

中山大学

硕士学位论文

基于聚类的羊群效应分析模型

姓名：郭旭超

申请学位级别：硕士

专业：应用数学

指导教师：张磊

20100527

基于聚类的羊群效应分析模型

专业： 应用数学

硕士生：郭旭超

指导教师：张磊(副教授)

摘 要

2007 年底，金融海啸席卷全球，我国股票市场亦不能幸免。短短的一年半时间内，上证指数从 6124 点的高位跌至 1664 点的低谷。市场尚不成熟、市场信息的高度不对称、投资者非理性的投资行为等因素导致我国股票市场经常处于大幅波动之中。极不稳定的股票市场将制约我国经济稳定的发展。

近年来，行为金融学的发展为解析投资市场中的异常现象提供了新的思路。作为行为金融学的一个重要分支，羊群效应从市场信息不对称、投资者心理认知偏差等方面出发，研究市场中非理性的投资行为对市场波动性的影响。

本文在研究市场整体羊群效应的 CH 模型、CCK 模型的基础上，提出基于聚类的羊群效应分析模型。尝试加入聚类分析模型对市场数据作特征提取，找出显著存在羊群效应的时间点，并对相关时间点作进一步的分析，提出有限波动性模型，能有效地解析我国市场波动的实际情况。

基于本文提出的模型，对我国 A 股市场 2006 年到 2009 年间的进行了实证分析。通过有效地对数据进行特征提取，聚类分析的结果将市场波动状况大致分为两类：稳定市场和羊群市场。进一步地，应用统计方法对结果进行了显著性检验，说明了本文的建模能有效地确定羊群效应发生的时间点。此外，对本文提出的有限波动性模型进行求解，结合 GARCH 模型的拟合结果，解析我国股票市场收益率在有限范围内波动的状况。

关键词： 羊群效应，信息不对称，聚类分析，市场波动率

Herd Behavior Analysis Based on Clustering Method

Major: Applied Mathematics

Name: Guo, Xuchao

Supervisor: Zhang, Lei (Associate Professor)

ABSTRACT

During 2007, the Financial Tsunami had great impact on Chinese stock market. The Shanghai Stock Exchange Composite Index (SSE Composite Index) fell seriously from 6124 to 1664. Due to factors such as information asymmetry and irrational finance behavior, Chinese stock market always suffers from heavy volatility. Unstable stock market would probably block the development of economic.

In recent years, the research of Behavior Finance provides a new direction for understanding anomaly of the investment market. As a main branch of Behavior Finance, the study of herd behavior attempts to explain the unusual market volatility that caused by irrational investment decisions from the perspectives of information asymmetry and investors' cognitive psychology.

Based on the CH model and CCK model, this paper adds clustering analysis, which helps extract characteristics from data, to the herding detection process and attempts to figure out when herding would happen. In addition, the paper analyzes the market volatility from the views of herding, which fits Chinese market efficiently.

In this paper, an empirical analysis about herd behavior is done with the A Stock Market Index data from year 2006 to 2009. As a result of the clustering analysis, characteristics of data are extracted effectively, which helps detect herding when it happens. Further information implies that the fluctuation of market volatility is limited and it varies by whether herding happens in the market.

Key Words: Herd Behavior, Information Asymmetry, Clustering, Volatility

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，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作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：2010年5月26日



学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅，有权将学位论文的内容编入有关数据库进行检索，可以采用复印、缩印或其他方法保存学位论文。

学位论文作者签名：

日期：2010年5月26日

导师签名：

日期：2010年5月26日



第一章 绪论

1.1 研究背景

1.1.1 行为金融学

在现代金融理论发展的数十年历史中,金融学家们一直试图将复杂的金融现象抽象为简单的数学模型,通过建立一系列合理的金融模型去理解金融资产的定价、单个理性的投资者如何选择最优的投资组合等问题。现时流行的传统金融理论包括:

- 1.现代资产组合理论(Modern Portfolio Theory, MPT)
- 2.资本资产定价理论(Capital Asset Pricing Theory, CAPM)
- 3.套利定价模型(Arbitrage Pricing Theory, APT)
- 4.期权定价模型(Options Pricing Theory)

这一系列的经典理论均是在以投资者是完全理性的和有效市场假设(Efficient Market Hypothesis, EMH)为公理化假设的基础下,通过一系列严谨的数理逻辑推理建立起来的。这些经典模型为金融学构建了比较完整的框架,并且在很大程度上合理地解释了市场中存在的许多金融现象。

然而,完全竞争的市场和完全理性的投资者等基础性的假设却一直备受质疑。20世纪90年代以来,许多经济学家与实证学家的实证研究发现,有效市场假设并不能解释股票市场中一些股票价格的异常情况。而另一方面,市场投资者的决策行为也并非完全理性、独立,传统的“理性人”假设已经无法解释现实投资者的行为。

于是,将认知心理学、社会学等引入到金融学的研究当中,试图通过人的情

感与认知的偏差来解释投资者决策行为的学科——行为金融学开始逐渐兴起。尽管行为金融学还没有建立一套完整的理论架构，但它突破了传统金融理论框架的局限性，从投资者实际决策的心理出发，将投资者这一影响金融市场的主体加入到其理论的研究当中。这种回归金融行为本源的研究将有助于了解、完善与规范金融市场。而随着心理学、社会学、行为经济学、决策理论等其他相关学科的进一步发展，行为金融学的发展前景将十分广阔。

1.1.2 数据挖掘与现代金融学

长期以来，资产收益率与其波动率之间的关系的研究以及如何寻找高收益、低风险的资产组合一直是现代金融经济学的核心课题。Markowitz的现代资产组合理论尝试用收益率方差来描述风险，进而寻找最优投资组合；Sharpe的CAPM模型则通过个体与市场收益率的协方差描述风险溢价系数；Black、Scholes以标的资产收益的波动率作为参数之一建立微分方程组对期权进行定价。然而，这一系列经典的理论却没有突破收益率波动性是独立同分布的静态性假设限制。

随着信息技术的不断发展，数据挖掘方法使得人们能够利用计算机处理大量数据，提取有效信息。正因如此，数据挖掘理论在金融经济、生物统计等领域中得到了广泛的应用。面对金融市场中大量的交易数据，数据挖掘研究以数理统计的理论为基础，将各类数据与相关的模型结合起来，能够有效地提取的市场中各类重要的信息。

