

学校代码: 10385

分类号:

研究生学号: 1511406014

密 级:



华侨大学
HUAQIAO UNIVERSITY

硕士专业学位论文

基于投资者情绪的股市隐马尔科夫模型预测

Forecasting the Stock Market With the Hidden Markov
Model Based on Investment Sentiment

作者姓名: 易 轩

指导教师: 胡石清 副教授

实践导师:

专业学位类别/领域: 金融硕士

研究方向: 证券投资

所在学院: 经济与金融学院

论文提交日期: 二零一七年五月十五日

学位论文独创性声明

本人声明兹呈交的学位论文是本人在导师指导下完成的研究成果。论文写作中不包含其他人已经发表或撰写过的研究内容，如参考他人或集体的科研成果，均在论文中以明确的方式说明。本人依法享有和承担由此论文所产生的权利和责任。

论文作者签名： 易柯 签名日期： 2017.5.14

学位论文版权使用授权声明

本人同意授权华侨大学有权保留并向国家机关或机构送交学位论文的复印件和电子版，允许学位论文被查阅和借阅。本人授权华侨大学可以将本学位论文的全部内容或部分内容编入有关数据库进行检索，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。

论文作者签名： 易柯 指导教师签名： 胡石青
签名日期： 2017.5.14 签名日期： 2017.5.14

摘要

行为金融学的研究表明，情绪对投资者的决策有显著的影响，在积极的情绪下，投资者倾向于高估投资机会、低估投资风险，从而交易频繁；在消极的情绪下，则反之。这样的投资行为最终使得股票市场产生了超乎寻常的波动，不利于资本市场的稳健发展。若能从“源头”着手，把投资者的情绪充分运用起来，在研究股票市场的涨跌态势方面将会开创新的思路。本文拟对投资者情绪进行量化，再构建隐马尔科夫模型对股票市场进行预测。

本文首先借用情绪-金融决策模型分析了投资者情绪影响股票市场的两条路径：“情绪-认知-行为”路径与“情绪-行为”路径。然后，从理论层面分析了如何以投资者情绪为基础，运用隐马尔科夫模型对股票市场进行预测。

在实证分析部分，本文首先对投资者情绪进行测量，运用偏最小二乘法（PLS）构建出了综合指标，该指标表明：消费者信心指数、新增股票投资者开户数、股票交易量和换手率与投资者情绪正相关，封闭式基金折价率与投资者情绪负相关。然后，考虑到我国股票市场“政策市”的实际情况，本文对九次利好政策事件、九次利空政策事件进行了研究，结果表明：政策因素对股票市场的影响已经被包含在了投资者情绪之中，运用投资者情绪对股票市场进行预测无需再单独考虑政策因素的影响。接着，本文根据 2007 年 1 月-2015 年 12 月的投资者情绪综合指标数值，计算出了隐马尔科夫模型的关键参数，再根据预测原理运用 MATLAB 软件实现了对股票市场的预测，结果表明：所预测的时间段越长、期数越多，预测结果出现的概率越小。进一步地，为了突出投资者情绪的作用，本文还运用马尔科夫链进行了预测，并对两种模型的预测结果进行了检验和对比，分析表明：隐马尔科夫模型产生的预测结果更加准确，也就是说，投资者情绪是有助于预测股票市场的。

关键词：投资者情绪 隐马尔科夫模型 股市 预测 马尔科夫链

Abstract

Studies of behavioral finance have shown that the sentiment has a significant impact on investors decision-making. When feeling positive emotions, investors tend to overestimate investment opportunities and underestimate investment risks, and thus they trade stocks frequently. When feeling negative emotions, investors behave conversely. This investment behavior eventually led to the stock market fluctuating abnormally, and is not conducive to the stable and steady development of the capital markets. Therefore, if the sentiment could be applied, there are new ideas to study the ups and downs of the stock market. This paper intends to quantify investor sentiment, and then build the hidden Markov model to forecast the stock market.

Firstly, this paper uses the sentiment-financial decision model to analyze the two paths of investor sentiment affecting the stock market: 'sentiment-cognition-behavior' path and 'sentiment-behavior' path. And then, this paper analyzes how to use the hidden Markov model to forecast the stock market based on investor sentiment.

In empirical analysis, this paper first applies the partial least squares(PLS) to construct a comprehensive investor sentiment index. This index presents that Consumer Confidence Index, New Investors Amount, Trading Volume and Turnover rate are positively related to investor sentiment, and the discount rate of closed-end funds is negatively related to investor sentiment.

Then, given that China's stock market is susceptible to the policy, this paper studies nine good and nine bad policy events. The results show that the influence of the policy on the stock market has been reflected in investor sentiment, so there is no longer need to consider the policy when applying investor sentiment to predict the stock market. no longer need to consider the impact of policy factors.

After that, according to the comprehensive investor sentiment index from 2007.01 to 2015.12, this paper calculates key parameters of the hidden Markov model,

and predicts the stock market based on the forecasting principle using MATLAB software. The results show that the longer the predicted time period, the smaller the probability of the outcome.

Furthermore, in order to highlight the role of investor sentiment, this paper also uses the Markov chain to predict the stock market, and tests and compares results of these two kinds of two models. The analysis shows that the result of the Hidden Markov model is more accurate. That is to say investor sentiment is effective in predicting the stock market.

Key Words: investment sentiment the Hidden Markov mode the stock
market predict the Markov chain

目录

第 1 章绪论	1
1.1 研究背景	1
1.2 研究意义	2
1.3 文献综述	4
1.3.1 投资者情绪对股市的影响	4
1.3.2 隐马尔科夫模型（HMM）与股市预测	5
1.4 研究方法及技术路线	6
1.4.1 研究方法	6
1.4.2 技术路线	8
1.5 论文篇章安排	9
1.6 创新点与不足之处	9
1.6.1 论文的创新点	9
1.6.2 论文的不足之处	9
第 2 章 相关理论基础	11
2.1 投资者情绪的理论基础	11
2.1.1 投资者情绪的定义	11
2.1.2 投资者情绪的测量指标	11
2.1.3 投资者情绪的理论解释	15
2.2 隐马尔科夫模型（HMM）理论	18
2.2.1 马尔科夫链	18
2.2.2 隐马尔科夫模型	19
第 3 章 基于投资者情绪的股市 HMM 预测的理论分析	23
3.1 投资者情绪影响股市的理论分析	23
3.1.1 情绪-金融决策模型	23
3.1.2 投资者情绪与股市	23
3.2 隐马尔科夫模型预测股市的理论分析	24
3.2.1 投资者情绪与隐马尔科夫模型	24
3.2.2 预测原理分析	25

第 4 章 基于投资者情绪的股市 HMM 预测的实证分析	27
4.1 构建投资者情绪综合指标	27
4.1.1 偏最小二乘法介绍	27
4.1.2 单个指标的选择	28
4.1.3 综合指标的构建	29
4.2 我国“政策市”对投资者情绪的影响	31
4.2.1 提出假设	31
4.2.2 数据来源	32
4.2.3 事件研究法与实施步骤	33
4.2.4 实证结果与分析	34
4.3 隐马尔科夫模型的预测	39
4.3.1 投资者情绪状态转移概率矩阵	39
4.3.2 观测概率矩阵	41
4.3.3 隐马尔科夫模型的 MATLAB 函数介绍	43
4.3.4 基于投资者情绪的隐马尔科夫模型预测实现	44
4.4 隐马尔科夫模型预测与马尔科夫链预测的对比分析	46
4.4.1 马尔科夫链预测的实证分析	46
4.4.2 预测结果的检验与对比	49
4.5 实证结论	51
第 5 章 结论与展望	53
5.1 研究结论	53
5.2 研究展望	53
参考文献	55
致谢	59

第 1 章 绪论

1.1 研究背景

2015 年前三个季度，我国股市的表现特别不合理，不仅上演了千股跌停、千股停牌、千股涨停的奇观，还多次刷新了最大单日涨、跌幅纪录，沪深两市成交量更是以 2.03 万亿元改写了全球股票现货市场单日成交纪录。就上证综指而言，从 2014 年 11 月的 2400 点上方大涨至 2015 年 6 月 12 日，在击穿 5178 点高位后大幅回落，并于 8 月 26 日最低触及 2850.71 点，九个月内波动幅度高达 53.65%。这不禁让人想起上一次的股灾，就在 07、08 年金融危机期间，上证综指从 6100 多点暴跌到 1600 多点，如此巨幅波动令人心惊肉跳。我国股市的异常之处可在这十年间的多次暴涨暴跌之中窥探一二。除了股指的波动，市盈率也能够体现股票市场是否合理。截至 2015 年 12 月末，我国股市中小板、创业板的平均市盈率分别高达 80.67 倍、103.09 倍，这些成长股、热点股“高烧不退”体现的无非是股民想乘机分得一杯羹的心理，这也就表明了我国股市存在的过度投机。

当透过这些数据，本文认为，我国股市表现出来的不合理究其原因在于投资者的非理性行为。从 2007 年至今，我国股票市场的投资者一直以个人投资者为主。根据中国证券登记结算有限责任公司编制的统计月报，截至 2016 年 8 月末，在我国证券市场上，自然人投资者数是 11,148.96 万，占比高达 99.72%。与机构投资者相比，个人投资者的投资决策总体而言是不够专业、不够客观的，是容易受到主观意识影响的，其交易行为往往具有过于频繁、跟风、投机性强、缺乏进行长期价值投资的信念等非理性特点，对股票市场的异常波动起到了推波助澜的作用。

在这种现实情况下，显然，传统金融理论对投资者的“理性人假设”是不恰当的，打破这种假设的是行为金融学。从上世纪 80 年代起，越来越多的金融学家将心理学的研究成果应用于对投资者的行为分析，他们以投资者的实际心理为出发点讨论其决策对市场的影响，在这方面贡献了大量高质量的理论和实证文献，使得行为金融学兴起并蓬勃发展。其中，马修·拉宾（Matthew Rabin）荣获 2001 年克拉克奖，丹尼尔·卡尼曼（Daniel Kahneman）和弗农·史密斯（Vernon Smith）荣获 2002 年诺贝尔经济学奖，这些代表人物的获奖表明行为金融学是受

到肯定的。由于行为金融学派认为人是非理性的，因此他们非常注重于研究这种非理性是如何表现的、对金融市场会产生什么影响，逐渐地，“投资者情绪（Investment Sentiment）”这一概念被提了出来，并随着行为金融学的发展，投资这情绪成为了该学科解释股票市场异常表现的基础之一。因此，本文把投资者情绪和我国股票市场的上证综指联系起来进行研究。

根据行为金融学家的研究，投资者的非理性有很多表现，比如过度自信、后悔厌恶、羊群效应等等，然而这些“情绪”是无法在股票市场观察到的，只能通过投资者的决策使得某些数据变动了来反映。这个过程，恰好可以运用隐马尔科夫模型（Hidden Markov Model, HMM）进行模拟。隐马尔科夫模型是一种统计分析模型，创立于 20 世纪 70 年代，其关键参数可以把隐藏起来的状态和直接观测变量连接起来。与此同时，该模型也可以对状态的转变，以及状态与观测变量的对应进行识别，从而实现学习和预测。由于具有强大的理论基础和非常好的可处理性，HMM 在 80 年代得到了传播和发展，成为信号处理的一个重要手段。如今，该模型被更加广泛地运用，所涉及的领域包括语音识别、生物序列分析（DNA、蛋白质等）、股市预测、地震预报等。在本文中，运用隐马尔科夫模型的关键在于把股市上不可观测的投资者情绪作为模型的隐状态，使其成为模型的参数之一。这种尝试将为股票市场数据的预测研究开阔新的思路，提供新的理论和技术支持。

1.2 研究意义

自 1990 年底上海证券交易所成立至今，我国股票市场才发展了二十多年的时间，在很多方面还不够完善、成熟，比如说目前存在着的过度投机。为什么会普遍出现这种急功近利的心理，本文认为是股市中大量的个人投资者的非理性行为造成的。一般地，股市上的投资者分为机构投资者与个人投资者。机构投资者由证券从业人员组成，具有专业的知识结构和分析能力，也具有雄厚的资金实力，甚至有可能获得内幕消息。而个人投资者，俗称“散户”，由于受到自身各方面条件的限制，在进行股票交易前，一般做不到如机构投资者那般的专业性分析，而且很有可能是听信传言，或者受到情绪影响冲动地买卖了股票。在我国，目前个人投资者是主流大军，而且在未来较长一段时期内仍然会占大多数，他们的非理性决策行为对股市的稳定发展已经产生了负面影响。因

此，仅仅借助以“理性人”为假设的传统金融理论来分析我国股市的现状和未来显然是不足够的，新兴的行为金融学进入了学者的视野。行为金融学以投资者的实际心理为出发点讨论其决策对股市的影响，值得对其进行充分地检验和发展，并将其应用到我国的证券市场。

目前国内对投资者情绪的研究主要集中在构建投资者情绪测度指标和投资者情绪在市场层面、对公司行为的影响。其中研究投资者情绪是否可以预测股票收益率的文章比较多，但是真正利用投资者情绪对股票市场进行预测的文章比较少，仅有的相关文献与本文的预测对象和预测方法也不相同。因此本文从投资者情绪出发对上证综指收益率进行预测，有助于丰富和完善对投资者情绪研究的相关内容；运用隐马尔科夫模型也有利于股票市场预测方法的进一步创新。

在现实意义上，一方面，发行股票是企业进行直接融资的重要方式。我国股市建立近三十年来，A股上市公司已超过3000家，总市值在全世界范围内名列前茅，是支持我国经济发展的中坚力量；在中小板和创业板的上市公司中，80%以上的是民营企业，更有数以万计的公司在新三板挂牌，其中有大量代表着创新、代表着前沿科技的新兴企业，他们在国家经济发展的过程中释放着更多的活力，在提升自主创新能力的路上进行着更加大胆的尝试。另一方面，股票是一种大众投资工具。在我国股市，个人投资者占了绝大多数，股票扩大了他们投资的选择范围，而且股票投资的流动性和灵活性比较强，投资者随时可以比较方便地将股票出售变现、收回投资资金。一般来说，通过购买股票可以获得较高收益、实现资本的增值。但是普通股民由于风险意识淡薄、非理性行为明显，往往被机构投资者的杠杆资金和假消息利用，最后被套牢、亏得血本无归，进而可能导致严重的社会问题。

因此，对股市进行适当的预测是无可非议、意义深远的，尤其是对股票价格指数。股票价格指数不仅是一个数字，表示着这个股票市场的总体价格水平以及走势情况，与此同时，透过股票价格指数，也可以看出其代表的国家或地区的政治是否稳定、经济状况是否良好。上证综指作为中国股票市场上的第一只指数，是国内最具影响力的指数，是国外主流媒体认可的最能代表中国股市的指数，也是股民最熟悉、最关心的指数。对上证综指相关数据进行预测，这不仅可能为企业提供一个相对合适的上市融资的时间点，还可能帮助投资者更好地了解股票市场的发展情况、做出更有利的投资决策，更在为政府及时有效

地对股市进行管理、保持股市健康发展、预防系统性风险等方面具有实际参考价值。

1.3 文献综述

1.3.1 投资者情绪对股市的影响

研究投资者情绪对股市的影响的国内外文献非常多，主要从影响股市的波动、收益率这两个角度进行。考虑到后文的预测对象是股指收益率，本文拟对较新的、影响较大的相关文献进行综述，并分成投资者情绪对股市收益是否有影响、投资者情绪能否预测股市收益这两个部分。

(1) 投资者情绪对股市收益是否有影响

Brown、Cliff（2004）运用 VAR 模型检验个人、机构投资者情绪对股市整体收益的影响，结果表明二者都与大盘股的当期收益具有正相关关系。Chuang 等人（2010）运用 GARCH-M 模型研究台湾股市中的投资者情绪对超额收益的影响，实证结果表明二者具有显著的负相关关系。

黄德龙等人（2009）利用异方差模型实证检验了情绪指标与深、沪市 A 股多个指数的当期收益的关系，得出情绪指标的正向变动对当期收益有正向影响的结论。池丽旭、庄新田（2011）将投资者情绪分为乐观和悲观，应用面板数据模型研究情绪指标对股票收益的影响，实证结果表明，相较于悲观情绪，乐观情绪对股票收益有更明显的影响。张宗新、王海亮（2013）应用多元回归法和脉冲响应函数检验情绪指标对市场收益率的影响，实证结果表明，投资者情绪对上证综指收益率具有较强的正向推动作用。文凤华等人（2014）将投资者情绪分为积极与消极，并采用多种模型分析情绪指标对上证综指收益率的非对称影响，结果表明，积极情绪、情绪的正向变动对收益率有显著的正向影响，但消极情绪、情绪的负向变动对其影响不明显。

