expectations Stock Price Predictive Model Using Closed-End Fund Premiums", *Journal of Finance*, Vol.28, PP67-87.

致谢

本论文的工作是在我的导师王明涛教授的悉心指导下完成的。从课题的选择到论文的最终完成，王老师都始终给予我细心的指导和不懈的支持。王老师为人随和热情，治学严谨细心，虽身负教学、科研重任，仍多次询问我的研究进程，给我很多宝贵意见，帮助我开拓研究思路，精心点拨、热忱鼓励。王老师一丝不苟的作风，严谨求实的态度，踏踏实实的精神，让我受益终身。王明涛教授严谨的治学态度和科学的工作方法给了我极大地帮助和影响，在此向两年来对我关心和指导的王老师致以崇高的敬意和衷心的感谢。

当初该研究课题因为涉及到技术难度较大的文本挖掘与分类技术，一度对该选题犹豫不决，但是王老师的鼓励和坚决的态度，让我不再动摇，坚定的继续研究。在处理数据期间，遇到过很多的问题，也花了很多的时间和精力，在这里我非常感谢王老师为我推荐的强大技术支持——信管学院的李欣苗教授和她的学生尹涤，他们给了我很大的技术指导和帮助，让我从对文本挖掘技术的一无所知到最终的圆满完成数据处理，在此表示衷心的感谢。

在论文完成之际，我的心情无法平静，从开始进入上海财经大学金融学院学习到论文的顺利完成，有很多可敬的师长、同学、朋友给了我无私的帮助，在这里请接受我诚挚的谢意！当然，也很感恩上海财经大学为我营造的无与伦比的学习氛围，也更加证明当初选择考研上海财经大学是明智的抉择。最后，我还要感谢一直默默支持我的家人，我所取得的每一点成绩和经验都离不开他们的支持！

学业即将完成，我将带着师长、家人的鼓励和期望，迈向人生崭新的阶段！

个人简历及在学期间发表的研究成果

1990年08月22日出生于浙江省宁波市。

2009年9月考入浙江工业大学建筑学专业，2010年9月转专业进入工商管理（技术经济）专业，2013年6月本科毕业并获得管理学学士学位。

2014年9月考研进入上海财经大学金融分析师专业攻读金融硕士学位至今。

在学期间参与金融学院“创新金融人才项目”的课题研究《P2P对小微企业融资支持力度的深度研究—基于对你我贷平台的分析》，没有发表过期刊论文。