比如说，基于Mandelbrot^[1] (1963)的研究，金融市场的收益率波动会出现一定时间序列的聚集性，数据挖掘中的聚类分析能根据过往的历史数据，通过聚类与时间序列模型相结合的方法提取出收益率波动性的特征，这将更有效地了解不同时期动态的市场风险。

可以预期，随着现代金融理论的深入发展，数据挖掘方法将不断与行为金融学理论相互融合、相互促进，在研究市场整体效率、个体投资者行为等方面提供重要的科学研究途径。

1.1.3 羊群效应

羊群效应(Herd Behavior)也称羊群行为、从众行为,是指原本毫无组织的羊群跟从“领头羊”行动的行为表现。在金融市场中,羊群效应描述了金融市场中由于信息不对称及投资者心理认知偏差所导致的盲目从众、跟风的非理性投资行为。羊群效应的理论是行为金融学理论中的重要分支,它旨在研究投资市场中投资者非理性的投资行为对市场波动性的影响。

Banerjee^[2] (1992)将羊群效应定义为:即便在拥有私有信息的情况下,人们还是会根据别人的行动来作出与私有信息相违背的行动决策。投资市场中这一系列的从众决策将引起“信息瀑布”(Information Cascades),并使市场投资者的决策趋向一致;最终导致资产价格的大幅度波动,并严重偏离正常水平。近十几年来,金融市场中羊群行为的存在性逐渐被实证研究所证实。而羊群效应导致的价格泡沫、对金融市场稳定性的影响等课题也引起了学术界及政府监管部门的广泛关注。

1.2 羊群效应的研究意义和现状

现代行为金融学的理论研究从兴起至今已有20余年的历史。尽管至今还没有形成一套完整的理论体系,但在金融学家的众多研究中,诸如“阿莱悖论”、“一月效应”、“羊群效应”等与有效市场假说相违的实证结论是值得重视的。而一些社会心理学的相关研究已证实人们普遍存在羊群心理。

在风险资产市场上,羊群行为反映了投资者心理和认知的偏差,这种非理性行为对风险资产的价格有着重大的影响。市场上的羊群心理过度发展和蔓延将引起巨大的价格泡沫,投资者将处于极端的心理之中。而无论投资者对投资市场的环境是过度乐观或过度悲观,极端的投资情绪将使市场产生超额的波动性,并可能导致投资市场的崩溃,引起新一轮的金融危机。

1.2.1 羊群效应与我国金融市场

在最近的十多年里，已先后发生多次全球性的金融危机，其中包括 1997 年的亚洲金融风暴、由 2001 年的“911 恐怖袭击”及 2007 年美国的次贷风暴所引起的金融危机。这一系列的金融危机对全球金融市场有着重大的打击，而对发展中国家金融市场的冲击更是毁灭性的。

在全球金融一体化的环境下，我国的金融市场亦不能幸免，投资者纷纷看空股市的走势。在极端的投资环境下，上证与深证两大综合指数先后遭遇多次恐慌性暴跌，证券市场与股民心理都变得极其脆弱。

而 Kindleberger^[3] (1989)发现金融危机的发生与羊群效应的不稳定性有着密切的联系。Chang、Cheng 和 Khurana^[4] (1998)应用 CCK 模型分别对上涨市场与下跌市场的实证研究中发现，羊群效应引起金融市场的大幅度波动在下跌时尤为明显。金融危机导致市场羊群效应，而羊群效应又进一步加剧金融危机，这样的恶性循环将可能对我国尚未完善的金融市场造成致命的打击。

因此，研究金融危机下我国股票市场中的羊群效应，对如何逐步完善我国证券市场的监管政策、提高投资者理性投资意识，从而建立成熟的证券市场等课题的研究都有着重要的意义。

1.2.2 羊群效应的研究现状

羊群效应主要是在承认一些经济学、投资学的基本假设的前提下，针对市场信息不对称、投资者对市场信息的认知偏差及基金经理重视个人声誉和回报等条件下，利用统计决策、Bayesian决策及实证分析等统计方法对市场波动性、风险资产定价等问题的研究。

现时学术界对羊群效应的研究主要有以下几个方面：

1. Banerjee^[2]提出的基于行动序列性的信息瀑布模型

该模型认为投资者的决策是有序的，序列性的决策过程中由于投资者的模仿行为造成了私有信息的损失，并形成信息瀑布流，最终导致总体的羊群效应。

2. Scharfstein和Stein^[5] (1990)、Graham^[6] (1999)等的个人信誉模型

信誉模型认为除了风险资产的收益率外，基金经理在选择投资决策过程中更重视其个人信誉的因素。从个人声誉出发的博弈模型得出的结论是：基金经理选择从众投资将降低错误决策对其声誉的影响。

3. 实证数据分析模型

其中包括研究具体投资者行为的LSV模型^[7] (1992)和研究股市整体价格波动的CH模型^[8] (1995)及CCK模型^[4]。实证模型主要通过市场数据针对羊群行为的存在性、影响的因素等方面做实证研究。

国内关于羊群效应的研究主要是在现有计量模型的基础上作修改并对我国证券市场做分析。其中主要包括：① 宋军、吴冲锋^[9] (2001)用股票收益率分散度指标对我国证券市场做实证研究；② 孙培源、施东辉^[10] (2004)以CAPM模型为基础，检测我国证券市场中真羊群行为；③ 陈浩^[11] (2004)应用LSV方法检验投资基金的羊群模仿模型等。

1.3 本文的结构、内容及创新之处

1.3.1 本文的主要工作及创新之处

鉴于金融危机环境下羊群效应对证券市场的打击可能是致命的,而现有的羊群效应实证分析模型中还存在一定的不足,本文首先对现有关于羊群效应研究的理论及实证分析模型作一次全面的整理,并尝试针对现时关于市场整体羊群效应研究模型中存在的不足之处提出改进的方法,利用我国股票市场 2007 年到 2009 年间 780 个交易日的数据,应用改进模型对我国的股票市场作羊群效应的实证分析。最后,本文还尝试在市场存在羊群效应的情况下,建立有限波动性模型,研究市场中的收益率波动情况。

基于现有的羊群效应研究方法,本文针对我国的股票市场数据做分析研究,其中涉及以下创新之处:

1. 提出在已有的检测市场整体羊群效应的模型基础上加入聚类分析的特征提取过程,使得经过特征提取处理的数据能符合回归模型的应用条件,并且能更有针对性、更有效地检测、了解市场中的羊群效应。