(2) 投资者情绪能否预测股市收益

Fisher、Statman（2000）将投资者分为小、中、大型三等，运用多元线性回归法研究各投资者情绪指标与 S&P 500 未来收益的关系，实证结果表示，小型、大型投资者的情绪与 S&P 500 未来收益呈显著负相关关系，两个指标具有预测性。Lemmom、Portniaguina（2006）以消费者信心指数为情绪指标，其研究发现该指标对小盘股的收益具有预测能力。

张丹、廖士光（2009）采用因果检验方法研究前一期的机构、个人投资者情绪对当期股市收益的影响，结果表明两种情绪都对未来市场的收益具有较强的预测能力。池丽旭、庄新田（2011）的结论还表示极端的投资者情绪具有独特的预测能力，池丽旭等（2012）利用向量自回归模型分析了情绪指标与深、沪市指数收益率的经验关系，实证结果表明，情绪指标的变化量更具预测性。刘伟奇、刘新新（2014）利用普通最小二乘回归模型分别考察个人、机构投资者对上证 A 股市场收益率的影响，研究表明，机构投资者更理性，他们的情绪可以预测后市，而个人投资者情绪不具有预测性。鹿坪等人（2015）分别构建了机构、个人投资者情绪指数，并基于向量自回归模型分析其对深沪市股指收益的影响，研究结果表明，机构、个人投资者情绪都对股指收益有显著影响，并具有一定的预测力。

通过对相关文献的梳理，本文认为投资者情绪对股市收益有显著影响这一结论是可以肯定的；至于投资者情绪是否具有预测性，目前尚无定论，本文认为这是由各学者构建的指标不同、选取的样本不同以及运用的实证方法不同导致的。另外，本文发现国内学者的研究脚步虽然慢于国外学者，但目前的趋势是对情绪、投资者以及研究对象进行细分来研究，而且检验投资者情绪对股市收益影响的方法多样。本文认为机构投资者亦是个人，亦受到情绪影响，并且我国股市中个人投资者占多数，故在实证环节不对投资者进行分类。但是本文拟对情绪进行分类。

1.3.2 隐马尔科夫模型（HMM）与股市预测

国内运用隐马尔科夫模型研究股市的期刊文献比较少，苗苗等人（2010）采用深交所 1998-2008 年的月度股价指数研究股市的状态转换，他们首先将原始数据处理成 ST 序列，然后根据股市的熊市、疯牛市和慢牛市三种状态对 ST 序列进行状态分类从而得到 HMM 的状态序列，最后通过计算得到股指状态的转移矩阵。他们的研究表明，运用这种非线性状态法比传统计量方法更能较好地对股市进行分析和预测。胡淑兰等人（2011）的研究对象和研究方法几乎与苗苗等人一致，区别在于他们采用的是上交所的月度股价指数，以及对股市状态的划分更加细致，他们在考虑了我国当期 GDP 的增长率和一年定期存款利率的情况下，把股市划分为大（疯）牛市、成长市、震荡市、弱熊市和大（疯）熊市。

相对于期刊文献,运用隐马尔科夫模型对股市进行预测的硕博士论文数量多一点。吴漫君(2011)应用隐马尔科夫模型对日内股价走势进行预测,她把涨和跌(即隐状态)取为股价的未来走势,观察值为四个股价指标。当模型在离散条件下,她把观察值离散化为涨与跌的16种组合情况,再通过四位数二进制编码得到最终观察值;而在连续条件下,观察值就是股价数值。其研究结论表明,连续HMM的预测平均准确率高于离散HMM的,但二者都没有体现模型的优势,也就是说,隐马尔科夫模型在短期市场中的应用效果不佳。张珍

(2012)应用隐马尔科夫模型预测S&P500的日波动率,她以收益率为观察值,波动率则由包含了五种状态的马尔科夫过程驱动,从而通过历史数据对模型参数进行估计,再利用参数和当前观察值进行一步向前波动率预测。同时,她还对比了隐马尔科夫模型、GARCH模型和T-GARCH模型,最后发现隐马尔科夫模型在对短期的波动率进行预测时更有效。黄冉(2015)以股票价格的开盘价、收盘价、最高价、最低价四组数据作为观察序列,通过BW算法进行参数估计、Viterbi算法进行模式识别,进而预测未来一天的收盘价。进一步地,她对原始数据进行小波去噪处理,并运用15日加权预测法,从而实现了模型的改进和优化。其实证结果表明,运用改进的HMM对股票价格进行预测的精度相对较高。

本文认为,这些硕士论文基本上是对高频数据来运用连续隐马尔科夫模型,技术含量比较高,通过阅读受益匪浅。但是考虑到自身情况,包括数据的可获取性以及投资者情绪的联系,本文在预测上主要是对隐马尔科夫模型的原理进行直接地运用。

1.4 研究方法及技术路线

1.4.1 研究方法

(1) 定性、定量分析相结合

本文首先在第3章从理论层面分析投资者情绪是如何影响股票市场的,以及如何运用隐马尔科夫模型预测股票市场,然后在第4章进行实证分析,以我国股票沪市2007年1月至2015年12月的相关数据为样本,运用偏最小二乘法构建投资者情绪综合指标、得到情绪数值,再以此计算出隐马尔科夫模型的参

数，最后运用该模型对沪市 2016 年涨跌情况进行预测。

（2）偏最小二乘法

偏最小二乘法是一种新型的多元统计数据分析方法，它可以同时解决三个问题：通过多元线性回归进行建模，通过主成分分析简化数据结构，以及对两组变量进行相关性分析。目前构建投资者情绪综合指标最常用方法是主成分分析法，但是该方法的缺陷十分明显：在根据相关程度选择主成分的时候，遗漏重要变量的风险很大，这将直接降低模型的可信程度。

因此，为提高情绪的测量精度，本文选择偏最小二乘法。偏最小二乘法在对两个变量都进行分解的基础上，提取共同的成分作为因子，再分析因子之间的相关程度，最后选择合适的因子进行建模，即该方法可以从情绪代理变量中尽可能地提取与投资者情绪相关的部分。

（3）隐马尔科夫模型预测法

目前可用于对股票市场进行预测的方法有很多种，包括基于统计原理的 GARCH 模型、LOGIT 模型等；人工智能预测模型，如 BP 神经网络；基于随机过程的马尔科夫链、隐马尔科夫模型等，本文选择的是隐马尔科夫模型。

隐马尔科夫模型是从马尔科夫链发展而来，该模型由具有一定隐状态数的马尔科夫链和显性随机函数集构成。其中，马尔科夫链表示隐状态序列，模型的关键参数转移概率矩阵可以表示状态之间的转移情况。在隐状态和显性观测值之间存在一一对应的关系，这种关系可以用观测概率转移矩阵表示。

与其他模型相比，一方面，隐马尔科夫模型通俗易懂，它以数据模式识别为原理，运用获得的历史数据计算初始参数、建立数学模型，然后通过不断的迭代，得到序列数据随着时间变化而形成的统计规律，预测未来的数据。另一方面，该模型中的参数隐状态序列，可以表示表面上看不到、实际上发挥重要作用的影响因素，充满研究想象力。如本文所考虑的，将投资者情绪视为股票市场上隐藏的因素，将其运用到隐马尔科夫模型当中，作为隐状态序列，并以此为基础对我国股市进行预测，开创了新思路。

同时，为进一步明确投资者情绪在预测方面所发挥的作用，本文设置了对比分析，在去掉投资者情绪这一参数的情况下，运用马尔科夫链来对股市进行预测，再比较两种情况下的预测准确度。

1.4.2 技术路线

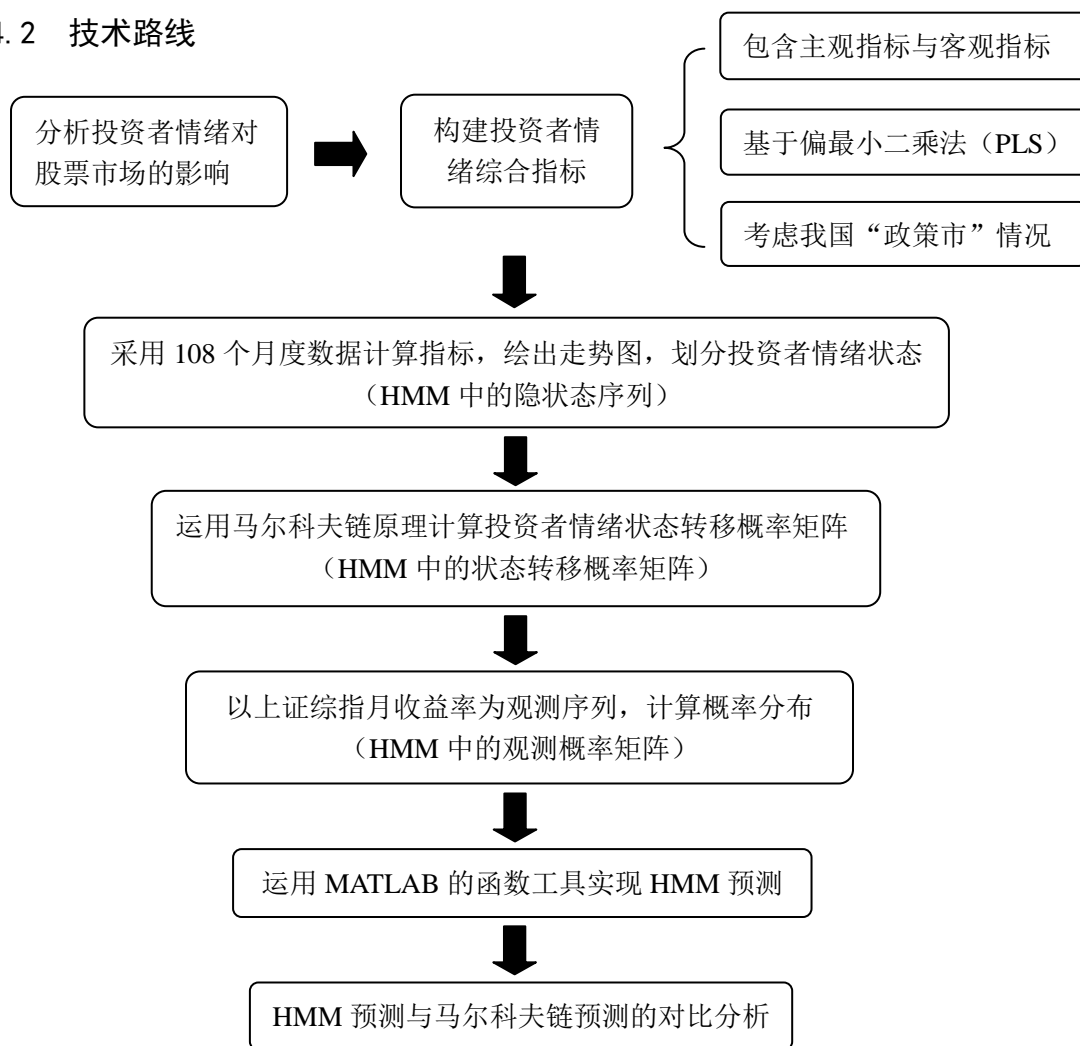


图 1.1 技术路线图

1.5 论文篇章安排

第 1 章绪论。本章首先表明了论文的研究背景和研究意义，然后对相关文献进行了介绍、评述，接着说明了论文拟用的研究方法和技术路线，并对论文进行了篇章安排，最后分析了论文可能的创新点与不足之处。

第 2 章相关理论基础。本章从投资者情绪和隐马尔科夫模型（HMM）两方面介绍相关理论。

第 3 章基于投资者情绪的股市 HMM 预测的理论分析。结合投资者情绪的理论基础分析其对股票市场的影响，并着重分析了投资者情绪与隐马尔科夫模

型的联系、构建基于投资者情绪的隐马尔科夫模型的思路，以及预测原理。

第 4 章基于投资者情绪的股市 HMM 预测的实证分析。本章在考虑了中国“政策市”的情况下，运用偏最小二乘法构建投资者情绪综合指标，并将其作为隐马尔科夫模型的参数之一，再运用该模型对股市涨跌情况进行预测，最后对比了在有投资者情绪这两种情况下的预测结果。

第 5 章结论与展望。撰写研究结论，并提出可进一步研究之处。

1.6 创新点与不足之处

1.6.1 论文的创新点

选题立足于当下，以 2015 年前三个季度中国股市，尤其是沪市的大幅起落为研究背景，对投资者行为和情绪展开思考，使得文章兼具时效性和学术性。并且本文继续深入，以投资者情绪的测量结果为基础，运用隐马尔科夫模型对上证综指月收益率进行预测，使得文章内容更加丰富、深度有所体现，使得研究结论相对而言具有实际参考价值。

本文的创新之处在于将投资者情绪的测量结果作为隐马尔科夫模型的初始参数之一，并运用隐马尔科夫模型对股市进行预测，丰富了相关研究。同时，设置了对比分析，在去掉投资者情绪这一参数的情况下，运用马尔科夫链来对股市进行预测，再比较两种情况下的预测准确度，从而明确了投资者情绪在预测股市方面的作用。

1.6.2 论文的不足之处

本文尝试基于投资者情绪建立隐马尔科夫模型，以此对我国股市的涨跌进行预测，但在研究过程中受到一些因素的限制，无法尽善尽美，存在不足：

（1）本文对投资者情绪是如何影响股票市场的、隐马尔科夫模型是如何预测股票市场的理论分析不够深入、不够透彻。

（2）本文对投资者情绪进行的量化存在偏差，无法完全准确地反映投资者的真实情绪。

（3）本文考虑到了政策因素对投资者情绪的影响、对预测股市的影响，但运用事件研究法的实证分析不够严谨。

第2章 相关理论基础

2.1 投资者情绪的理论基础

2.1.1 投资者情绪的定义

投资者情绪（investment sentiment），或者投资者心态，其前身是指非理性的噪音交易者对风险资产未来收益的错误信念，并且这种信念具有不可预测性。Shleifer、Summers（1990）明确使用投资者情绪（investment sentiment）这一名词，但并未给出详细定义，只表示其无法完全被基本面消息证明是合理的。Barberis 等人（1998）认为投资者情绪是投资者形成投资决策理念的过程，并基于心理学的表征性启发和守旧性构建模型进行分析。Brown、Cliff（2004）认为情绪代表的是投资者相对于平均收益水平的预期，即看涨（看跌）的投资者预期收益将高于（低于）平均收益。Baker、Wurgler（2007）将投资者情绪定义为关于资产未来现金流和投资风险的、无法被现有事实证明合理的信念。

在国内学者中，饶育蕾、刘达峰（2003）把投资者情绪定义为对未来带有带有系统性偏差的预期。王美今、孙建军（2004）认为因为心理上或认知上的偏差，投资者会做出“基于情感的判断”，即投资者情绪，这会影响他们对风险资产未来收益分布的正确认识。黄德龙等人（2009）给出了比较具体的定义，他们认为投资者情绪不仅包含了对实体经济的预期，即对上市公司盈利前景的预期，还包含了对虚拟经济的预期，即对投机性需求的满足。

由此可知，对投资者情绪的定义，目前没有一个统一的表述，因为它实在是让人感觉“只可意会不可言传”。但是不难看出，大多数学者都认为投资者情绪是非常主观的，是投资者的一种预期，即“我认为它将...”，从而影响了投资者的投资行为。本文比较赞同王美今、孙建军（2004）的看法，故本文引用他们对投资者情绪的定义。

2.1.2 投资者情绪的测量指标

在投资者情绪的存在被认同之后，学者们开始研究它是如何表现的、该如何捕捉和测量。在三十多年的研究中，投资者情绪的测量指标经历了一个间接代理变量、直接调查指标和综合指标的过程。

（1）间接代理指标

间接代理指标指的是由金融市场的交易数据构建而成，从侧面间接地表示投资者情绪。

① 封闭式基金折价率

封闭式基金折价率指的是每份封闭式基金份额代表的转让价格，低于平均每份基金持有资产的市场价格的比例，这种贴水转让的现象被称为“封闭式基金之谜”。

Zweig (1973) 指出，封闭式基金的贴水转让反映的是投资者的预期。De long 等人 (1990) 将构建的 DSSW 模型用于解释该现象：当投资者对封闭式基金收益看涨时，基金以升水或小幅贴水的价格卖出；当看跌时基金卖出的贴水幅度将增大。同时，即使不假设噪音投资者是看涨或看跌的，他们对封闭式基金预期收益变化的难以预测（即上文所提的投资者情绪的前身）造成了折价，以及折价率的波动。Lee 等人 (1991) 进一步地运用投资者情绪解释了封闭式基金之谜，他们表示折价率的随机变动反映的是个人投资者不同的情绪，并且认为该情绪是系统性的，不同基金的折价率会同步变化；而且投资者情绪也会影响到与封闭式基金没有任何关系的其他资产。他们的实证结果支持了这一理论结论。之后，有部分学者质疑、反对他们的结论，但都无法撼动这篇文献的重要地位。