2. 本文还尝试提出一个能较好地反映羊群效应特征的度量指标用于聚类分析,并将该度量指标应用于研究市场收益率波动状态的模型之中。

3. 在分析得到市场存在羊群效应的结果下,结合我国股票市场的波动状况,创新地在羊群市场意义下提出有限波动性模型,并利用我国股票市场数据作了实证拟合分析。

4. 通过有限波动性模型及GARCH模型都验证了我国股票市场有限波动的性质,并且还发现市场中可能存在短期频繁振动的现象。

1.3.2 本文的内容及结构

除了本章绪论以外，本文还包括下列内容：

第二章为关于羊群效应的概述。主要论述了羊群效应的定义和分类，并对现时国内外学术界中的经典文献作了梳理和概述。

第三章为羊群效应的研究模型简介。主要介绍了各种应用较为广泛的针对不同研究方向、研究对象的羊群效应模型。

第四章为基于聚类分析的羊群效应建模。这章提出加入聚类分析的建模思路，并结合原有的模型结构和相关聚类分析的理论，对比讨论基于聚类的羊群效应分析模型的优劣。

第五章为中国股票市场羊群效应的实证分析。此章应用改进后的建模方法，利用中国股票市场的数据作实证分析，并对结果作分析和总结。

第六章为羊群市场有限波动性模型。此章中创新地提出了有限波动性模型，尝试在羊群效应意义下，理解市场收益率波动性。

第七章为全文总结。对本文的研究进行梳理和总结。

第二章 羊群效应的概述

在现实生活中人们往往面临着许许多多的选择，而他们所作出的选择往往会受到周围人群的影响，并做出与之相类似的决定。比如说在选择餐馆就餐时，人们会参考朋友或报刊杂志的推荐作出选择；在购买生活用品时，人们会倾向于选择那些贴有“销售冠军”标签的物品；在面临升学及就业方向的选择时，当前学校里的热门专业、社会中倍受尊敬、收入丰厚的职业通常会受到广泛的关注和青睐。以上所提到的这些例子正是羊群效应在现实生活中的表现。

2.1 羊群效应的定义与分类

2.1.1 理性羊群与非理性羊群

羊群效应或羊群行为一词源于生物学对动物聚群性行为的研究，引申到人类社会中常指人们在选择过程中受到多数人的影响，放弃自身独立的思考判断，并盲目模仿、追随他人决策的行为。

正如前文所提及的，羊群行为在生活中的例子屡见不鲜。由于人们心理认知的偏差往往会导致对自己的决定盲目自信、过度乐观（悲观）、基于少量信息作主观判断等行为现象，并很可能使人们做出非理性的决策。

据此，**非理性羊群定义为**：在决策过程中基于经验或不完全信息，盲目地相互模仿，并做出非理性分析和决策的行为。诸如对某种热门商品的过度追求、投资时不做具体分析，盲目追涨杀跌、从众预测选举或选美结果等行为均属于非理性的羊群行为。非理性的羊群行为会对供求关系产生一定的影响，使商品、风险资产价格偏离正常水平，从而大大地降低市场的效率。

而值得注意的是，尽管还没有准确的严格定义，但在金融投资市场中的羊群

效应可能是理性的,并非完全没有思考的盲目模仿。**Banerjee^[2] (1992)**将羊群效应定义为人们模仿他人的行动,做出与自身私有有效信息相违背的决策行为。**Bikhchandani 和 Sharma^[12] (2001)**也有类似定义,若别人的决策行动使投资者改变了其原来根据私有信息得出的决策行为,则可定义为羊群行为。这些羊群行为是在信息不完全的基础上,基于 Bayesian 准则下的最优决策。这些理性羊群行为应该与盲目模仿的羊群行为有所区别。

现时行为金融学当中,关于羊群行为的理论研究主要是以理性羊群为主体的。总的来说,这些理论中定义的理性羊群效应具有以下一些特点:

1. 投资者模仿市场上其他人的行动做出决策。
2. 羊群效应的产生会导致市场中私有信息的缺失,并使风险资产偏离实际价格。
3. 在市场信息不完全的情况下,投资者的羊群行为是最大化其效用函数(收益或声誉等方面)的理性决策。相对来说,由于实证模型主要以市场整体的数据为基础,因此实证研究大多局限于羊群效应存在性的测量,却很难对理性与非理性作出显著区分。

2.1.2 真实羊群与伪羊群

根据模仿行为是否具有主观性,**Bikhchandani^[12] (2001)**的文章中提出了应对理性羊群作出以下区分:真实羊群(Intentional Herding)和伪羊群(Spurious Herding)。

该文指出**伪羊群应定义为**:基于相似的资料信息的情况下,投资者面对类似问题时所作出相类似的决策行为;而**真实羊群则是指**在信息不对称的情况下,投资者抛弃自身拥有的信息,为最大化其个人期望收益,模仿、跟从别人的投资决策的行为。

根据 **Banerjee^[2] (1992)**、**BHW^[13] (1992)**等的研究显示真实羊群是脆弱的,金融证券市场中的这种羊群效应的形成往往会造成市场有效信息的损失,资产价

格偏离正常水平，市场波动性加剧，并加剧市场效率的下降和扭曲。而伪羊群则可能利用市场信息，抓住市场中的套利机会，对自身投资组合作出调整并从中获利。这种套利模型下的羊群使市场的新信息能被资产价格迅速消化，并重新达到平衡状态。尽管现时的理论研究认为这两类羊群对市场波动性有着截然不同的影响，但区分这两类羊群却有着相当大的困难。

2.2 羊群效应模型的分类及相关文献综述

金融市场中的羊群效应最早由 Keynes^[14] (1936) 在研究资本市场的价格波动时提出的，他借用“选美博弈”来描述投资者之间的相互影响。在“选美博弈”中，参赛者要从 100 张照片中选出 6 中最美照片，若这与全体参赛者的平均选择最相近就能获奖。这样的规则使参赛者都努力地猜测别人的想法，而不一定都选自己认为最美的照片。Keynes 认为投资市场也有类似的情况，股价并不一定完全反映股票内在价值，而是投资者群体心理的产物，股价会因群体心理的突然变化而剧烈波动。