国内学者是从 2000 年开始探讨封闭式基金折价率能否反映投资者情绪的，即投资者情绪能否解释我国封闭式基金之谜。他们的研究方法经历了一个由生搬硬套 Lee 等人的分析思路，到注重我国市场的实际情况并进行深入研究的过程。他们得到的结论也从最初的不确定或者与 Lee 等人的结论相反，到即使考虑我国封闭式基金市场的特殊性，结论也与 Lee 等人的结论相符，即封闭式基金折价率可以作为投资者情绪的衡量指标。在众多文献中，本文认为伍燕然、韩立岩 (2007) 的研究是较严谨、全面、深刻的。他们首先运用数据检验了封闭式基金价格的过度波动；然后用国内 1997 年 4 月到 2005 年 12 月的数据简单套用 Lee 等人的二元回归模型，如此结果却表明两个自变量存在共线性；于是作者构建了自己的一元回归模型，其稳健性检验分析结果支持 Lee 等人的结论，同时作者考虑到当时我国封闭式基金主要由机构投资者持有这一现实，对实证结果做出了新的解释：封闭式基金投资者整体受情绪影响，并且小公司股票较大公司的而言，更容易受投资者情绪影响，他们对此也进行了论证；进一步地，作者用两个新指标（IPO 首日收益率和月新开户数）从侧面证明了封闭式基金折

价率是投资者情绪指标。随后的相关文献，如杨元泽(2010)和李静、李杰(2012)，他们实证分析也证明封闭式基金折价率可衡量投资者情绪。

② IPO 数量及 IPO 首日收益率

IPO 是指普通股的首次公开发行，多年数据表明新股上市首日的收盘价大幅高于发行价，即首日收益率非常高，这在过去是无法解释的谜团。学者们通过研究投资者情绪能否解释该现象，来证明 IPO 数量和 IPO 首日收益率是否为投资者情绪指标。

Ritter (1991) 的研究表明，IPO 首日收益率之所以过高是因为投资者对新股保持着过度乐观的情绪。Pagano 等 (1998) 对意大利米兰交易所 11 年 IPO 的发行量进行了研究，结果发现公司发行股票的时机大多选在市场情绪高涨的时期。Lowry (2002) 认为公司对资本的需求和投资者情绪是 IPO 数量剧烈波动的原因。Dorn (2003) 的研究表明，投资者过度乐观是发行市场火爆的最主要原因。Ljungqvist 等人 (2006) 认为，IPO 价格被投资者的乐观情绪推动着，出现了高于其内在价值的情况，但是长期来看公司收益率比较低。

韩立岩、伍燕然 (2007) 开始从投资者情绪角度解释 IPO 之谜。他们不仅用噪音交易者现象和投资者的乐观偏差、锚定心理解释了国内新兴市场 IPO 首日收益率超高的现象，而对于某阶段 IPO 数量增加的原因，作者认为是监管部门在调节发行节奏；还利用数据验证与换手率、月新开户数指标的关系，说明新股首日收益率也反映了投资者情绪。邵新建 (2010) 以此为基础，利用 1993 年到 2008 年的 A 股市场 IPO 月度数据，证明了在我国实行核准制、无卖空交易制度的背景下，IPO 首日与 20 日（初期）收益率含有投资者情绪因素，且前期 IPO 初期收益率与后期 IPO 数量呈正相关关系。

但是本文认为，我国的新股发行基本上被监管当局控制着，而且具有不连续性，相关数据缺失严重，故 IPO 数量、首日收益率在通过实证分析来构建投资者情绪指标时，并不是合格的变量，部分国内学者也确实没有采用这两个指标。

③ 换手率

换手率指的是当期成交股数与市场流通总股数之比，它从流动性的角度、在一定程度上体现出投资者情绪高涨或低迷。

Baker、Stein (2003) 总结了之前对市场流动性指标与股票收益之间关系的研究，并以此为基础，从投资者情绪的角度研究了换手率。利用 1927-1998 年间

纽约证券交易所的交易换手率年数据，他们的实证结果表明：换手率与价值加权资产组合、等权重资产组合的下一期收益都呈负相关关系，故换手率可以是一个情绪指标。也就是说，当非理性投资者情绪乐观时，换手率变高、市场流动性增强，从而导致股票估值过高，进而导致未来收益降低。韩立岩、伍燕然（2007）认为高换手率意味着投资者的过度投机。邵建新（2010）也选用了换手率作为投资者情绪的代理指标。

事实上，以市场流动性变量，如交易量、换手率，作为投资者情绪的指标是最直接、最被经验支持的，因为在大多数情形下，投资者情绪的高涨或低迷总是与市场流动性密切相关。

④ 投资者新开户数

该指标目前只有国内学者在使用。伍燕然、韩立岩（2007）提出，交易所月新开户数代表场外投资者对证券的需求，能更直观的反映情绪。当情绪高涨，投资者进入市场的热情就高，故月新开户数就高；反之则反。随后，多数学者纷纷引用该指标，把其纳入综合指标的构建中。

（2）直接指标

直接指标主要通过问卷的形式直接调查投资者对未来市场行情的预期，一般分为看涨、看跌和看平三种情况，可直接反映投资者的态度。在早期，投资者智慧指数和美国个体投资者协会指数为国外学者所用，好淡指数和“央视看盘”数据指标为国内学者所用。目前，消费者信心指数（CCI）的使用最广泛。

Fisher、Statman（2000）对会议委员会消费者信心指数和纳斯达克收益、小公司股票收益进行相关性分析，结果表明指数与收益之间存在显著的负相关关系。Qiu、Welch（2004）比较了封闭式基金折价率与消费者信心指数，认为后者更适合作为投资者情绪的直接测量指标。薛斐（2005）在分析国内数据后得到了与前二位国外学者相同的结论。王汝芳、田业钧（2009）利用2001年至2007年的消费者信心指数月度数据，直接考察其与沪市收益率的关系，得出二者具有较强的协同性这一结论。

本文认为，直接指标虽然是投资者当下情绪最真实的反映，但是却具有易变的特点，投资者未来发生的实际行为可能与接受调研时的涨跌想法不一致，这样就造成了运用投资者情绪直接指标进行研究在一开始就产生了偏误，因此只靠直接指标反映投资者情绪明显是不够的。

（3）综合指标

投资者情绪受多方面的影响，单一的间接、直接指标无法对其进行全面测量，所以学者们逐渐转向对投资者情绪综合指标的构建。目前，国内外影响相对较大、较新的综合指标包括以下三个。

① BW 指数

目前认可度最高的的综合指标是 Baker、Wurgler（2006）构建的 BW 指数。构建该指数运用的是主成分分析法，提取封闭基金折价、IPO 数量和上市首日收益、交易量、股利收益和股票证券发行量之比的第一主成分，然后为消除宏观层面的影响，作者将第一主成分与 GDP、工业生产指数等进行了回归分析，最后取残差构建了 BW 指数。

② CICS I 指数

易志高、茅宁（2009）受到 BW 指数的启发，结合国内的实际情况，同样采用主成分分析法，在提取了第一至五主成分的加权平均值、控制了宏观经济变量的同时，更考虑到不同的指标对投资者情绪的反应可能存在时间差，最后采取交易量的滞后变量、封闭基金折价率、IPO 数量及上市首日收益、消费者信心指数、新增投资者开户数滞后变量六个主客观相结合的指标。

③ CIS^{PLS} 指数

这是王镇、郝刚（2014）针对 CICS I 指数、运用偏最小二乘法（PLS）构建的指数。他们选取了与易、茂文中相同的单个指标：封闭基金折价率、交易量、IPO 数量的滞后变量、上市首日收益、消费者信心指数、新增投资者开户数，所构建的指标通过了稳健性检验，并且他们通过严谨的比较分析发现，采用 PLS 法构建的投资者情绪指数的效果确实比 CICS I 指数更优。

综上，虽然 CIS^{PLS} 指数仍然存在一些缺陷，如构建时过于强调与 CICS I 指数的对比、没有替换在理论上更适合的单个指标，但它确实在构建方法上有所改进，故本文拟引用王镇、郝刚（2014）基于偏最小二乘法构建的指数表达式来构建文本的投资者情绪综合指数。

2.1.3 投资者情绪的理论解释

由投资者情绪的定义可知，本文偏向于从个体心理的角度探讨投资者情绪，而这一现象也确实是从大量的心理学实验中发现的。这些实验成果表明，在人们的实际决策行为中，预期的形成存在着一些非理性的局限，即偏差；而且在很多情况下，这种偏差是具有普遍性的，它会在总体上对人们的决策行为和决

策结果产生长远的影响。基于这些研究发现，本节将从认知偏差、选择偏好以及群体决策偏差这三个方面对投资者情绪进行解释。

（1）认知偏差

当人们在形成预期时，通常会根据自己的直觉或者一些常识、经验。这些直觉帮助人们确定的“答案”不一定是正确的，可能很接近，但也可能出现一些系统性的偏差。

① 代表性偏差

代表性偏差是指当面对不确定的情况时，人们会根据问题某个最突出的点直接做出判断，并不进行逻辑思考。在股票市场上，投资者对涨势良好的股票表现得过于乐观，认为它们的价格会继续上涨，进而买入甚至加仓；而对处于价格下跌当中的股票表现得过于悲观。这种偏差会导致股票价格偏离其基本价值。但是随着时间的推移，被高估的股票价格会回调，跌破市场平均水平的股票也会反弹。

② 保守性偏差

保守性偏差是指人们的预期一旦形成就将长期固守，新信息对原有预期的修正往往不足。一些学者将保守性偏差用于解释投资者对新信息作出的不充分反应，正好与代表性偏差相反。其实面对同一新消息，这两种偏差是统一的。当人们认为新信息对于做出判断是有效的，就会把关注点转移到新信息上，进而产生代表性偏差。如果人们认为新信息并没有什么作用，则会忽略它，产生保守性偏差。

③ 过度自信

过度自信是指相信自己的判断到了超越相信事实的程度。大量的研究表明，投资者对自己的投资能力过度自信，这主要表现在以下三个方面：

第一，过度自信的投资者会高估自己评估股票的能力。如果投资者已经购买了某只股票，他就会坚持认为这种投资是正确的选择，同时他会忽视这只股票的负面信息，比如最初不应该购买的信息，或者应该卖掉的信息等等。

第二，过度自信的投资者会认为自己比其他投资者提前掌握了绝对正确的信息。在这种情况下，他们会十分频繁地进行交易，从而被交易成本拉低回报。

第三，过度自信的投资者一般会“把全部鸡蛋放在同一个篮子里”，而不会进行风险分散化的投资。

（2）选择偏好

① 前景理论

前景理论由 Kahneman、Tversky(1979)提出，根据一系列的实验结果，他们发现人们在收益和损失情况下对风险的偏好是不同的：当面临损失时人们通常会变得爱好风险，而有所收益时是厌恶风险的。这在股票市场上就表现为投资者急于卖出盈利的股票，而不轻易卖出亏损股票。与此同时，人们对于收益和损失的感觉依赖于某个参考点（由过去或现在的经验背景确定），人们的决策行为随着参考点的变化而变化；而在决策时，人们又常常会强调小概率事件而忽视一般或者高概率事件，比如说投资者往往过于看重涨停板的股票。

② 后悔厌恶

后悔厌恶是指人们在犯错误之后都会感到后悔，后悔带来的痛苦可能大于由于错误本身引起的损失，而且采取行动的后悔程度要远高于没有采取行动的后悔程度。从长期来看，即使没有做出决策，也会产生后悔心理。比如投资者没有相信分析师对某只股票的推荐买入建议，后来该股票价格上涨，投资者可能后悔当时没有采取行动。因此，人们在决策时的目标可能是最小化未来的后悔程度，而这既有可能导致其规避风险，也有可能导致其追求风险。

（3）群体决策偏差

投资者作为一个社会人，难免受到社会其他人的影响，而这种影响有可能改变投资者个人独自进行决策时预期的形成和选择偏好的确定。羊群效应就是这种影响的典型表现。

羊群效应可以理解为市场中的投资者在进行投资的过程中会被其他投资者的行为所影响，具体表现为在股票市场上跟随其他投资者进行持仓。产生这一现象的原因是因为投资者对自己的决策没有信心，而认为其他人的决策可能包含了一些内幕消息或具有更强的技术分析。同理，投资者在卖出股票时也存在这一效应。当跟随他人进行投资的投资者人数逐渐增多时，这一效应趋于明显，从而使市场上的投资者的行为出现单一性。羊群效应是一种跟风行为，跟风对象往往是证券市场上的专家，如股票分析师和所谓的股神。这种跟风行为还具有无意识性，即投资者并不知道自己的行为属于羊群效应。

2.2 隐马尔科夫模型（HMM）理论

2.2.1 马尔科夫链

马尔科夫链是一个时间与状态都离散的随机过程，其未来状态的概率分布只与当前状态有关，与历史状态无关。

（1）基本概念

设随机过程由 3 个不同的状态（ S_1, S_2, S_3 ）组成，如图 2.1 所示。

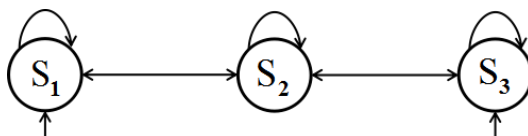


图 2.1 3 个不同状态的马尔科夫链

记状态变化的时间常数为 $t=1,2,\dots$ ， t 时刻的状态为 q_t 。若在任一时刻 t ，有

$$P\{q_t = S_j / q_{t-1} = S_i, q_{t-2} = S_k, \dots\} = P\{q_t = S_j / q_{t-1} = S_i\} \quad (2.1)$$

则称该随机过程为马尔科夫链。

在 t 时刻处于状态 i 的条件下，到 $t+1$ 时刻变为处于状态 j 的条件概率为

$$a_{ij}(t) = P\{q_{t+1} = S_j / q_t = S_i\} \quad (2.2)$$

称为一步转移概率。若马尔科夫链的一步转移概率 a_{ij} 与转移的起始时刻 t 无关，则为齐次马尔科夫链。

状态 i 经过 n 步转移后，处于状态 j 的概率为

$$a_{ij}^{(n)} = P\{q_{m+n} = S_j / q_m = S_i\} \quad (2.3)$$

称为 n 步转移概率。

由一步转移概率为元素构成的矩阵为一步转移矩阵 $A_1 = (a_{ij})$ ，由 n 步转移概率为元素构成的矩阵为 n 步转移概率矩阵 $A_n = (a_{ij}^{(n)})$ 。

当处于 0 时刻时，若 $a_i(0) = P\{q_0 = S_i\}$ 满足

$$a_i(0) \geq 0, \quad \sum_{i=1}^N a_i(0) = 1 \quad (2.4)$$

则称 $\{a_i(0)\}$ 为马尔科夫链的初始分布。马尔科夫链在初始时刻有可能处于任意状

态，因此初始分布就是马尔科夫链在初始时刻的概率分布。

当处于 t 时刻时，若 $a_i(t) = P\{q_t = S_i\}$ 满足

$$a_i(t) \geq 0, \sum_{i=1}^N a_i(t) = 1 \quad (2.5)$$

则称 $\{a_i(t)\}$ 为该马尔科夫链的绝对分布，可以行矩阵表示为：

$$A(t) = \{a_1(t), a_2(t), \dots, a_N(t)\} \quad (2.6)$$

(2) 马尔科夫链的预测原理

若一个马尔科夫链是齐次马尔科夫链，包含 N 个状态，则符合两个定理：

① 马尔科夫链的绝对概率由其初始分布和 n 步转移概率完全确定，即

$$\begin{aligned} A(t+1) &= A(t)A_1 \\ A(t) &= A(0)A_n \end{aligned} \quad (2.7)$$

② C-K 方程，即

$$\begin{aligned} A_{m+n} &= A_m A_n \\ A_n &= A_1^n \end{aligned} \quad (2.8)$$

将上述两个定理结合起来即为马尔科夫链预测模型：

$$A(t) = A(0)A_n = A(0)A_1^n \quad (2.9)$$

即计算出一步转移概率矩阵 A_1 、确定初始分布 $A(0)$ 之后，可求出各时刻的绝对分布 $A(t) = \{a_1(t), a_2(t), \dots, a_N(t)\}$ ，从而所预测的状态 j 满足

$$a_j(t) = \max\{a_i(t)\} \quad (2.10)$$

即所预测的状态就是概率值最大的那种状态。

2.2.2 隐马尔科夫模型

隐马尔科夫模型 (Hidden Markov Model, HMM) 是马尔科夫链的一种。该模型由具有一定隐状态数的马尔科夫链和显性随机函数集构成。其中，马尔科

夫链表示隐状态序列，模型的关键参数转移概率矩阵可以表示状态之间的转移情况。在隐状态和显性观测值之间存在一一对应的关系，这种关系可以用观测概率转移矩阵表示。

(1) 基本概念

在马尔科夫链的基础上，一个隐马尔科夫模型具有以下元素：

- ① 模型的隐状态数目 N
- ② 每个隐状态所具有的不同观测值的数目 M
- ③ 状态转移概率矩阵 $A = \{a_{ij}\}$ ，且有

$$a_{ij} = P\{q_{t+1} = S_j | q_t = S_i\}, \quad 1 \leq i, j \leq N \quad (2.11)$$