到了二十世纪九十年代，行为金融学的研究开始广泛兴起，包括 Banerjee、BHW、LSV、CH 和 CCK 等从不同角度研究羊群效应的模型纷纷涌现。一般而言，关于金融市场中羊群效应的模型主要分为两大方面——理论模型和数据实证分析研究模型。羊群效应的理论模型又分为信息瀑布模型(Information Cascades)和个人信誉模型(Reputation-Based)；而实证模型则根据研究对象分为市场整体和具体投资者两大方向。

2.2.1 信息瀑布模型

Banerjee^[2] (1992) 建立了一个连续序列决策模型，通过贝叶斯决策分析的方法对羊群行为中的信息瀑布模型进行研究。该模型有效地解释了投资者的决策趋同性问题，并指出这样的羊群效应可能导致超额波动。

Bikhchanbani、Hirshleifer和Welch^[13] (1992)认为, 序列中初始几位投资者的决策将决定整体的决策方向; 而在投资者采取了错误策略的情况下, 过往的经验和新获得的信息将使他们可能采取相反的决策, 并形成新的羊群效应, 从而增加市场的波动性。

Avery、Zemsky^[15] (1998)认为Banerjee模型中固定资产价格的假设并不符合实际情况, 并尝试放宽这样的假设。他们发现: 假如将固定资产价格的假设放宽为资产价格会根据投资者的决策而调整, 则羊群效应不复存在; 但若将信息质量的不确定性加入模型中, 则即便在理性人的假设下, 依然会产生羊群效应。

2.2.2 个人信誉风险模型

Scharfstein和Stein (1990)、Graham (1999)等针对基金经理投资过程中个人信誉方面的问题, 提出了另一种羊群效应模型——个人信誉风险模型。个人信誉风险模型描述的是基金经理将个人声誉归入其投资效用当中, 因而可能作出模仿他人投资决策的羊群现象。

Scharfstein和Stein^[5] (1990)提出关于基金经理信誉的模型, 该模型认为无论基金经理管理投资组合的能力如何, 他们都倾向于选择与其他同行一致的投资, 从而产生羊群效应。而选择与其他同行一致的投资决策是保护基金经理声誉的最佳决策。

Graham^[6]的模型进一步指出, 在基金经理拥有高信誉或低能力、或者市场信息高度一致的情况下, 基金经理都倾向于仿效Value Line来进行投资决策。

Zwiebel^[16] (1995)则通过实证分析基金经理间的相互模仿行为, 并发现为了能得到更高更准确的评价, 除了极少数能力极强或极差的经理外, 其他基金经理都倾向于保守地遵循行业中标准进行投资活动。

2.2.3 数据分析实证模型

羊群效应的数据分析实证模型最具代表性的有：研究具体投资机构行为的 LSV 模型、以股票价格分散度为指标，研究股市整体价格波动的 CH 模型及 CCK 模型等。

Lakonishok、Shleifer、Vishny^[7] (1992)提出的 LSV 模型对美国证券市场中养老金方面的羊群行为作实证研究，分别利用买卖交易的均衡程度来作为度量，并发现市场中的羊群效应并不明显。但 LSV 模型有两个不足之处：

1. 仅用买卖数而非买卖量作度量。
2. 不能反映不同时期的交易状况。

Grinblatt、Titman、Wermers^[17] (GTW)(1995) 在 LSV 模型的基础上作了简单的修正，提出了通过投资组合量的变化测度去克服 LSV 模型的局限性。在对 274 只共同基金(Mutual Funds) 投资组合的买卖量变化情况的研究中发现，共同基金中存在一定程度的羊群效应。

Christie、Huang^[8] (1995)提出的 CH 模型利用个股收益率的横截面标准差 (Cross-Sectional Standard Deviation of Returns, CSSD)作为股市整体羊群性的度量指标，通过市场大幅变化的哑变量与 CSSD 值做回归分析。假如市场中的大幅变化使 CSSD 的值显著下降，则表明此时市场中的个股收益有明显的趋同性，股市中存在羊群效应。尽管用到的计算方法相当简单，然而 CH 模型只能度量股票市场中较强烈的羊群行为，模型的回归的效果往往并不显著。

为了克服 CH 模型在度量市场中小幅度羊群行为方面的不足，Chang、Cheng、Khorana^[4] (1998)提出了 CCK 模型。该模型用横截面绝对偏离度 (Cross-Sectional Absolute Deviation of Returns, CSAD)代替了 CSSD，并证明了在 CAPM 模型的框架下，CSAD 与市场收益率 R 为线性递增关系。当市场中存在羊群效应时，CSAD 与 R 的线性相关性消失，利用简单的回归模型便可以检测到羊群效应。

2.2.4 国内的相关研究现状

进入 2000 年以来,在上述一些早期的外国文献的基础上,我国的学者也开始对羊群效应作相关研究。而这些研究主要集中在对我国股票市场中是否存在羊群效应的数据分析实证研究。

宋军和吴冲锋^[9] (2001)利用 CH 模型,对我国证券市场的羊群行为进行数据分析研究,发现在极端市场收益率的情况下,我国证券市场羊群效应现象明显存在。

孙培源、施东晖^[20] (2002)则用 CCK 模型做实证研究,发现在多变的市场政策、投资者信息不对称等因素的影响下,我国股市存在一定程度的羊群行为,并且在下跌市场中尤为显著,这种羊群行为导致市场的系统风险有所增大。孙培源、施东晖^[10] (2004)发现单纯通过 CH 模型或者 CCK 模型都不能区别出市场中羊群的真伪。因此,该文在 CAPM 模型的框架下,提出用包含 β_{it} 系数的指标代替 CSSD、CSAD 作为羊群行为度量指标,剔除个股信息和市场整体重大信息等因素对羊群行为指标的影响,有效地检测出市场中的真羊群行为。

在研究具体投资者行为方面,宋军和吴冲锋^[21] (2003)对中国股评家的羊群行为进行了实证分析,发现我国股评家有明显的羊群行为。

陈浩(2004)应用 LSV 方法检验发现羊群行为显著存在于投资基金之间,并且与股票流通股规模、股票历史收益率等属性有一定程度的相关性。

第三章 羊群效应的研究模型简介

自上世纪末以来,羊群效应的研究不断深入,包括 Banerjee、Bikhchandani、LSV、CH 等人分别从不同的角度出发得到了一些有效地解释市场上羊群效应的模型。这些模型从研究的方向(理论和实证)、研究的对象(市场整体和机构投资者之间)等方面都有所不同。其中最具代表性的包括 Banerjee 提出的研究信息瀑布流的序列羊群模型、研究基金投资者羊群行为的 LSV 模型及测量市场整体是否存在羊群效应的 CH、CCK 模型。本章将对这些模型作简单的介绍。