- ④ 观测概率矩阵 $B = \{b_j(k)\}$ ，且有

$$b_j(k) = P\{v_k | q_t = S_j\}, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (2.12)$$

也就是说，当状态为 S_j 时，观测结果为 v_k 的后验概率。

- ⑤ 初始状态概率矩阵 $\pi = \{\pi_i\}$ ，且有

$$\pi_i = P\{q_0 = S_i\}, \quad 1 \leq i \leq N \quad (2.13)$$

一般地，可以用 $\lambda = (A, B, \pi)$ 简洁地表示一个隐马尔科夫模型。给定了 N, M, A, B, π 后，隐马尔科夫模型就可以产生一个观测序列 $O = O_1 O_2 \dots O_T$ 。

(2) 评估问题与前向-后向算法

评估问题是应用隐马尔科夫模型时的一个基本问题：对于观测序列 $O = O_1 O_2 \dots O_T$ 和模型参数 $\lambda = (A, B, \pi)$ ，怎样有效计算观测序列的输出概率，即 $P\{O|\lambda\}$ ？

这实际上是一个计算问题，即给定模型参数和观测序列，如何计算由该模型产生的观测序列的概率，也可以把这个问题看做评估给定模型与给定观测序列的匹配程度，通常使用前向-后向算法能够求解。

- ① 隐马尔科夫模型的前向概率为

$$a_t(i) = P\{O_1 O_2 \dots O_t, q_t = i | \lambda\} \quad (2.14)$$

表示给定 HMM 参数，部分观测序列 $\{O_1 O_2 \dots O_t\}$ 在时刻 t 处于状态 i 的概率。它的递推计算公式如下：

初始化:

$$a_t(i) = \pi_i b_i(O_t), \quad 1 \leq i \leq N \quad (2.15)$$

迭代计算:

$$a_{t+1}(j) = \left(\sum_{i=1}^N p_{ij} a_t(i) \right) b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N \quad (2.15)$$

终止计算:

$$P(O|\lambda) = \sum_{i=1}^N a_t(i) \quad (2.16)$$

② 隐马尔科夫模型的后向概率为

$$\beta_t(i) = P\{O_{t+1}O_{t+2}\dots O_T, q_t = i | \lambda\} \quad (2.17)$$

表示给定 HMM 参数，观测序列在时刻 t 处于状态 i ，系统输出部分观测序列 $\{O_{t+1}O_{t+2}\dots O_T\}$ 的概率。它的递推计算公式如下：

初始化:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (2.18)$$

迭代计算:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-1, \quad 1 \leq i \leq N \quad (2.19)$$

终止计算:

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \quad (2.20)$$

③ 这样，可以根据前向概率和后向概率得到整个观测序列的输出概率：

$$P(O|\lambda) = \sum_{i=1}^N \sum_{j=1}^N a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-1 \quad (2.21)$$

第3章 基于投资者情绪的股市 HMM 预测的理论分析

3.1 投资者情绪影响股市的理论分析

3.1.1 情绪-金融决策模型

大量的行为金融研究表明，情绪对金融决策有显著性影响，尤其是存在不确定性和风险时。赖凯声等人（2014）修改的情绪-金融决策模型，可以很好地解释投资者情绪在金融决策中的作用。

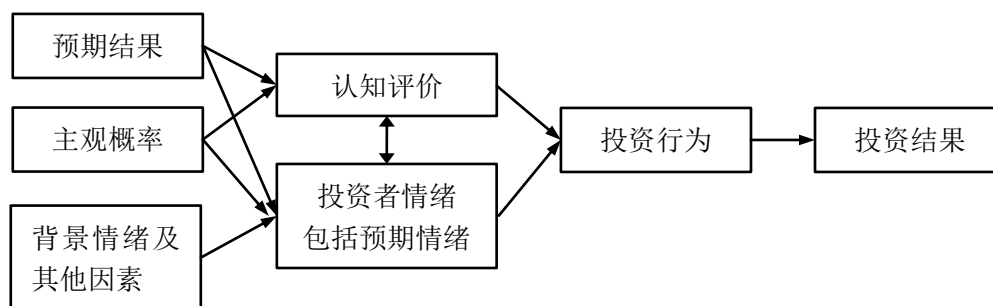


图 3.1 情绪-金融决策模型

如图 3.1 所示，从投资者出现投资意愿到产生投资结果一共经历了四个环节。首先，投资者会对投资结果进行预测，并猜测可能出现的结果。这两者会影响投资者对其决策的评价与认知。其次，投资者进行投资时会受到环境因素影响从而产生背景情绪，它与前述的两种因素会共同干预投资者决策时的心理状态，即投资者情绪。随后，投资者对决策行为的认知与评价和投资者情绪会相互影响，最终达到动态平衡并产出投资行为。一般情况下，当投资者的情绪处于较好的状态时会对投资结果做出较好的评价，反之亦然。

3.1.2 投资者情绪与股市

由市场规律不难发现，投资者会买入自己看涨的股票而卖出看跌的股票。而不同的投资者对同一支股票的预期往往是不同的，当买卖双方的数量不平衡时即出现了股票供需失衡，此时股价会出现变动进而调节供需双方达到平衡。因此可以认为股票的价格实际上包含了市场上整体投资者对其价格的预期。上一节所述的情绪-金融决策模型表明，情绪能影响金融决策，那么从理论上讲，

情绪也能影响股票市场的运行状况，比如股票的价格、收益率等。

投资者情绪通过两种方式影响投资者的行为。一是影响投资者对投资结果的认知行为，然后通过认知的变化间接影响投资者的行为；二是直接干预投资者的决策过程。具体而言，当投资者处于较好的情绪下时，他们更容易忽略投资中包含的风险而只看到可能的收益，这使投资者的交易动机大幅度增加。反之，当投资者处于消极的情绪下时，往往会减少自己的投资行为以避免其主观意愿所高估的风险。最后，这样的投资结果就表现为股票市场受到了影响、出现了波动，甚至是大起大落、暴涨暴跌。

因此，投资者情绪影响股票市场是具有理论基础的。那么若能捕捉到投资者情绪、将其量化，再运用合适的方法，以投资者情绪为基础对股票市场进行预测，在理论上就是可行的。

3.2 隐马尔科夫模型预测股市的理论分析

3.2.1 投资者情绪与隐马尔科夫模型

正如上节所述，以投资者情绪为基础对股票市场进行预测需要合适的应用方法，这种方法要能够直接把投资者情绪运用起来，隐马尔科夫模型是很好的选择，具体原因如下：

(1) 在股票市场上，呈现出来的是一组组的数据，是投资者的投资结果，并没有直接表现出投资者的喜怒哀乐等情绪，换言之，投资者情绪在股票市场上是隐藏在数据背后的东西，用陆剑清（2015）的话说就是一种“看不见、摸不着但却时刻能够感受到其存在的市场暗物质”，因此，若要运用隐藏起来的投资者情绪对股票市场上直观的数据进行预测，就需要一座桥把二者连接起来。

如第2章所述，隐马尔科夫模型是一个双重随机过程：具有一定隐状态数的马尔科夫链和显性随机函数集。这个模型的妙处就在于“隐”，这个“隐”是指一条隐状态序列，恰好可以代表股票市场上隐藏起来的投资者情绪；被预测的直观数据可以由模型的观测序列代表；模型的关键参数：状态转移概率矩阵和观测概率矩阵，可以把投资者情绪和股市数据连接起来。

投资者情绪与隐马尔科夫模型的联系可以用图 3.2 表示。

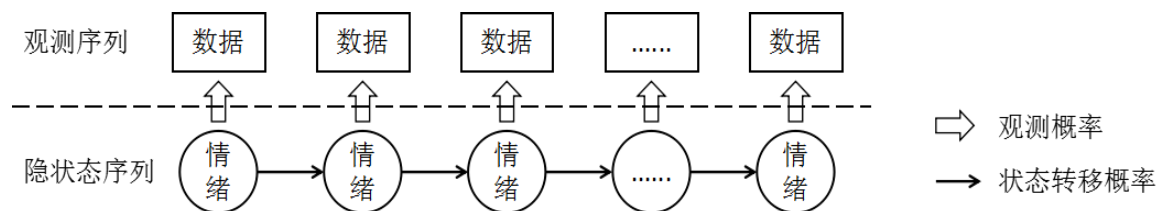


图 3.2 投资者情绪与隐马尔科夫模型联系图

(2) 运用投资者情绪进行股市数据预测的前提是测量投资者情绪，对投资者情绪进行量化。同时，这样也有利于在隐马尔科夫模型中区分隐状态，进而得到状态转移概率矩阵和观测概率矩阵。因此，运用投资者情绪构建隐马尔科夫模型的思路是很清晰的：

把投资者情绪利用进隐马尔科夫模型，首先要量化投资者情绪，根据第 2 章投资者情绪的理论基础，也就是构建投资者情绪综合指标，得到投资者情绪指标数值；然后，根据数值的大小划分投资者情绪的状态，比如乐观、悲观、平淡，并将各数值转化为状态，这就是隐马尔科夫模型中的隐状态序列。接下来，由于隐状态序列本质是一条马尔科夫链，所以可以统计出各隐状态之间相互转移的一步转移频率，并写成矩阵的形式，这就是隐状态之间的状态转移概率矩阵。

对于观测序列的直观数据，同样需要根据数值大小对结果进行分类，比如说涨、跌，然后把隐状态和数据的类别一一对应起来，再统计出各隐状态对应不同数据类别的频率，并写成矩阵的形式，这就是观测概率矩阵。

这样，就得到了基于投资者情绪的隐马尔科夫模型：投资者情绪状态序列、股市直观数据序列、投资者情绪状态转移概率矩阵、观测概率矩阵。

3.2.2 预测原理分析

隐马尔科夫模型以数据模式识别为原理，在设定初始参数后，通过不断的迭代，识别出最佳数据模式。而且与传统的模型相比，隐马尔科夫模型无需进行复杂的数学模型推导和参数寻优。

因此，基于投资者情绪运用隐马尔科夫模型对股市进行预测，就是以投资者情绪作为隐状态、以预测对象作为观测序列，当确定了状态转移概率矩阵和

观测概率矩阵，模型就可以产生新的观测序列，并计算出该序列出现的概率，从而实现预测。

隐马尔科夫模型产生观测序列 $O = O_1 O_2 \dots O_T$ 的过程如下：

- (1) 根据初始状态概率矩阵选择一个初状态 $q_1 = S_i$
- (2) 令 $t=1$
- (3) 根据状态 S_i 的观测概率，输出一个观测值 O_1
- (4) 根据状态 S_i 的状态转移概率，转移到一个新的状态 $q_{t+1} = S_j$
- (5) 令 $t=t+1$ ，若 $t < T$ ，返回 (3)，否则结束

计算该序列出现概率的方法，在第 2 章已经介绍过，也就是前向-后向算法，通过初始化、迭代计算和终止计算，可以得到整个序列的输出概率：

$$P(O/\lambda) = \sum_{i=1}^N \sum_{j=1}^N a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-1 \quad (3.1)$$

这也就是预测结果出现的可能性。

第4章 基于投资者情绪的股市HMM预测的实证分析

4.1 构建投资者情绪综合指标

4.1.1 偏最小二乘法介绍

在把偏最小二乘法(PLS)运用于构建投资者情绪之前,有两个重要的假设。第一个是假设投资者预期的下一期收益率主要受本期情绪影响,两个变量之间是标准的线性关系,即

$$E(R_{t+1}) = \alpha + \beta S_t \quad (4.1)$$

其中, S_t 表示与预期收益率相关的投资者情绪,此处为投资者情绪综合指标。

由此,下一期已实现的收益率可表示为:

$$R_{t+1} = E(R_{t+1}) + \varepsilon_{t+1} = \alpha + \beta S_t + \varepsilon_{t+1} \quad (4.2)$$

其中, ε_{t+1} 是与 S_t 无关的误差项。

第二个假设是关于单个投资者情绪代理变量的。假设每个投资者情绪代理变量都有如下结构:

$$x_t = \eta_0 + \eta_1 S_t + \eta_2 E_t + e_t \quad (4.3)$$

其中, S_t 表示与预期收益率相关的投资者情绪; E_t 表示与投资者情绪有关、与预期收益率无关的各情绪代理变量的共同近似偏差; η_1 、 η_2 分别表示 x_t 对 S_t 、 E_t 变动的敏感度; e_t 表示 x_t 特有的误差项。

现在,关键问题是能否把单个投资者情绪代理变量 x_t 中的投资者情绪 S_t 分离出来,同时在这个过程中消除共有近似偏差 E_t 和特有误差项 e_t 。

偏最小二乘法可以通过两步普通最小二乘法(OLS)回归来解决这个问题。

第一步,对于单个投资者情绪代理变量 x_t ,建立如下回归方程:

$$x_{t-1} = \pi_0 + \pi_1 R_t + u_{t-1} \quad (4.4)$$

其中, π_1 通过已实现的下一期收益率 R_t 反映了单个投资者情绪代理变量 x_{t-1} 对投资者情绪 S_{t-1} 的敏感度。因为由式(4.2)、式(4.3)可知, R_t 由预期收益

率和误差项组成，而预期收益率主要受投资者情绪 S_{t-1} 的影响，因此， π_1 可以反映出单个投资者情绪代理变量 x_{t-1} 在多大程度上取决于真实的投资者情绪综合指标 S_{t-1} 。

第二步，对于每一个时间 t ，建立如下回归方程：

$$x_i = c + S^{PLS} \hat{\pi}_{1,} + v_i \quad (4.5)$$

其中， S^{PLS} 表示所估计的投资者情绪综合指标。在式 (4.5) 中，第一步所得到的 $\hat{\pi}_{1,}$ 成了解释变量，而投资者情绪综合指标 S^{PLS} 成了待估参数。

经过数学运算，投资者情绪综合指标 S^{PLS} 可以表示为 $T \times 1$ 阶的向量：

$$S^{PLS} = XJ_N X' J_T R (R' J_T XJ_N X' J_T R)^{-1} R' J_T R \quad (4.6)$$

其中， X 表示的是单个投资者情绪测量的 $T \times N$ 阶矩阵，即 $X = (x_1', \dots, x_T')$ ； R 表示收益率的 $T \times 1$ 阶向量，即 $R = (R_2, \dots, R_{T+1})'$ ；矩阵 $J_T = I_T - \frac{1}{T} 1_T 1_T'$ ，而 I_T 是 T 阶单位矩阵、 1_T 是元素全为 1 的 T 阶向量。

最后，再经过一系列的整理之后，PLS 构建的投资者情绪综合指标可以表示为：

$$S^{PLS} = \beta X \quad (4.7)$$

其中， $X = (X_1, \dots, X_N)'$ 表示单个投资者情绪指标序列， $\beta = (\beta_1, \dots, \beta_N)$ 表示每个单个投资者情绪指标在综合指标中所占的权重。

4.1.2 单个指标的选择

单个投资者情绪指标主要分为间接代理指标和直接调查指标。间接代理指标虽然能客观地反映实际参与到交易中的投资者的情绪，但它们能在多大程度上真实反映仍然不能确定；直接调查指标虽然真实地反映了投资者的情绪，但接受调查的投资者不一定参与交易、不一定有所行动。因此，本文认为，投资者情绪综合指标里一定既要包括间接指标，又要包括直接指标。

本文基于数据的可获取性和连续性，选择以封闭式基金折价率、股票成交量、换手率、新增股票投资者开户数四个间接代理指标，和消费者信心指数一个主观调查指标来构建本文的投资者综合指标。

(1) 封闭式基金折价率 (CEFD)

封闭式基金折价率反映的是投资者的悲观情绪，预期与综合指标呈负相关关系。本文采用的封闭式基金月折价率是由各月的日折价率取平均值得到。

（2）股票成交量（TV）

成交量是流动性指标，大量相关研究表明其与投资者情绪是正相关关系。本文采用的是沪市月度交易量的数据。

（3）换手率（TURN）

换手率也是流动性指标，高换手率既可以反映投资者参与交易的情绪积极，又可以反映出他们情绪的变动性。本文采用的是沪市流通市值加权市场月换手率。

（4）新增股票投资者开户数（NIA）

该指标目前只有国内学者在使用，他们认为，新增投资者开户数越多，表明投资者进入股票市场的情绪越高涨。本文采用沪市新增投资者开户数的月度数据。

（5）消费者信心指数（CCI）

虽然投资者信心指数比消费者信心指数更直接地反映投资者的情绪，但由于前者数据的不可获得性，加上相关文献支持后者对情绪的衡量，故本文采用消费者信心指数的月度数据，并认为消费者信心指数越高，投资者情绪越积极。

本文数据的样本期为2007年1月至2015年12月，其中CEFD取自Wind中国金融数据库，TV、TURN取自锐思数据库，NIA摘自中国证券登记结算有限公司网站，CCI摘自东方财富网。此外，上证综指月收益率R取自锐思金融数据库。为消除量纲的影响，以上样本数据在进行综合指标构建之前都进行了标准化处理。

4.1.3 综合指标的构建

考虑到所选单个指标对投资者情绪的反应存在时间上的提前和滞后关系，易志高、茅宁（2009）在构建综合指标时还包括了各单个指标的滞后一期，这样共计5对10个变量，再通过相关性分析选取各对中与综合指标相关程度较高的，最终构建成投资者情绪综合指标。本文采用类似的处理方法。

首先，将10个变量代入到式（4.6）中，从而得到PLS法构建的投资者情绪综合指标，用 S^{PLS} 来表示。然后，把 S^{PLS} 和这10个变量进行相关性分析，比较当期、滞后一期变量和 S^{PLS} 之间的相关程度，选出各对变量中相关系数较高的、

共计 5 个变量作为构建 S^{PLS} 的最终单个指标。相关性分析结果如表 4.1 所示：

表 4.1 S^{PLS} 与 10 个变量的相关性分析

	S_t	CCI _t-1	NIA _t-1	TA _t-1	TURN _t-1	CEFD _t-1	CCI _t	NIA _t	TA _t	TURN _t	CEFD _t
Pearson	1	.428**	.731**	.690**	.767**	-.310**	.433**	.864**	.728**	.868**	-.335**
相关性											
显著性											
S_t (双侧)		.000	.000	.000	.000	.001	.000	.000	.000	.000	.000
N	108	108	108	108	108	108	108	108	108	108	108

**. 在 .01 水平（双侧）上显著相关。

由表 4.1 可知，在所选的 5 对、10 个变量中，与 S^{PLS} 相关程度较高的变量分别是 CCI_t、NIA_t、TV_t、TURN_t 和 CEFD_t，即均为当期变量更好地反映了投资者情绪。将这 5 个优胜变量重新代入到式（4.6）、式（4.7）中，并对新的 S^{PLS} 再次进行相关性分析，结果如表 4.2 所示：

表 4.2 S^{PLS} 构建的最终结果

	Beta		0.1370	0.2865	0.2437	0.2566	-0.0054
		S_t	CCI _t	NIA _t	TA _t	TURN _t	CEFD _t
Pearson	相关性	1	.437**	.896**	.763**	.832**	-.280**
S_t	显著性（双侧）		.000	.000	.000	.000	.003
N		108	108	108	108	108	108

**. 在 .01 水平（双侧）上显著相关。

根据表 4.2 中的 Beta 值，可以得到本文构建的投资者情绪综合指标 S^{PLS} 的表达式：

$$S^{PLS} = 0.137 CCI_t + 0.2865 NIA_t + 0.2437 TV_t + 0.2566 TURN_t - 0.0054 CEFD_t \quad (4.8)$$

由式（4.8）中各变量前面的系数符号可知，消费者信心指数、新增股票投资者开户数、股票交易量和换手率与投资者情绪正相关，各变量的数值越大，表示投资者情绪越高涨；而封闭式基金折价率与投资者情绪负相关，表示投资

者情绪的消极。本文若单独考虑更专业的机构投资者，其坚持止盈、止损的心理可能会使得成交量与换手率的相关性呈现另一种情况。

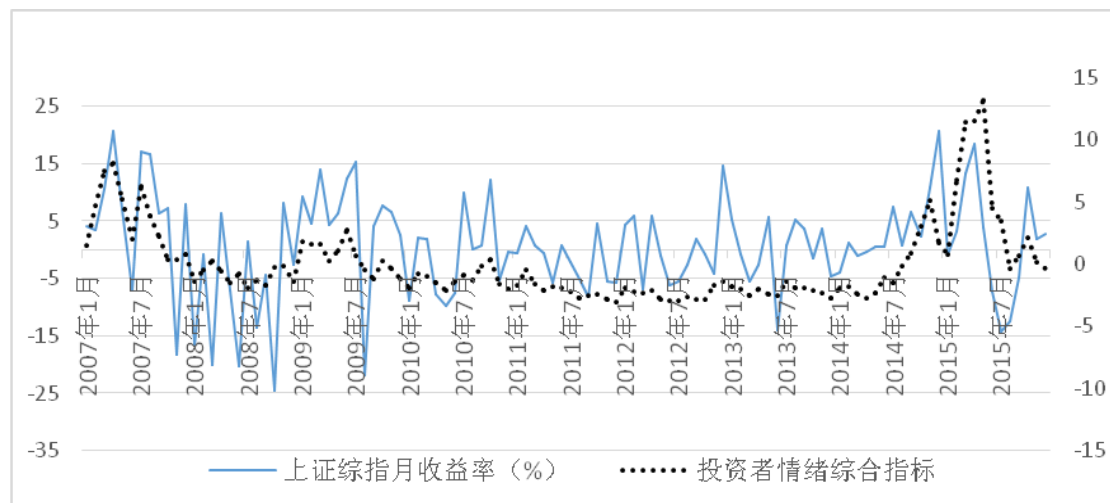


图 4.1 上证综指月收益率与投资者情绪综合指标变动关系图

4.2 我国“政策市”对投资者情绪的影响

4.2.1 提出假设

我国股市历来被称为“政策市”，主要是指相关政策的出台会导致股市不同程度的波动。在这个方面，我国学者已经进行了充分地论证，比如王明涛等人（2012）将 1993 年至 2009 年的股市政策划分为供求政策、制度性政策、投资者预期政策和其他政策四大类、共 16 个变量并进行量化，通过建立多元回归方程分析得到了政策因素是影响我国股市整体波动的主要因素这一结论；杨继平等人（2012）根据 1997 年至 2010 年股指日收益率的 16 个结构变点选取出政策性事件，通过回归分析发现国家股市政策、宏观经济政策对股市波动产生了巨大的影响，并进一步分析了这些事件对于股市波动性增加或减少的影响。大量的研究表明，中国股市是典型的“政策市”，相关政策的出台会导致中国股市整体波动。

因此，若对我国股市进行预测，就不得不考虑政策这个因素，也就不得不思考政策因素与投资者情绪之间的关系。李静（2012）通过事件研究法分析 2005 年至 2010 年股市的利好、利空政策对个人投资者行为的影响，其结论表明，政

策性因素通过影响投资者的交易投资活跃程度和出入市意愿，进而影响了交易频率和成交量变动率。张前程、杨德才（2015）分析认为，货币政策可以通过改变投资者的心理预期和认识偏差，从而直接塑造投资者情绪，并构建回归模型对此进行了实证。靳光辉等人（2016）则站在政策不确定性的角度，实证检验了投资者情绪的中介作用。

基于这样的认识，本文提出一个大胆的假设：政策因素是通过投资者情绪对股市价格产生影响的，投资者情绪已经包含了对政策的反应。又由于投资者情绪可被成交量、换手率这两个指标捕捉，而二者对股市价格所起的重要性作用不言而喻，故本文只需实证政策的出台会影响成交量和换手率即可。

4.2.2 数据来源

对于成交量和换手率，本文选取的是上证综指在 2007 年 1 月 1 日-2015 年 12 月 31 日相关交易日的日数据，取自锐思金融数据库。

对于政策因素，本文选取了 2007 年至 2015 年期间发生、对股市造成重大影响的政策事件，为了在后续的实证环节避免事件窗口的重合，对于一段时间内出台的多条政策会进行筛选。这些事件收集自中国人民银行官方网站、中国经济网、和讯网、凤凰网等，最终确定了 18 件具有代表性的政策事件，包括 9 次利好事件、9 次利空事件，如表 4.3 所示。

表 4.3 2007 年-2015 年我国股市重大政策事件

序号	时间	政策内容	影响
1	2007.5.30	证券交易印花税税率由 1‰上调为 3‰	利空
2	2007.9.15	人民币存贷款基准利率上调 0.27 个百分点	利空
3	2008.4.24	证券交易印花税税率由 3‰下调为 1‰	利好
4	2008.9.19	人民币贷款基准利率下调 0.27 个百分点；存款准备金率下调 1 个百分点；印花税调整为单边征收	利好
5	2008.11.21	为应对全球金融危机政府推出 4 万亿元人民币投资计划	利好
6	2009.8.13	银行理财资金禁入股票二级市场；证监会公布《证券投资基金评价业务管理暂行办法》，监管手法加强	利空

续表 4.3 2007 年-2015 年我国股市重大政策事件

7	2010.4.16	国务院发布《关于坚决遏制部分城市房价过快上涨的通知》（新国十条）；股指期货上市	利空
8	2010.11.30	首批创业板公司限售股解禁流通	利空
9	2011.7.7	半年内共 9 次上调人民币存款准备金率、存贷款基准利率	利空
10	2012.7.5	一个月内第 2 次下调人民币存贷款基准利率	利好
11	2012.11.18	财政部、国税总局、证监会发布《关于实施上市公司股息红利差别化个人所得税政策有关问题的通知》	利好
12	2013.10.8	严格借壳上市，出台并购重组分道审核制	利空
13	2014.5.9	国务院发布《关于进一步促进资本市场健康发展的若干意见》（新国九条）	利好
14	2014.11.21	沪港通；人民币存贷款基准利率下调 0.4 个百分点	利好
15	2015.1.16	证监会整顿券商融资融券业务	利空
16	2015.4.19	人民币存款准备金率下调 1 个百分点	利好
17	2015.6.12	中国证券业协会发布《证券公司外部接入信息系统评估认证规范》，严查券商场外配资	利空
18	2015.7.15	暂停 IPO、主要券商启动稳定市场基金、央行给予证金公司流动性支持	利好

4.2.3 事件研究法与实施步骤

在本节，本文采用事件研究法来分析相关政策出台对成交量和换手率两个指标的影响。事件研究法是一种运用统计思想，针对某一特定事件对研究对象造成影响的程度和持续时间进行度量以及检验的方法，主要通过计算异常变化率和累积异常变化率等指标来分析事件的影响。其步骤包括定义事件和事件窗口，选择估算正常值的模型，计算并检验异常值、分析结果等。具体实施步骤如下。

（1）定义事件

见表 4.3。

(2) 事件窗口

事件窗口包含估计期、事件期和事后检验期，本文以政策公布日为事件期，以公布日前 15 个交易日为估计期，以后 15 个交易日为事后检验期，故整个事件窗口为 31 天。

(3) 正常值的计量模型

正常值定义为事件发生后的估计值，计量模型主要是均值调整模型和市场模型。由于本节研究的是成交量和换手率的异常变化，故采用较简单的均值调整模型：

$$E(X_{it}) = \mu_i + \varepsilon_{it} \quad (4.9)$$

其中， $E(\varepsilon_{it})=0$ $D(\varepsilon_{it})=\sigma^2$ ， X 代表本节研究的两个变量：成交量变化率（ Q ）和换手率变化率（ E ）。

(4) 计算异常值并检验

异常值定义为事件发生后的实际值与正常值之差。先使用对数变化率反应成交量和换手率的变化情况：

$$Q_{it} = \log(TV_{it}) - \log(TV_{it-1}) \quad (4.10)$$

$$E_{it} = \log(TURN_{it}) - \log(TURN_{it-1}) \quad (4.11)$$

再计算异常值、累积异常值：

$$AX_{it} = X_{it} - E(X_{it}) \quad (4.12)$$

$$CAX_{it} = \sum AX_{it} \quad (4.13)$$

最后对 AQ 、 AE 、 CAQ 、 CAE 进行检验，使用 T 统计量，如果检验 T 统计量大于显著性水平下的临界值，就拒绝原假设，即事件使得成交量和换手率发生了异常变动。

4.2.4 实证结果与分析

本文选取了 18 次政策事件，包括 9 次利好事件、9 次利空事件，每个事件的事件窗口是 31 天，故参与检验的共有 558 个有效样本。实证结果与分析如下。

由表 4.4 可知,对于利空政策,异常成交量变化率在事后检验期小于 0 的天数要多于事件发生前的天数;由 t 值可知,在事件发生前第 8 天和第 2 天的异常成交量变化率显著小于 0,这很有可能是因为政策信息在公布之前就被泄露或者被预估;在事件发生后的第 8 天和第 11 天,异常成交量变化率同样显著小于 0 并且在第 11 天是最小值,这说明投资者对于公布的利空政策比较谨慎,一开始处于观望状态,在确认后才出现明显缩量,表明投资者对利空政策的反应不足,具有延迟性。累积异常成交量变化率在利空政策出台后的第 13、14 天才显著小于 0 也可以说明这一点。对于利好政策,异常成交量变化率正负互现,表现并不突出,但是累积异常成交量变化率在利好政策发生后的第 1、2、4、5、6 天都显著大于 0,并且在第 1 天达到最大值,随后小幅减少,说明投资者对于利好政策的反应是热烈的而且具有持续性。对于全部政策,异常成交量变化率在事件发生后多次显著异于 0,说明政策的出台确实对成交量的变动产生了明显的影响。

表 4.4 政策事件对异常成交量变化率、累积异常成交量变化率的影响

事件窗口	利空政策				利好政策				全部政策			
	AQ	t	CAQ	t	AQ	t	CAQ	t	AQ	t	CAQ	t
-15	0.025	1.093	0.025	1.093	-0.006	-0.206	-0.006	-0.206	0.010	0.537	0.010	0.537
-14	0.010	0.378	0.035	1.002	0.006	0.219	0.000	0.004	0.008	0.432	0.018	0.792
-13	0.025	1.087	0.060	1.457	-0.001	-0.061	-0.001	-0.037	0.012	0.850	0.030	1.209
-12	0.003	0.165	0.063	1.675	0.031	0.874	0.030	0.939	0.017	0.877	0.046	1.916
-11	-0.017	-0.546	0.046	0.963	-0.013	-0.622	0.017	0.521	-0.015	-0.822	0.031	1.114
-10	-0.008	-0.243	0.038	1.240	-0.037	-1.539	-0.021	-0.713	-0.023	-1.137	0.009	0.401
-9	0.015	0.420	0.052	1.234	0.049	1.537	0.029	0.750	0.032	1.365	0.041	1.455
-8	-0.034	-2.254	0.018	0.376	0.022	1.028	0.051	1.576	-0.006	-0.430	0.034	1.217
-7	-0.014	-0.549	0.004	0.110	-0.039	-1.051	0.012	0.287	-0.027	-1.201	0.008	0.299
-6	0.024	2.232	0.028	0.763	0.006	0.155	0.018	0.340	0.015	0.752	0.023	0.736
-5	-0.005	-0.251	0.023	0.544	0.005	0.248	0.023	0.405	0.000	0.015	0.023	0.671
-4	0.009	0.442	0.033	0.619	0.003	0.236	0.026	0.436	0.006	0.515	0.029	0.759
-3	-0.031	-1.154	0.001	0.021	0.019	0.631	0.045	0.729	-0.006	-0.287	0.023	0.550
-2	-0.057	-2.697	-0.056	-1.019	-0.021	-1.413	0.024	0.421	-0.039	-2.940	-0.016	-0.406
-1	0.007	0.381	-0.049	-1.013	0.037	1.404	0.061	0.934	0.022	1.371	0.006	0.139
0	0.039	1.407	-0.010	-0.246	0.043	1.047	0.103	1.458	0.041	1.706	0.047	1.105
1	-0.009	-0.208	-0.019	-0.333	0.051	0.807	0.154	2.383	0.021	0.568	0.068	1.454

续表 4.4 政策事件对异常成交量变化率、累积异常成交量变化率的影响

2	-0.016	-0.732	-0.035	-0.838	-0.055	-1.679	0.099	2.321	-0.035	-1.803	0.032	0.972
3	-0.017	-0.673	-0.051	-0.895	-0.056	-2.168	0.043	1.170	-0.036	-2.019	-0.004	-0.114
4	0.018	0.976	-0.033	-0.628	0.081	2.139	0.124	1.990	0.049	2.271	0.046	1.034
5	-0.022	-1.727	-0.055	-1.052	-0.018	-0.515	0.106	1.982	-0.020	-1.106	0.025	0.611
6	0.003	0.219	-0.052	-1.023	-0.008	-0.305	0.098	1.996	-0.002	-0.156	0.023	0.590
7	-0.005	-0.173	-0.057	-1.075	-0.011	-0.425	0.086	1.655	-0.008	-0.429	0.015	0.369
8	-0.041	-2.039	-0.098	-1.834	-0.005	-0.217	0.081	1.305	-0.023	-1.453	-0.009	-0.189
9	0.045	1.475	-0.054	-1.482	0.035	1.745	0.116	1.805	0.040	2.253	0.031	0.753
10	0.021	1.553	-0.033	-0.742	-0.007	-0.215	0.109	1.846	0.007	0.420	0.038	0.966
11	-0.047	-3.178	-0.080	-1.672	0.010	0.412	0.119	1.792	-0.019	-1.219	0.020	0.422
12	-0.005	-0.122	-0.085	-1.476	0.015	0.359	0.134	1.618	0.005	0.179	0.025	0.444
13	-0.026	-0.994	-0.110	-1.981	-0.062	-1.794	0.072	1.190	-0.044	-2.050	-0.019	-0.417
14	0.022	0.708	-0.088	-1.937	0.014	0.496	0.086	1.391	0.018	0.884	-0.001	-0.030
15	0.039	0.963	-0.049	-0.919	-0.037	-1.829	0.049	0.834	0.001	0.054	0.000	0.000