3.1 序列羊群模型

Banerjee^[2]在其文章中提出了决策序列性的假设,即投资者在金融市场中的决策过程是有序的,这种序列性同时会导致市场中私有信息的不对称。一般而言,由于序列前面投资者的行动信息对于后面的投资者是公开的,后者将根据序列前面的决策信息去作理性的分析,试图找出市场中的私有信息并从中获利。

然而 Banerjee 的模型却发现,序列中每一位投资者都会根据他所获得的所有信息做理性的分析,并在最大化自身投资收益的前提下,依 Bayesian 准则作出最优的决策。然而,序列决策的过程却会形成信息瀑布流,使市场私有信息丢失,并导致羊群效应的产生。由于序列中的理性决策是建立在不完全的市场信息之中的,因此市场上资产的价格并不完全反映市场中的所有信息,这样的市场是无效的。下面我们尝试用一个简单的模型来直观地了解信息瀑布及羊群效应是如何产生的。

3.1.1 Banerjee 序列羊群模型

假设有大量投资者将决定是否投资某一股票，该投资的收益 V 等概率地分布为 1 或 -1。而每位投资者在决策前都将独立、随机地获得一个该投资收益大小的私有信息，若 $V = 1$ ，则该信息记为 G 信息，其正确、错误的概率分别为 p 和 $1 - p$ ，其中 $p > 0.5$ ；类似地，若 $V = -1$ ，则该信息质量是 B 信息，正确、错误的概率也为 p 和 $1 - p$ 。投资者被随机排列并按顺序地作出决策，后面的投资者可以知道前面投资者的决策结果，但无法知道他们是否获得有效的信息。

值得注意的是，假如市场中有 M 个好的信息 G 和 N 个坏的信息 B。下面分下列情形：

1. 不妨设 $M > N$ ，根据 Bayesian 法则，我们知道该投资收益 V 的后验分布与仅有 $M - N$ 个质量好的信息的后验分布是一致的；
2. 类似地，若 $M < N$ ，则 V 的后验分布与仅有 $N - M$ 个质量坏的信息的后验分布是一致的；
3. 若 $M = N$ ， V 的后验分布是等概率地分布为 1 或 -1。

首先考虑只有一个 G 信息的情况，根据 Bayesian 公式有：

$$\begin{aligned}
 P(V=1|G) &= \frac{P(G|V=1) \cdot P(V=1)}{P(G|V=1) \cdot P(V=1) + P(G|V=-1) \cdot P(V=-1)} \\
 &= \frac{p \times 0.5}{p \times 0.5 + (1-p) \times 0.5} = p > 0.5
 \end{aligned}$$

假如该信号是 B 信息，则有：

$$P(V=1|B) = 1 - P(V=1|G) = 1 - p < 0.5$$

因此，第一位理性投资者 P 必定会根据他所获得的信息选择是否投资，即若信号为 G 则投资，否则不然。对于序列中第二位 Q，由于他知道 P 的决策，他

也等价于知道 P 的私有信息。若 Q 获得一个 G 信息，那么他的期望收益为：

$$\begin{aligned} E[V|G] &= 1 \times P(V=1|G) - 1 \times P(V=-1|G) \\ &= P(V=1|G) - [1 - P(V=1|G)] \\ &= 2 \times P(V=1|G) - 1 = 2p - 1 > 0 \end{aligned}$$

由于 Q 的投资收益为正，Q 同样会选择投资该资产；另一方面，前文提到若好坏信号的数量一样，则 V 的后验分布等概率分布为 1 或 -1，那么 Q 会随机选择投资或不投资。

在投资者 R 选择是否投资前，他能观察到 P、Q 两人的决策选择，这其中包含 4 种情况：

1. P、Q 都投资；
2. P、Q 都不投资；
3. P 投资，Q 不投资；
4. P 不投资，Q 投资。

由于对称性，我们不妨只考虑 1、3 两种情况。

在情况 1 中，P 选择投资表明 P 获得的是 G 信息；而对于 Q，他选择投资可能是因为获得 G 信息(正确概率 $p > 0.5$)或获得坏信息 B 并随机选择，故 Q 倾向于获得 G 信息。因此，R 即使获得坏信息 B，这信息会与 P 的 G 信息的作用相抵消，而 Q 获得的是 G 信息的概率为 $p > 0.5$ ，故无论 R 获得什么信息他都同样会选择投资。

而对于情况 3，R 由 P、Q 的选择可以推测出 P、Q 分别得到 G、B 的信息，这时 V 为 1 或 -1 的先验概率都是 0.5，则 R 处于和 P 一样位置，他会根据自己获得的信息做出决策。