从图 4.2 也可以看出，累积异常成交量变化率在利空政策出台之后的十五个交易日都小于 0，并且整体呈下降趋势，而在利空政策出台之后都大于，并由较大波动趋向稳定。这说明投资者是先解读政策再有所行为，对不同的政策有不同的反应。

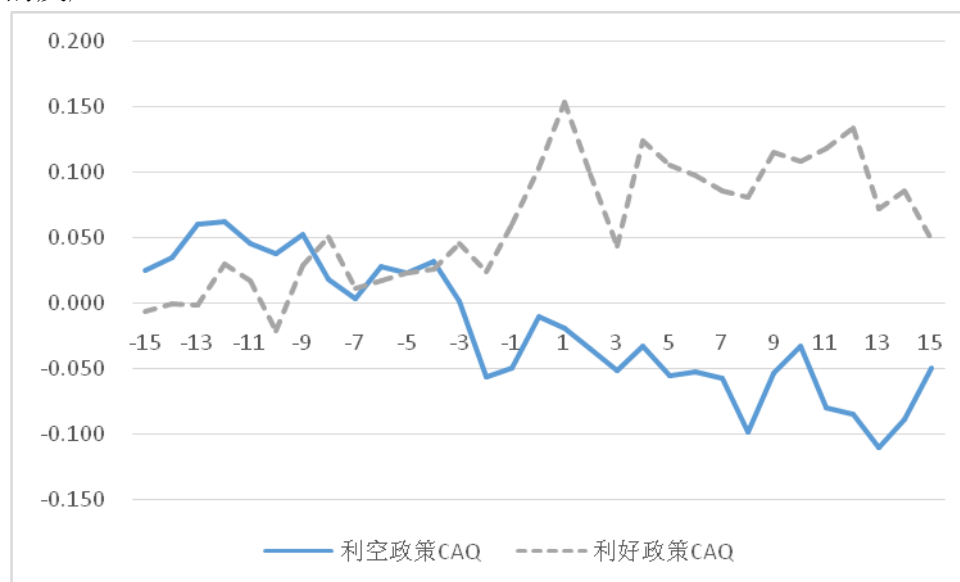


图 4.2 政策事件前后的累积异常成交量变化率

由表 4.5、图 4.3 可知，对于政策事件，换手率的变动与前文分析的成交量

非常类似。对于利空政策，异常换手率变化率在事后检验期小于0的天数更多；同样可能由于小道消息或预期的原因，在事件发生前的第8天和第2天，异常换手率变化率显著小于0；在事件发生后的第8天和第11天，异常换手率变化率显著小于0并且在第8天是最小值，这说明交投活跃程度在利空政策公布的后期才明显降低，表明投资者对利空政策的反应不足。累积异常换手率变化率在利空政策出台后的第13天显著小于0也可以说明这一点。对于利好政策，异常换手率变化率正负互现，表现并不突出，但是累积异常换手率变化率在利好政策发生后的第1、2、4、11天都显著大于0，并且在第1天达到最大值，随后小幅减少，说明在利好政策下投资者的交易意愿是强烈的，并且具有持续性。对于全部政策，异常换手率变化率在事件发生后多次显著异于0，说明政策的出台确实对换手率的变动产生了明显的影响。

表4.5 政策事件对异常换手率变化率、累积异常换手率变化率的影响

事件窗口	利空政策				利好政策				全部政策			
	AE	t	CAE	t	AE	t	CAE	t	AE	t	CAE	t
-15	0.024	1.245	0.024	1.245	-0.007	-0.228	-0.007	-0.228	0.009	0.482	0.009	0.482
-14	0.014	0.595	0.038	1.295	0.000	0.002	-0.007	-0.242	0.007	0.406	0.015	0.755
-13	0.026	1.288	0.063	1.760	0.003	0.160	-0.004	-0.154	0.014	1.109	0.029	1.252
-12	-0.002	-0.153	0.061	1.895	0.030	0.929	0.025	0.850	0.014	0.793	0.043	1.988
-11	-0.026	-0.933	0.035	0.899	-0.015	-0.689	0.010	0.338	-0.021	-1.191	0.023	0.940
-10	-0.001	-0.016	0.035	1.334	-0.032	-1.233	-0.021	-0.786	-0.016	-0.813	0.007	0.343
-9	0.013	0.395	0.048	1.272	0.037	1.194	0.015	0.448	0.025	1.128	0.032	1.266
-8	-0.032	-2.100	0.016	0.389	0.025	1.324	0.040	1.371	-0.003	-0.251	0.028	1.131
-7	-0.010	-0.458	0.006	0.212	-0.040	-1.232	0.000	-0.003	-0.025	-1.291	0.003	0.133
-6	0.028	1.804	0.034	1.080	0.005	0.118	0.005	0.096	0.017	0.786	0.020	0.694
-5	-0.009	-0.427	0.026	0.703	0.014	0.689	0.018	0.369	0.003	0.181	0.022	0.736
-4	0.007	0.327	0.032	0.684	0.001	0.047	0.019	0.372	0.004	0.311	0.026	0.760
-3	-0.025	-0.969	0.008	0.134	0.022	0.669	0.041	0.738	-0.001	-0.065	0.024	0.630
-2	-0.056	-3.191	-0.049	-0.922	-0.027	-1.775	0.015	0.293	-0.042	-3.514	-0.017	-0.478
-1	0.013	0.675	-0.036	-0.811	0.039	1.564	0.054	0.929	0.026	1.671	0.009	0.237
0	0.035	1.438	-0.002	-0.037	0.037	0.934	0.091	1.417	0.036	1.595	0.045	1.167
1	-0.013	-0.331	-0.014	-0.260	0.056	0.888	0.147	2.123	0.022	0.587	0.066	1.407
2	-0.006	-0.309	-0.020	-0.477	-0.052	-1.688	0.095	1.984	-0.029	-1.568	0.038	1.102
3	-0.023	-0.994	-0.043	-0.740	-0.056	-2.148	0.040	0.955	-0.039	-2.272	-0.002	-0.050

续表 4.5 政策事件对异常换手率变化率、累积异常换手率变化率的影响

4	0.012	0.705	-0.031	-0.558	0.081	2.174	0.120	1.879	0.047	2.157	0.045	0.994
5	-0.012	-0.995	-0.043	-0.794	-0.021	-0.594	0.100	1.703	-0.016	-0.918	0.028	0.668
6	-0.009	-0.752	-0.052	-1.004	-0.012	-0.487	0.088	1.813	-0.011	-0.793	0.018	0.460
7	-0.003	-0.140	-0.055	-0.977	-0.009	-0.380	0.079	1.485	-0.006	-0.384	0.012	0.285
8	-0.038	-2.003	-0.093	-1.746	-0.003	-0.135	0.076	1.232	-0.021	-1.356	-0.009	-0.200
9	0.047	1.455	-0.047	-1.288	0.034	1.779	0.110	1.775	0.040	2.219	0.032	0.796
10	0.009	0.715	-0.038	-0.915	-0.010	-0.350	0.099	1.742	-0.001	-0.063	0.031	0.801
11	-0.035	-2.116	-0.073	-1.747	0.018	0.776	0.118	1.911	-0.009	-0.551	0.022	0.514
12	-0.003	-0.092	-0.077	-1.428	0.010	0.243	0.128	1.627	0.003	0.125	0.026	0.486
13	-0.026	-1.044	-0.103	-1.943	-0.058	-1.721	0.070	1.260	-0.042	-2.030	-0.017	-0.390
14	0.027	0.874	-0.076	-1.687	0.012	0.452	0.082	1.511	0.020	0.983	0.003	0.076
15	0.032	0.854	-0.044	-0.948	-0.038	-1.745	0.044	0.852	-0.003	-0.131	0.000	0.000

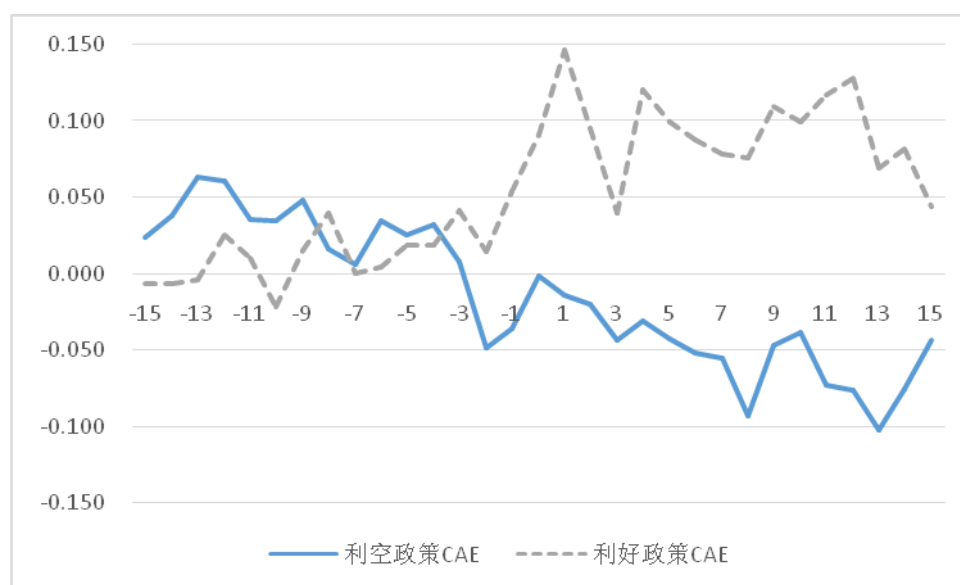


图 4.3 政策事件前后的累积异常换手率变化率

因此，通过上述分析可以得到的结论是：政策出台对成交量和换手率的变动有显著的影响，在利好政策下成交量和换手率都有明显上升，而在利空政策下二者则呈下降趋势。因此，本文在 4.2.1 节提出的假设是成立的，投资者情绪（成交量、换手率）包含了对政策事件的反应，在下文运用投资者情绪对我国股市进行预测时，无需再额外考虑政策因素。

4.3 隐马尔科夫模型的预测

4.3.1 投资者情绪状态转移概率矩阵

对投资者情绪综合指标数值分布进行分析，并参考相关投资者情绪的状态划分，本文认为可以把投资者情绪状态分为五个区间： $S < -2$ 时，投资者过度悲观，此时股市表现出触底反弹； $-2 \leq S < -0.5$ 时，投资者情绪悲观； $-0.5 \leq S \leq 0.5$ 时，投资者情绪平淡，即看平； $0.5 < S \leq 5$ 时，投资者情绪乐观； $S > 5$ 时，投资者过度乐观，大盘可能出现反转。

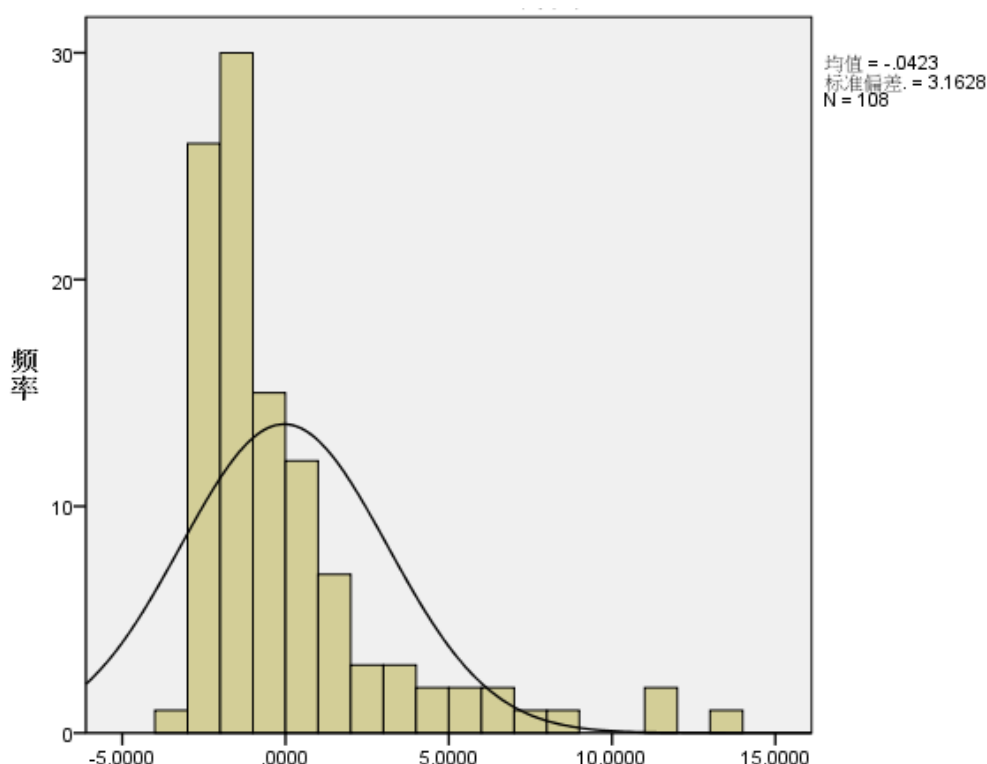


图 4.4 投资者情绪综合指标数值分布图

根据这五种投资者情绪状态，将数值转化为状态，如表 4.6 所示：

表 4.6 2007 年 1 月-2015 年 12 月投资者情绪状态

2007-01	乐观	2009-04	平淡	2011-07	过度悲观	2013-10	过度悲观
2007-02	乐观	2009-05	乐观	2011-08	过度悲观	2013-11	过度悲观
2007-03	过度乐观	2009-06	乐观	2011-09	过度悲观	2013-12	过度悲观
2007-04	过度乐观	2009-07	乐观	2011-10	过度悲观	2014-01	悲观

续表 4.6 2007 年 1 月-2015 年 12 月投资者情绪状态

2007-05	过度乐观	2009-08	平淡	2011-11	过度悲观	2014-02	悲观
2007-06	乐观	2009-09	悲观	2011-12	过度悲观	2014-03	过度悲观
2007-07	过度乐观	2009-10	平淡	2012-01	悲观	2014-04	过度悲观
2007-08	乐观	2009-11	平淡	2012-02	过度悲观	2014-05	过度悲观
2007-09	乐观	2009-12	悲观	2012-03	过度悲观	2014-06	悲观
2007-10	平淡	2010-01	悲观	2012-04	过度悲观	2014-07	悲观
2007-11	平淡	2010-02	悲观	2012-05	过度悲观	2014-08	平淡
2007-12	乐观	2010-03	悲观	2012-06	过度悲观	2014-09	乐观
2008-01	悲观	2010-04	悲观	2012-07	过度悲观	2014-10	乐观
2008-02	平淡	2010-05	过度悲观	2012-08	过度悲观	2014-11	过度乐观
2008-03	平淡	2010-06	悲观	2012-09	过度悲观	2014-12	乐观
2008-04	悲观	2010-07	悲观	2012-10	过度悲观	2015-01	乐观
2008-05	悲观	2010-08	悲观	2012-11	悲观	2015-02	过度乐观
2008-06	悲观	2010-09	平淡	2012-12	悲观	2015-03	过度乐观
2008-07	悲观	2010-10	平淡	2013-01	悲观	2015-04	过度乐观
2008-08	悲观	2010-11	悲观	2013-02	悲观	2015-05	过度乐观
2008-09	悲观	2010-12	过度悲观	2013-03	过度悲观	2015-06	乐观
2008-10	平淡	2011-01	悲观	2013-04	悲观	2015-07	乐观
2008-11	平淡	2011-02	悲观	2013-05	过度悲观	2015-08	平淡
2008-12	悲观	2011-03	悲观	2013-06	过度悲观	2015-09	乐观
2009-01	乐观	2011-04	过度悲观	2013-07	悲观	2015-10	乐观
2009-02	乐观	2011-05	悲观	2013-08	悲观	2015-11	平淡
2009-03	乐观	2011-06	悲观	2013-09	悲观	2015-12	平淡