如此类推，可以到一个结论就是：倘若已经决策的序列中选择投资（不投资）的人数比选择不投资（投资）人数多超过两个时，羊群便开始形成，投资者

将不再根据私有信息做出决策。图 3-1 给出了对 Banerjee 的序列羊群决策模型的直观理解：

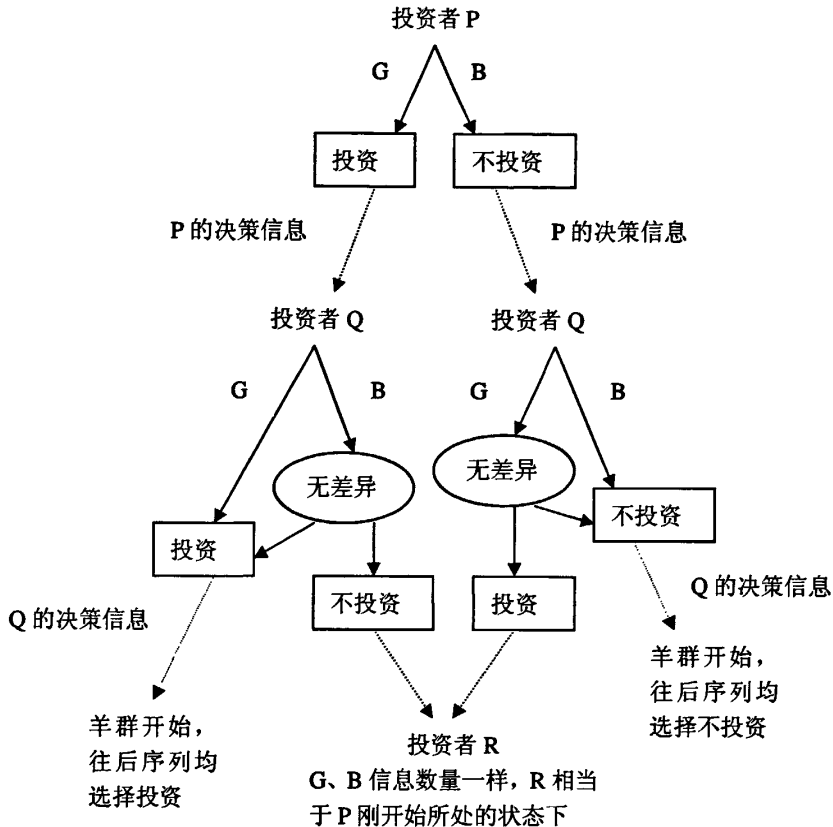


图 3-1： Banerjee 的序列羊群模型

Banerjee 的研究中还发现：无论投资者的样本数量是多少，这种由序列产生的所有决策都失败的概率显著地不为 0，即这样的理性决策可能是无效的。而上述序列性模型还表明，前面少数几位投资者的决策将决定了羊群的走向，而羊群一旦形成，投资者拥有的私有信息将丢失，股票价格将有可能偏离其真实值。由于决策过程中市场信息的大量丢失，这样形成的羊群将十分脆弱，市场中任何新的信息介入都将可能使羊群做出反方向的决策，导致市场的大幅度波动。

此外, Bikhchandani^[12] (2001)关于该序列羊群模型还有下列一些结论:

1. 对于噪声信息 (即信息正确的概率接近 0.5), 信息瀑布在 4(8)个投资决策后形成的概率为 0.93(0.996), 即信息瀑布必然会形成。
2. 所有决策者都选择错误的概率略低于个人独立决策的概率, 即观察其他人的决策行动几乎不能使投资者获益。
3. 若投资者能知道前面所获得的私有信息, 则市场是信息有效的, 每位投资者都能作出最佳的决策。

3.1.2 序列羊群模型的进一步讨论

Avery, Zemsky^[15] (1998) 在他们的研究中指出, 尽管上述过程中投资者的决策在 Bayesian 理论下是最优的, 信息瀑布及羊群序列的形成是 Bayesian 均衡的产物, 但这个在序列决策下产生的市场均衡是无效的。

另一方面, Banerjee、Bikhchandani 等人在建立信息瀑布模型研究羊群效应时加入了一个很强的假设 —— 资产的收益率对于所有人是固定不变的。由于收益率并不随着前面投资者的决策而调整, 那么股票价格便不能完全反映市场上的信息。因此, Avery, Zemsky 在其模型中修改了收益率固定的假设, 试图研究在股票价格反映所有市场信息的情况下, 是否仍然存在羊群效应。

Avery, Zemsky 的模型假设初始的风险资产价格为 0。当投资者 P 决定投资的信息进入市场时, 其私有的 G 信息也同时反映在资产价格之中。由于在一个 G 信息的条件下, V 为 1 的后验分布概率为 p, 根据套利模型的理论(APT), 此时资产价格迅速地调整为其期望收益 $2p - 1$, 套利机会不再存在。

若投资者 Q 持有一个 G 的信息, 在资产价格为 $2p - 1$ 的情况下, 他选择投资的期望收益为:

$$E[V|G] = \frac{2p-1}{p^2 + (1-p)^2} > 2p-1$$

因此他同样选择投资，而资产价格再次作出调整变为：

$$\frac{2p-1}{p^2+(1-p)^2}$$

相反，若 Q 持有的是 B 信息，其期望收益为 0，小于其投资成本 $2p - 1$ 。故 Q 将选择放弃投资，而资产价格会再次调整为 0。对于序列后面的投资者而言，不管前面投资作出哪些决定，其私有的信息 G 或 B 已经决定了他的期望收益与调整后资产价格相比的大小。

因此，序列中每位投资者都将根据其私有信息作出投资决策。风险资产市场的信息是完全对称的，而信息瀑布也将不复存在。理性投资者会根据市场中的所有信息作出最大化自身效用的决定，这样的市场是有效的。

在此基础上，Avery, Zemsky 进一步地将投资者获得信息的能力加入到模型中。假设市场中有 2 种投资者 L 和 H，并且他们获得信息的能力是有差异的。投资者 H 有接近 1 的概率能获得准确的私有信息，而 L 信息准确的概率只有 0.5，并且市场中 H 与 L 人数的比例是未知的。这种情况下，尽管风险资产的价格完全地反映了市场的所有私有信息，而这些信息的质量不确定性将导致风险资产的价格可能错误地反映市场的信息。具体来说，当市场中 L 投资者的比例占大多数时，由于 L 投资者很可能误认为其他投资者都是 H 类型的，那么 L 投资者将放弃其私有信息并跟随其他人的决定，羊群效应随之产生。

总的来说，金融市场中的羊群效应产生之源在于信息的不确定性，这其中包括信息的质量和对于所有投资者是否对称、完全等方面。私有信息在序列模型的传递过程中很可能会丢失，并形成信息瀑布流，最终导致市场信息的不对称。由于后面的投资者并不能完全掌握市场信息，只能透过别人的决策去揣摩丢失的市场信息，最终作出模仿的投资决策。

3.2 机构投资者羊群模型

在研究具体机构投资者羊群行为的模型中,最经典的是 Lakonishok、Shleifer 和 Vishny^[7] (LSV)(1992)提出的 LSV 模型。LSV 提出通过测量一群机构投资者对某一股票买卖的不均衡程度作为衡量市场中羊群效应的指标,并对 1985 至 1989 年间美国 769 家投资机构的交易数据对羊群效应作实证研究。LSV 的实证研究并没有发现市场中存在显著的羊群行为,但在一定程度上证实了小的投资机构更倾向于做出模仿、羊群的投资行为。

3.2.1 LSV 模型

在 LSV 模型中,用作度量羊群行为的指标为 $H(i,t)$, 并有

$$\begin{cases} H(i,t) = |P(i,t) - P(t)| - AF(i,t) \\ P(i,t) = \frac{B(i,t)}{B(i,t) + S(i,t)} \\ AF(i,t) = E[P(i,t) - P(t)] \\ P(t) = E[P(i,t)] \end{cases} \quad (3-1)$$