对投资者情绪的这五种状态：过度悲观、悲观、平淡、乐观、过度乐观，分别记为 $i = \{1, 2, 3, 4, 5\}$ ， x_i 表示第 i 状态的频数。从表可知，处于过度悲观状态的频数 $x_1 = 27$ ，处于悲观状态的频数 $x_2 = 36$ ，处于平淡状态的频数 $x_3 = 16$ ，处于乐观状态的频数 $x_4 = 20$ ，处于过度乐观状态的频数 $x_5 = 9$ 。因为最后一期状态为平淡，无后续状态，所以 $x_3 = 15$ 。对各个状态的转移频数进行统计，其中过度悲观状态转移到过度悲观状态的频数为 18，则由过度悲观状态转移到过度悲观状态的频率是 $a_{11} = 18/x_1 = 18/27$ 。由过度悲观状态转移到悲观状态的频数为 9，则由过度悲观状态转移到悲观状态的频率是： $a_{12} = 9/x_1 = 9/27$ 。由过度悲观状态转移到平淡状态、乐观状态和过度乐观状态的频数都为 0，因此由过度悲观状态转移到平淡状态、乐观状态和过度乐观状态的频率是：

$a_{13} = a_{14} = a_{15} = 0/x_1 = 0$ 。同理，可求其余状态的转移频率。最后，得到隐状态之间的状态转移概率矩阵，如表 4.7 所示：

表 4.7 投资者情绪状态转移概率

	过度悲观	悲观	平淡	乐观	过度乐观
过度悲观	0.6667	0.3333	0	0	0
悲观	0.2500	0.5833	0.1389	0.0278	0
平淡	0	0.3333	0.4000	0.2667	0
乐观	0	0.0500	0.2500	0.5000	0.2000
过度乐观	0	0	0	0.4444	0.5556

4.3.2 观测概率矩阵

由图 4.5 可知，观测序列上证综指月收益率的分布相对较均匀，本文将其划分为四个区间： $R \leq -10\%$ 为大跌， $-10\% < R < 0$ 为跌， $0 \leq R < 10\%$ 为涨， $R > 10\%$ 为大涨。

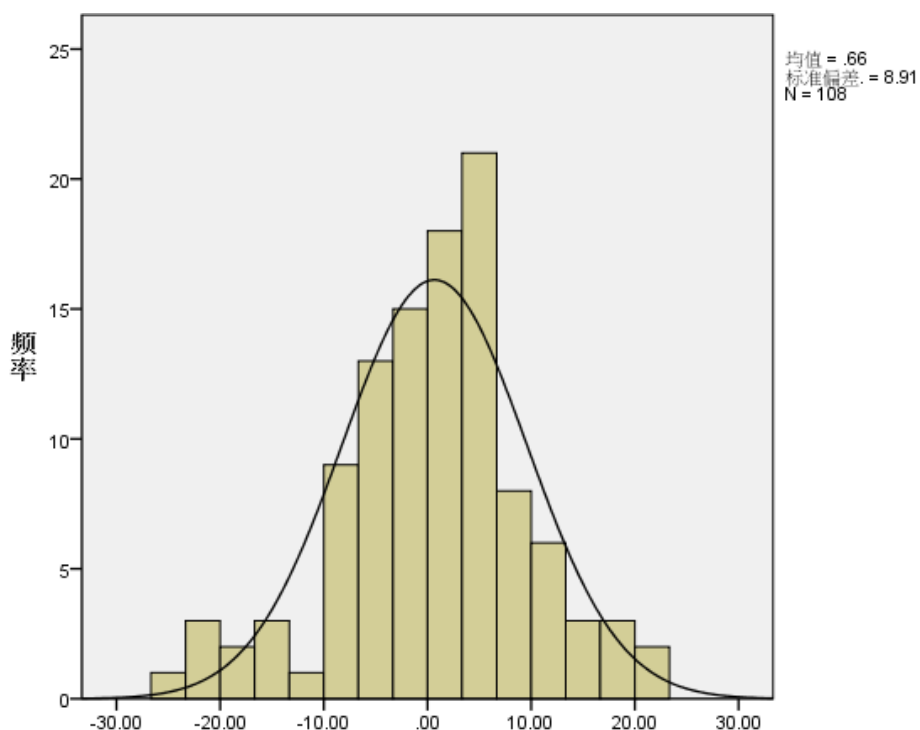


图 4.5 上证综指月收益率分布图

将不同投资者情绪状态与上证综指月收益率的分类一一对应起来，如表 4.8 所示；然后通过统计投资者情绪对应不同情况的月收益率的频率，就可以计算得到观测概率矩阵，如表 4.9 所示。

表 4.8 投资者情绪状态与上证综指月收益率涨跌情况

2007-01	乐观	涨	2009-04	平淡	涨	2011-07	过度 悲观	跌	2013-10	过度 悲观	跌
2007-02	乐观	涨	2009-05	乐观	涨	2011-08	过度 悲观	跌	2013-11	过度 悲观	涨
2007-03	过度 乐观	大涨	2009-06	乐观	大涨	2011-09	过度 悲观	跌	2013-12	过度 悲观	跌
2007-04	过度 乐观	大涨	2009-07	乐观	大涨	2011-10	过度 悲观	涨	2014-01	悲观	跌
2007-05	过度 乐观	涨	2009-08	平淡	大跌	2011-11	过度 悲观	跌	2014-02	悲观	涨
2007-06	乐观	跌	2009-09	悲观	涨	2011-12	过度 悲观	跌	2014-03	过度 悲观	跌
2007-07	过度 乐观	大涨	2009-10	平淡	涨	2012-01	悲观	涨	2014-04	过度 悲观	跌
2007-08	乐观	大涨	2009-11	平淡	涨	2012-02	过度 悲观	涨	2014-05	过度 悲观	涨
2007-09	乐观	涨	2009-12	悲观	涨	2012-03	过度 悲观	跌	2014-06	悲观	涨
2007-10	平淡	涨	2010-01	悲观	跌	2012-04	过度 悲观	涨	2014-07	悲观	涨
2007-11	平淡	大跌	2010-02	悲观	涨	2012-05	过度 悲观	跌	2014-08	平淡	涨
2007-12	乐观	涨	2010-03	悲观	涨	2012-06	过度 悲观	跌	2014-09	乐观	涨
2008-01	悲观	大跌	2010-04	悲观	跌	2012-07	过度 悲观	跌	2014-10	乐观	涨
2008-02	平淡	跌	2010-05	过度 悲观	跌	2012-08	过度 悲观	跌	2014-11	过度 乐观	大涨
2008-03	平淡	大跌	2010-06	悲观	跌	2012-09	过度 悲观	涨	2014-12	乐观	大涨
2008-04	悲观	涨	2010-07	悲观	涨	2012-10	过度 悲观	跌	2015-01	乐观	跌

续表 4.8 投资者情绪状态与上证综指月收益率涨跌情况

2008-05	悲观	跌	2010-08	悲观	涨	2012-11	悲观	跌	2015-02	过度乐观	涨
2008-06	悲观	大跌	2010-09	平淡	涨	2012-12	悲观	大涨	2015-03	过度乐观	大涨
2008-07	悲观	涨	2010-10	平淡	大涨	2013-01	悲观	涨	2015-04	过度乐观	大涨
2008-08	悲观	大跌	2010-11	悲观	跌	2013-02	悲观	跌	2015-05	过度乐观	涨
2008-09	悲观	跌	2010-12	过度悲观	跌	2013-03	过度悲观	跌	2015-06	乐观	跌
2008-10	平淡	大跌	2011-01	悲观	跌	2013-04	悲观	跌	2015-07	乐观	大跌
2008-11	平淡	涨	2011-02	悲观	涨	2013-05	过度悲观	涨	2015-08	平淡	大跌
2008-12	悲观	跌	2011-03	悲观	涨	2013-06	过度悲观	大跌	2015-09	乐观	跌
2009-01	乐观	涨	2011-04	过度悲观	跌	2013-07	悲观	涨	2015-10	乐观	大涨
2009-02	乐观	涨	2011-05	悲观	跌	2013-08	悲观	涨	2015-11	平淡	涨
2009-03	乐观	大涨	2011-06	悲观	涨	2013-09	悲观	涨	2015-12	平淡	涨

表 4.9 观测概率

	大跌	跌	涨	大涨
过度悲观	1/27	19/27	7/27	0
悲观	1/12	13/36	5/9	0
平淡	5/16	1/16	9/16	1/16
乐观	1/20	1/5	9/20	3/10
过度乐观	0	0	1/3	2/3

4.3.3 隐马尔科夫模型的 MATLAB 函数介绍

根据第 3 章介绍的隐马尔科夫模型预测的原理，在确定了主要参数之后，

模型可以通过产生新的观测序列，并计算该序列出现的概率来实现预测。这两个步骤虽然不难理解和实现，但是计算量却比较大。**MATLAB** 是一种高级计算机语言，具有灵活的编程方法和极高的编程效率，运用其包含的隐马尔科夫模型及统计工具箱中的函数，可以简便、快速地实现预测分析。

(1) `hmmgenerate`

根据给定的状态转移概率矩阵和观测概率矩阵，该函数可以生成一个隐马尔科夫模型序列。其调用格式如下：

```
[seq,states]=hmmgenerate(len,TRANS,EMIS)
hmmgenerate(...,'Symbols',SYMBOLS)
hmmgenerate(...,'Statenames',STATENAMES)
```

其中，`len` 是序列的长度；`TRANS` 是状态转移概率矩阵；`EMIS` 是观测概率矩阵；`seq` 返回一个观测序列；`states` 返回一个状态序列。

(2) `hmmdecode`

该函数可以计算给定观测序列的概率。其调用格式如下：

```
PSTATES=hmmdecode(seq,TRANS,EMIS)
[PSTATES,logpseq]=hmmdecode(...)
hmmdecode(...,'Symbols',SYMBOLS)
```

其中，`seq` 是观测序列；`TRANS` 是状态转移概率矩阵；`EMIS` 是观测概率矩阵；`PSTATES` 返回后验概率；`logpseq` 返回观测序列概率的对数值。

4.3.4 基于投资者情绪的隐马尔科夫模型预测实现

本文以投资者情绪状态转移概率矩阵和观测概率矩阵作为模型参数，根据第 3 章所述的预测原理，利用 `hmmgenerate` 函数和 `hmmdecode` 函数，对观测序列，即股市涨跌情况进行预测，输出如下：

表 4.10 2016 年 1 月股市预测结果

时间	seq	logp
2016-01	跌	-0.5285

上述输出表明，对 2016 年 1 月的股市行情进行预测时，其结果是跌，该结果出现的概率是 $e^{-0.5285}$ ，即 58.95%。

表 4.11 2016 年上半年股市预测结果

时间	seq	logp
2016-01	跌	-3.6141
2016-02	跌	
2016-03	跌	
2016-04	跌	
2016-05	跌	
2016-06	跌	

上述输出表明, 对 2016 年上半年的股市行情进行预测时, 6 个月的预测结果全部都是跌, 该结果出现的概率是 $e^{-3.6141}$, 即 2.69%。

表 4.12 2016 年全年股市预测结果

时间	seq	logp
2016-01	跌	-8.5773
2016-02	跌	
2016-03	跌	
2016-04	跌	
2016-05	跌	
2016-06	涨	
2016-07	跌	
2016-08	跌	
2016-09	跌	
2016-10	涨	
2016-11	涨	
2016-12	涨	

上述输出表明, 对 2016 年的股市行情进行预测时, 12 个月的结果依次是跌、跌、跌、跌、涨、跌、跌、跌、涨、涨、涨, 该结果出现的概率是 $e^{-8.5773}$,

即 0.02%。

观察这三次预测情况，可以发现：观测序列概率的对数值随着预测时间段的变长而减小。换言之，利用隐马尔科夫模型对股市进行预测，当只预测未来一期的情况时，该预测结果出现的概率是最大的，达到了 58.95%；当预测未来半年、一年的情况时，预测结果出现的可能性明显降低，分别只有 2.69%、0.02%。对于预测结果的准确性检验，将在下一节的对比分析中进行。

4.4 隐马尔科夫模型预测与马尔科夫链预测的对比分析

如第 2 章所言，隐马尔科夫模型与马尔科夫链的区别就在于前者多了隐状态这一部分，因此，为了突显投资者情绪在预测我国股市方面的作用，本节将先运用马尔科夫链的预测原理对股市行情进行预测，再把两种方法预测出来的结果进行对比。

4.4.1 马尔科夫链预测的实证分析

（1）数据来源

为确保除投资者情绪这一变量之外的因素的一致性，本节数据同样选取 2007 年 1 月至 2015 年 12 月的上证综指月收益率，取自锐思金融数据库。

（2）一步转移概率矩阵

首先，划分上证综指月收益率的状态，根据同样的标准， $R \leq -10\%$ 为大跌， $-10\% < R < 0$ 为跌， $0 \leq R < 10\%$ 为涨， $R > 10\%$ 为大涨，结果如表 4.13 所示。

表 4.13 上证综指月收益率的涨跌情况

2007-01	涨	2009-04	涨	2011-07	跌	2013-10	跌
2007-02	涨	2009-05	涨	2011-08	跌	2013-11	涨
2007-03	大涨	2009-06	大涨	2011-09	跌	2013-12	跌
2007-04	大涨	2009-07	大涨	2011-10	涨	2014-01	跌
2007-05	涨	2009-08	大跌	2011-11	跌	2014-02	涨
2007-06	跌	2009-09	涨	2011-12	跌	2014-03	跌
2007-07	大涨	2009-10	涨	2012-01	涨	2014-04	跌

续表 4.13 上证综指月收益率的涨跌情况

2007-08	大涨	2009-11	涨	2012-02	涨	2014-05	涨
2007-09	涨	2009-12	涨	2012-03	跌	2014-06	涨
2007-10	涨	2010-01	跌	2012-04	涨	2014-07	涨
2007-11	大跌	2010-02	涨	2012-05	跌	2014-08	涨
2007-12	涨	2010-03	涨	2012-06	跌	2014-09	涨
2008-01	大跌	2010-04	跌	2012-07	跌	2014-10	涨
2008-02	跌	2010-05	跌	2012-08	跌	2014-11	大涨
2008-03	大跌	2010-06	跌	2012-09	涨	2014-12	大涨
2008-04	涨	2010-07	涨	2012-10	跌	2015-01	跌
2008-05	跌	2010-08	涨	2012-11	跌	2015-02	涨
2008-06	大跌	2010-09	涨	2012-12	大涨	2015-03	大涨
2008-07	涨	2010-10	大涨	2013-01	涨	2015-04	大涨
2008-08	大跌	2010-11	跌	2013-02	跌	2015-05	涨
2008-09	跌	2010-12	跌	2013-03	跌	2015-06	跌
2008-10	大跌	2011-01	跌	2013-04	跌	2015-07	大跌
2008-11	涨	2011-02	涨	2013-05	涨	2015-08	大跌
2008-12	跌	2011-03	涨	2013-06	大跌	2015-09	跌
2009-01	涨	2011-04	跌	2013-07	涨	2015-10	大涨
2009-02	涨	2011-05	跌	2013-08	涨	2015-11	涨
2009-03	大涨	2011-06	涨	2013-09	涨	2015-12	涨

对于上证综指月收益率的四种状态：涨、大涨、跌、大跌，分别记为 $i = \{1, 2, 3, 4\}$ ， x_i 表示第 i 状态的频数。从表 4.7 可知，处于涨状态的频数 $x_1 = 47$ ，处于大涨状态的频数 $x_2 = 14$ ，处于跌状态的频数 $x_3 = 37$ ，处于大跌状态的频数 $x_4 = 10$ 。因为最后一期状态为涨，无后续状态，所以 $x_1 = 46$ 。对各个状态的转移频数进行统计，其中涨状态转移到涨状态的频数为 20，则由涨状态转移到

涨状态的频率是 $a_{11} = 20/x_1 = 20/46$ 。由涨状态转为大涨状态的频数为 6，则由涨状态转移到大涨状态的频率是： $a_{12} = 6/x_1 = 6/46$ 。由涨状态转为跌状态的频数为 16，则由涨状态转移到跌状态的频率是： $a_{13} = 16/x_1 = 16/46$ 。由涨状态转为大跌状态的频数为 4，则由涨状态转移到跌状态的频率是： $a_{14} = 4/x_1 = 4/46$ 。同理，可求其余状态的转移频率。最后，得到上证综指月收益率 R 的一步转移概率矩阵：

$$A_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} 0.4348 & 0.1304 & 0.3478 & 0.0870 \\ 0.4286 & 0.3571 & 0.1429 & 0.0714 \\ 0.3784 & 0.0811 & 0.4324 & 0.1081 \\ 0.6000 & 0 & 0.3000 & 0.1000 \end{bmatrix}$$