其中 $P(i,t)$ 是在给定时间 t 内净买入某资产 i 的机构经理的比例; $B(i,t)$ 、 $S(i,t)$ 分别为 t 时间内买入、卖出资产 i 的机构经理的数量; $B(i,t)$ 服从参数为 $p(t)$ 的二项分布。

因而在投资机构中不存在羊群效应、买卖资产行为是相互独立的原假设下,某机构投资者买入资产 i 的概率应为 $P(t)$ 。当买卖总数 $n(i,t) = B(i,t) + C(i,t)$ 很大时, $H(i,t)$ 应显著趋于 0, 否则拒绝原假设,投资机构间买卖资产行为并非相互独立。而对于整个市场的情况,对于所有资产 i 及时间 t 对应的 $H(i,t)$ 作加权平均便可得到一段时期内市场整体的羊群效应度量指标 \bar{H} , \bar{H} 越大说明市场中羊

群效应越明显。

LSV 的实证研究发现, 羊群效应度量指标 $\bar{H}=0.027$ 。即在一个季度内的 100 个机构投资者买卖当中, 同方向只比反方向的交易多 2.7 宗, 市场中并不存在显著的羊群行为。尽管如此, LSV 认为倘若不知道证券市场的交易需求弹性状况, 轻微的羊群效应也可能对资产价格有重大的影响。

3.2.2 LSV 模型的缺陷及 Wermers 修正

应该指出的是, 用 LSV 模型测量市场的羊群效应有以下两大缺陷:

1. LSV 提出的模型只考虑到了买卖的投资机构的数量而非涉及的股票的数量。如不同的机构对同一股票做出不同的操作: 其中总的卖出数量只有 100 股, 而买入数量为 10000 股, 这种情况下市场很有可能存在羊群效应。因此用买卖的股票数量作参数的效果应该会比 LSV 模型好。

2. LSV 模型只能检测出一定时期内市场整体羊群效应的持续性, 而表现为羊群的到底是否一些相同的机构则不能在 LSV 模型中反映出来。

另一方面, 对于 LSV 模型中测量时期 t 的选取也十分重要。若时间间隔选取太短, 由于机构投资者的持股状况并不一定能在该时期内获得。这种情况下, 机构间不太可能存在相互模仿的羊群效应, LSV 模型检测出的可能是投资者对经济基本面普遍乐观的整体行为。若时间间隔选取太长, 期间许多市场信息的变化和新信息的获得都有可能影响到投资者的决策, 因而很难判断相同的操作策略是否羊群行为。

基于上述提及的 LSV 模型的缺陷, Wermers^[17] (1995)提出了在 LSV 模型基础上的一些修正。Wermers 的修正模型用买卖股票的数量作为 $B(i,t)$ 和 $S(i,t)$, 并且进一步地提出买、卖两个度量:

$$BH(i,t)=H(i,t)\left\{p(i,t)>E[p(i,t)]\right\}=H(i,t)\left[p(i,t)>p(t)\right] \quad (3-2)$$

$$SH(i,t)=H(i,t)\left\{p(i,t)<E[p(i,t)]\right\}=H(i,t)\left[p(i,t)<p(t)\right] \quad (3-3)$$

其中 $BH(i,t)$ 表示在 t 时期内买入股票 i 的数量比例 $p(i,t)$ 大于其期望 $p(t)$ 的条件下 $H(i,t)$ 的值； $SH(i,t)$ 则代表在 t 时期内买入股票 i 的数量比例 $p(i,t)$ 小于其期望 $p(t)$ 的条件下 $H(i,t)$ 的值。这两个指标分别度量机构投资者在买入和卖出股票时的羊群效应程度。

Wermers 用修正的 LSV 模型对 1975 至 1994 年间共同基金的买卖数据做了实证分析，并检测出共同基金投资者中存在明显的羊群行为。Wermers 还发现：

1. 羊群效应更显著地出现在那些前一时期内收益率绝对值（即正或负收益）较大的股票。
2. 市场中“追涨”比“杀跌”的羊群行为更加明显。

3.3 整体羊群行为模型

对股市整体羊群行为的实证研究主要通过股市数据对股票收益的分散度指数做回归分析及显著性检验。Christie、Huang(1995) 提出的 CH 模型及 Chang、Cheng、Khorana(1998) 的 CCK 模型是现时研究股市整体羊群行为中比较完整的实证分析模型，他们分别运用这些模型对美国、日本、韩国等股票市场做实证分析，并在不成熟的市场检测到羊群效应的存在。而国内施东辉、孙培源(2004) 指出了 CH、CCK 模型中的一些不足之处并作出改进，该模型在对我国股票市场的实证研究也得到了较好的效果。

3.3.1 CH 模型

Christie、Huang^[8] (1995) 认为市场中若存在显著的羊群行为，那么大部分的投资者的投资买卖交易将有一定的趋同性，因而个股的收益率应该接近股市整体的收益率。故他们在 CH 模型中提出利用个股收益率的横截面标准差 (Cross-Sectional Standard Deviation, CSSD) 作为度量个股收益率与股市整体

收益率差异的指标。其中 $CSSD_t$ 定义为：

$$CSSD_t = \sqrt{\frac{\sum_{i=1}^N (R_{i,t} - R_{m,t})^2}{N - 1}} \quad (3-4)$$

其中 $R_{i,t}$ 表示在 t 时刻市场中某股票 i 的收益率；而 $R_{m,t}$ 表示 t 时刻股市整体的收益率。

CH 模型认为，在资产定价模型的框架下，不同股票之间的相关性强弱之差将使个股收益率呈分散状。在理性、有效的市场中，在股市整体收益率绝对值很高的情况下，个股收益率的分散状态应更加明显，即度量指标 $CSSD_t$ 应该随整体收益率绝对值的增加而增加。相反，若市场中存在羊群效应，投资交易将集中在某些个股、板块之中。此时，相应个股及板块在整体收益率的加权计算中将占较大比例，因而 $CSSD_t$ 将不会很大。又或者市场中交易者都进行相同的买入或卖出操作，股市会整体地涨或跌，个股收益率与整体收益率差异不大，故 $CSSD_t$ 也不大。因此，在市场出现大幅度上涨或下跌时，通过 $CSSD_t$ 值能够测量是否存在羊群效应。