根据上证综指月收益率的一步转移概率矩阵 A_1 ，可得二步转移概率矩阵：

$$A_2 = A_1^2 = \begin{bmatrix} 0.4288 & 0.1315 & 0.3463 & 0.0934 \\ 0.4363 & 0.1950 & 0.2833 & 0.0854 \\ 0.4278 & 0.1134 & 0.3626 & 0.0962 \\ 0.4344 & 0.1026 & 0.3684 & 0.0946 \end{bmatrix}$$

同理，可得 n 步转移概率矩阵：

$$A_n = A_1^n = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}^n$$

(3) 马尔科夫链预测实现

根据第 2 章马尔科夫链预测模型 $A(t) = A(0)A_n = A(0)A_1^n$ ，首先对 2016 年 1 月的涨跌情况进行预测，此时的初始状态 $A(0)$ 为 2015 年 12 月的涨状态，即 $A(0) = [1 \ 0 \ 0 \ 0]$ ，经过计算可以得到如下预测结果：

表 4.14 2016 年 1 月股市预测结果

时间	预测状态				预测结果
	涨 (%)	大涨 (%)	跌 (%)	大跌 (%)	
2016-01	43.48	13.04	34.78	8.7	涨

上述预测结果表明了 2016 年 1 月股市涨跌的概率，其中涨的概率最大，达到了 43.48%，因此以涨作为该期的预测结果。

同理,经过计算可以得到 2016 年上半年和全年的预测结果,如表 4.15 所示:

表 4.15 2016 年股市预测结果

时间	预测状态				预测结果
	涨 (%)	大涨 (%)	跌 (%)	大跌 (%)	
2016-01	43.4800	13.0400	34.7800	8.7000	涨
2016-02	42.8748	13.1470	34.6346	9.3435	涨
2016-03	42.9886	13.0945	34.5696	9.3472	涨
2016-04	42.9932	13.0854	34.5747	9.3467	涨
2016-05	42.9929	13.0831	34.5771	9.3469	涨
2016-06	42.9928	13.0825	34.5777	9.3470	涨
2016-07	42.9928	13.0823	34.5779	9.3470	涨
2016-08	42.9928	13.0822	34.5779	9.3470	涨
2016-09	42.9928	13.0822	34.5780	9.3470	涨
2016-10	42.9928	13.0822	34.5780	9.3470	涨
2016-11	42.9928	13.0822	34.5780	9.3470	涨
2016-12	42.9928	13.0822	34.5780	9.3470	涨

上述预测结果表明了 2016 年全年股市涨跌的概率,12 个月都是涨的概率最大,故都以涨作为预测结果。

观察这 12 个月的预测情况,可以发现:马尔科夫链预测模型的结果在很大程度上受到初始状态的影响。当初始状态为涨时,再加上从涨状态转移到涨状态的概率比较大,就导致了后续的预测结果基本上与第一期所预测出来的涨状态相似,在保留了四位小数点的情况下,才能看出各期预测状态出现的概率之间的细微区别。

4.4.2 预测结果的检验与对比

将隐马尔科夫模型预测的结果、马尔科夫链预测的结果与股市在 2016 年的实际情况进行对比,可以得到运用这两种方法进行预测的准确程度,如表 4.16 所示。

表 4.16 HMM、马尔科夫链预测结果的检验与对比

时间	实际情况	HMM 预测结果	准确程度	马尔科夫链 预测结果	准确程度
2016-01	大跌	跌	不准确	涨	不准确
2016-02	跌	跌	准确	涨	不准确
2016-03	大涨	跌	不准确	涨	不准确
2016-04	跌	跌	准确	涨	不准确
2016-05	跌	跌	准确	涨	不准确
2016-06	涨	涨	准确	涨	准确
2016-07	涨	跌	不准确	涨	准确
2016-08	涨	跌	不准确	涨	准确
2016-09	跌	跌	准确	涨	不准确
2016-10	涨	涨	准确	涨	准确
2016-11	涨	涨	准确	涨	准确
2016-12	跌	涨	不准确	涨	不准确

由表 4.16 可知，运用隐马尔科夫模型对股市进行预测的准确度比较高，有 7 个月的预测结果与实际情况一致，若只划分成涨和跌两种情况，可能预测结果的准确度会更高。但是运用马尔科夫链进行预测的准确度不是很高，即使有准确的结果，也只能说是受到初始状态的影响而碰巧正确，基本不具备可信度。

因此，在保持其他变量一致的情况下，有无投资者情绪对于预测股市是有重大影响的。当把投资者情绪加入到模型当中时，即运用隐马尔科夫模型（HMM），得到的预测结果是更加准确的。这充分体现了投资者情绪在预测股市方面的作用，表明投资者情绪是有助于预测股市的。

4.5 实证结论

本章首先运用偏最小二乘法（PLS）对2007年1月至2015年12月我国股票沪市的投资者情绪进行了测量，由投资者情绪综合指标

$S^{PLS} = 0.137 CCI_t + 0.2865 NIA_t + 0.2437 TA_t + 0.2566 TURN_t - 0.0054 CEFD_t$ 可知，消费者信心指数、新增股票投资者开户数、股票交易量和换手率与投资者情绪正相关，各变量的数值越大，表示投资者情绪越高涨；而封闭式基金折价率与投资者情绪负相关，表示投资者情绪的消极。

然后，在对我国股市进行预测之前，本文考虑到了我国股市是一个“政策市”的情况，于是运用事件研究法分析了政策因素与投资者情绪的关系，实证结果表明，政策出台对成交量和换手率的变动有显著的影响，又由投资者情绪综合指标可知，成交量与换手率是指标的组成部分，因此投资者情绪已然包含了对政策事件的反应，进而在运用投资者情绪对我国股市进行预测时，无需再额外考虑政策因素。

本文在进行预测时，先根据2007年1月至2015年12月的 S^{PLS} 数值，把投资者情绪划分为过度悲观、悲观、平淡、乐观、过度乐观五种状态，并计算得出了投资者情绪状态转移概率矩阵 S ；再根据上证综指月收益率的值将其分类为涨、大涨、跌、大跌四种情况，并计算得出观测概率矩阵；然后以投资者情绪状态转移概率矩阵和观测概率矩阵为模型参数，运用隐马尔科夫模型（HMM）预测原理，辅助以Matlab提供的函数，对2016年1月、上半年和全年的股市涨跌情况进行了预测，输出结果表明：所预测的时间段越长、期数越多，预测结果出现的概率越小。

为了突出投资者情绪在预测方面的作用，在控制其他变量一致的情况下，本文又去掉投资者情绪变量、运用马尔科夫链对2016年全年的股市行情进行了预测，并检验、对比了两种情况下的预测结果。结果表明，运用隐马尔科夫模型对股市进行预测的准确度比较高，马尔科夫链的预测结果由于受到初始状态的严重影响，准确度、可信度都不太高。也就是说，在投资者情绪参与预测股市的情况下，模型能够产生与实际情况更相符的结果，投资者情绪是有助于预测股市的。

第 5 章 结论与展望

5.1 研究结论

通过对投资者情绪进行理论分析和实证分析，本文得到如下几点研究结论：

第一，投资者的情绪之所以能够影响股市，根据情绪-金融决策模型，是因为从投资者情绪出发，它既可以选择“情绪-认知”路径，通过影响根据预期效用理论得到的认知结果，从而影响投资行为；也可以通过“情绪-行为”路径直接影响投资行为，最后，受情绪影响所产生的投资结果就促使股票市场产生了异常的波动。

第二，隐马尔科夫模型（HMM）的隐状态序列可以表示股票市场上看不见的、隐藏的投资者情绪；观测序列可以表示直观的数据；模型的状态转移概率矩阵和观测概率矩阵，可以把情绪和数据连接起来。因此，隐马尔科夫模型在本文中是适合用于预测股市的。

第三，本文引用偏最小二乘法（PLS）构建了投资者情绪综合指标，其数值走势与上证综指月收益率的走势大体一致，体现了情绪对股市的影响。进一步地，本文研究 9 次利好政策事件、9 次利空政策事件后发现，政策公布后股市的成交量和换手率的变动会受到显著地影响，而这两个数据也是投资者情绪综合指标的构成部分，换言之，政策因素对股市的影响已经被包含在了投资者情绪之中，运用投资者情绪对股市进行预测无需再单独考虑政策因素的影响。

第四，本文基于投资者情绪构建了隐马尔科夫模型，并运用 MATLAB 软件提供的函数工具实现了预测，结果表明，所预测的时间段越长、期数越多，预测结果出现的概率越小。进一步地，为了突出投资者情绪对于预测股市所发挥的作用，本文在去掉投资者情绪的情况下，运用马尔科夫链进行了预测。经过对比分析，本文初步发现，在保持其他变量一致的情况下，有无投资者情绪对于预测股市是有重大影响的。当把投资者情绪加入到模型当中时，即运用隐马尔科夫模型（HMM），得到的预测结果是更加准确的。这充分体现了投资者情绪在预测股市方面的作用，表明投资者情绪是有助于预测股市的。

5.2 研究展望

本文基于投资者情绪建立隐马尔科夫模型，并对我国股市的涨跌进行预测，是一次较新的尝试，但在研究过程中受到一些因素的限制，无法尽善尽美，因此，后续的研究可以从以下几个方面进一步完善：

- （1）对投资者情绪影响股市的理论分析可以更深入；
- （2）投资者情绪综合指标的构建方法、构成指标的选择可以更优化；
- （3）应用高频数据，而且隐马尔科夫模型的预测过程可以更加严谨。

参考文献

- [1] Alexander P. Ljungqvist, Vikram , et al. Hot Markets, Investor Sentiment, and IPO Pricing[J]. Journal of Business, 2006, 79 (4) :1667-1702.
- [2] Andrei Shleifer, Richard H. Thaler, Charles M C Lee. ‘Investor Sentiment and Closed-End Fund Puzzle’[J]. Journal of Finance, 1991, 46 (1) :75-109.
- [3] Baker M, Stein J C. Market liquidity as a sentiment indicator [J]. Ssrn Electronic Journal, 2002, 7 (3) :271-299.
- [4] Baker, M. , and Wurgler, J. 2007. Investor Sentiment in the Stock Market [J] . Journal of Economic Perspectives, 21 : 129 - 151
- [5] Barberis, N. , Shleifer, A. , and Vishny, R. A Model of Investor Sentiment [J] . Journal of Financial Economics, 1998,49 : 307 - 343.
- [6] Brown, G. W., and Cliff, M.T. 2004. Investor Sentiment and the Near- term Stock Market [J] . Journal of Empirical Finance, 11 : 1 - 27.
- [7] Chuang, Wu-Jen, Ouyang, et al. The Impact of Investor Sentiment on Excess Returns: A Taiwan Stock Market Case[J]. International Journal of Information & Management Sciences, 2010, 21(1):13-28.
- [8] DeLong, J. B. , Shleifer, A. , Summers, L , and Waldman, R. Noise Trader Risk in Financial Markets [J] . Journal of Political Economy, 1990.98: 703 - 738.
- [9] Dorn D. Does Sentiment Drive the Retail Demand for IPOs?[J]. Journal of Financial & Quantitative Analysis, 2003, 44 (1) :85-108.
- [10] Fisher K L, Statman M. Investor Sentiment and Stock Returns[J]. Financial Analysts Journal, 2000, 2 (2) :11-13.
- [11] Lemmon M, Portniaguina E. Consumer Confidence and Asset Prices: Some Empirical Evidence[J]. Review of Financial Studies, 2006, 19(4):1499-1529.
- [12] Malcolm B, Jeffrey W. Investor Sentiment and the Cross-Section of Stock Returns[J]. Journal of Finance, 2006, 61 (4) :1645-1680.
- [13] Michelle L, William S G. IPO market cycles: bubbles or sequential learning[J]. Journal of Finance, 2002, 57 (3) :1171–1200.
- [14] Qiu L X, Welch I. Investor Sentiment Measures[J]. Ssrn Electronic Journal, 2004, 117 (35) :367-377.
- [15] Ritter J R. The Long-Run Performance of initial Public Offerings[J]. Journal of Finance, 1991, 46 (1) :3-27.

- [16] Shleifer A, Summers L H. The Noise Trader Approach to Finance[J]. Journal of Economic Perspectives, 1990, 4 (2) :19-33.
- [17] 池丽旭,庄新田. 我国投资者情绪对股票收益影响——基于面板数据的研究[J]. 管理评论,2011,06:41-48.
- [18] 池丽旭, 张广胜, 庄新田,等. 投资者情绪指标与股票市场——基于扩展卡尔曼滤波方法的研究[J]. 管理工程学报, 2012, 26(3):122-128.
- [19] 黄德龙,文凤华,杨晓光. 投资者情绪指数及中国股市的实证[J]. 系统科学与数学,2009,01:1-13.
- [20] 黄冉. 基于隐马尔科夫模型 (HMM) 的股票价格预测分析[D].青岛大学,2015.
- [21] 胡淑兰,魏捷,黄晟. 基于 HMM 的中国股市状态转换及预测[J]. 统计与决策,2011,22:152-154.
- [22] 靳光辉,刘志远,花贵如. 政策不确定性、投资者情绪与企业投资——基于战略性新兴产业的实证研究[J]. 中央财经大学学报,2016,05:60-69.
- [23] 赖凯声, 陈浩, 乐国安,等. 情绪能预测股市吗?[J]. 心理科学进展, 2014, 22(11):1770-1781.
- [24] 李静. 基于行为金融学的股票市场投资者行为研究[D].中国社会科学院研究生院,2012.
- [25] 刘维奇,刘新新. 个人和机构投资者情绪与股票收益——基于上证 A 股市场的研究[J]. 管理科学学报,2014,03:70-87.
- [26] 鹿坪,田甜,姚海鑫. 个人投资者情绪、机构投资者情绪与证券市场指数收益——基于 VAR 模型的实证分析[J]. 上海金融,2015,01:65-70.
- [27] 陆剑清, 陈明珠. 投资者情绪对股市波动影响的内在机制探析[J]. 山东社会科学, 2015(S2).
- [28] 苗苗,洪潇,付立新. 基于 Markov 转移模型的中国股市状态转换研究[J]. 统计与决策,2010,12:87-89.
- [29] 王美今,孙建军. 中国股市收益、收益波动与投资者情绪[J]. 经济研究,2004,10:75-83.
- [30] 王明涛,路磊,宋锴. 政策因素对股票市场波动的非对称性影响[J]. 管理科学学报,2012,12:40-57.
- [31]王镇,郝刚. 投资者情绪指数的构建研究——基于偏最小二乘法[J]. 金融理论与实践,2014,(07):1-6.
- [32] 文凤华,肖金利,黄创霞,陈晓红,杨晓光. 投资者情绪特征对股票价格行为的影响研究[J]. 管理科学学报,2014,03:60-69.
- [33] 伍燕然, 韩立岩. 不完全理性、投资者情绪与封闭式基金之谜[J]. 经济研究, 2007 (3) :117-129.
- [34] 吴漫君. 基于隐马尔科夫模型的股价走势预测[D].华南理工大学,2011.

- [35] 阎春宁, 刘瑞辉, 张欢欢. 基于马尔可夫链改进股市的预测精度[J]. 统计与决策, 2016(8):76-79.
- [36] 杨元泽. 封闭式基金的折价能否作为投资者情绪有效衡量——基于深圳股票市场的实证研究[J]. 中央财经大学学报, 2010 (5) :26-31.
- [37] 杨继平, 陈晓暄, 张春会. 中国沪深股市结构性波动的政策性影响因素[J]. 中国管理科学, 2012, 06:43-51.
- [38] 张丹, 廖士光. 中国证券市场投资者情绪研究[J]. 证券市场导报, 2009, 10:61-68.
- [39] 张前程, 杨德才. 货币政策、投资者情绪与企业投资行为[J]. 中央财经大学学报, 2015, 12:57-68.
- [40] 张珍. 基于隐马尔科夫模型的波动率预测[D]. 华南理工大学, 2012.
- [41] 张宗新, 王海亮. 投资者情绪、主观信念调整与市场波动[J]. 金融研究, 2013(4):142-155.

致谢

于我而言，这篇论文，不只是一项毕业任务，更多的是对自己当初放弃工作机会、选择离家读研的一个交待。

完成本文，我最感谢的是导师。感谢他要求我们去听他给本科生的授课，因为这个机会我学习到了隐马尔科夫模型，从而有了论文的选题思路；感谢他组织师门的学术交流活动，从中我接触到了 MATLAB 软件；感谢他关注我的论文进度、对我一再地督促，不然我就不会有时间进行暑假实习，也无法专心找工作，面对突如其来的意外事故也不能安心养伤；感谢他对本文的指导，我有很多版的初稿，很多版的开题报告，无论是假期还是深夜，导师都及时提出宝贵的修改意见、指点迷津；感谢他在每次汇报前都组织试讲，为我们的演说和 PPT 把关，只有经过训练才能出色表现；最后，也感谢他给予的生活上的关心和帮助。

同时，我对为本文提出建设性建议的各位老师也心存感激。