根据 $CSSD_t$ 值与大幅度上涨或下跌的关系 CH 的回归模型为：

$$CSSD_t = \alpha + \beta_1 D_t^H + \beta_2 D_t^L + \varepsilon_t \quad (3-5)$$

D_t^H 和 D_t^L 分别表示市场在 t 时刻出现大幅上涨、下跌时的哑变量。

由上述讨论知道，若发生大幅度涨跌时， $CSSD_t$ 的值随之而下降，则股票价格的变动趋向于股市整体收益的变化，此时可以认为市场中存在羊群效应。相应地在回归模型中，回归值 β_1 和 β_2 应显著为负。

Christie、Huang(1995)利用 CH 模型对美国股票日收益率做羊群效应的实证研究。结果发现美股市场中整体大幅度的价格变动时，个股收益率也呈现出大幅度的分散状态，羊群效应并不存在。

3.3.2 CCK 模型

Chang、Cheng、Khorana^[4] (1998)的 CCK 模型提出了与 CH 模型中的度量指标 CSSD 相类似的收益率横截面绝对偏离差(Cross-Sectional Absolute Deviation, CSAD)作为收益率趋同性的衡量标准。而且 CCK 模型还证明了在资本资产定价模型(Capital Asset Pricing Model, CAPM)的理论下, $CSAD_t$ 与市场收益率 $R_{m,t}$ 线性相关。

CCK 模型将 $CSAD_t$ 定义为:

$$CSAD_t = \frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_{m,t}| \quad (3-6)$$

在 CAPM 模型中, 风险资产的价格 $E(R_{i,t})$ 可以用无风险资产的收益率 R_f 、市场组合的收益率 $E(R_{m,t})$ 及衡量风险的 β_i 系数来表示:

$$E(R_{i,t}) = R_f + \beta_i [E(R_{m,t}) - R_f] \quad (3-7)$$

其中, $E(R_{m,t}) - R_f$ 称为市场的风险溢价; $\beta_i = \frac{\text{Cov}(R_{i,t}, R_{m,t})}{\text{Var}(R_{m,t})}$ 描述的是个股风险溢价与市场风险溢价之比。假如设市场的风险系数 β_m 为个股风险系数 β_i 的平均值:

$$\beta_m = \frac{1}{N} \sum_{i=1}^N \beta_i \quad (3-8)$$

对于 $E[CSAD_t]$ 有,

$$E[CSAD_t] = E\left[\frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_{m,t}|\right] = \frac{1}{N} \sum_{i=1}^N E|R_{i,t} - R_{m,t}| \quad (3-9)$$

另一方面, 由

$$E|R_{i,t} - R_{m,t}| = |\beta_i - \beta_m| \times (E(R_{m,t}) - R_f) \quad (3-10)$$

得到:

$$\begin{aligned} E[CSAD_t] &= \frac{1}{N} \sum_{i=1}^N E[R_{i,t} - R_{m,t}] \\ &= \frac{1}{N} \sum_{i=1}^N |\beta_i - \beta_m| \cdot (E(R_{m,t}) - R_f) \end{aligned}$$

由 $E[CSAD_t]$ 的表达式可知, $CSAD_t$ 的期望与市场组合收益率 $R_{m,t}$ 的期望是线性相关的, 等价地 $CSAD_t$ 与 $R_{m,t}$ 呈线性正相关。

在上一节的讨论中已提到, 当市场中存在显著的羊群效应现象时, 个股收益率趋向于市场收益率。因此, 若市场整体收益率绝对值较大时, 度量指标 $CSAD_t$ 仍偏小, 则 $CSAD_t$ 与 $R_{m,t}$ 之间的线性关系可能不复存在。

在极端情况下, $CSAD_t$ 甚至可能随着 $R_{m,t}$ 绝对值的增大而减少。股票市场整体出现极端收益率的出现可能引致市场中投资者过分的乐观或悲观情绪, 并做出非理性的决策行动, 盲目的追涨杀跌就是典型的例子。因此, CCK 模型除了通过 $CSAD_t$ 与 $|R_{m,t}|$ 之间线性关系外, 还加入了二次项 $R_{m,t}^2$ 来建立回归模型, 检测市场的羊群效应。具体的回归模型如下:

$$CSAD_t = \alpha + \beta_1 |R_{m,t}| + \beta_2 R_{m,t}^2 + \varepsilon_t \quad (3-11)$$

在正常的市场中 $CSAD_t$ 与 $|R_{m,t}|$ 表现为线性关系, β_2 应显著较小; 而若市场中出现羊群现象, 则 β_2 应显著为负。此外, 为了分别检测市场处于上涨或下跌时的羊群效应, 回归模型可进一步分为如下两个:

$$CSAD_t^u = \alpha + \beta_1^u |R_{m,t}^u| + \beta_2^u (R_{m,t}^u)^2 + \varepsilon_t \quad (3-12)$$

$$CSAD_t^d = \alpha + \beta_1^d |R_{m,t}^d| + \beta_2^d (R_{m,t}^d)^2 + \varepsilon_t \quad (3-13)$$

Chang、Cheng、Khorana(1998)的实证结果与 Christie、Huang(1995)的结果相类似, 在如美国、香港等成熟的股票市场中, 羊群效应并不存在。而在对台湾、韩国等未成熟市场的实证研究却发现, 在整体收益率剧烈的时候, 个股收益率的离散度明显减少, 这说明这些投资市场中存在羊群现象。新兴的投资市场中,

投资经验并不丰富的投资者占有相当大的比例。而这部分投资者对市场私有信息的获得及分析处理的能力不高,他们往往将市场整体的收益作为买卖决策的主要参考因素,因此这些市场中更容易出现羊群效应。

总体来说,CH 模型与 CCK 模型均通过度量个股收益分散度的指标与市场极端收益建立回归模型检测市场中的羊群效应。模型中所用到的数据均为市场公开的日交易信息。CH、CCK 模型的优势所用的数据获取更为便捷,而且所研究的时间段选取更加灵活。由于只能检测出市场相对极端的羊群行为,CH、CCK 模型一般应用在市场整体风险的检测、市场的监管规范及完善等方面。而对于轻微羊群现象对市场波动性的影响、潜在的投资机会检测等方面,整体的羊群模型却无能为力。

3.3.3 施东辉、孙培源(2004)的羊群模型

施东辉、孙培源^[10] (2004)指出 CH、CCK 模型能检测出市场整体的羊群效应现象,却不能将这些羊群加以区分。在本文第二章的羊群行为分类中曾提到应将真实羊群与伪羊群区别对待。比如说,市场检测出基本面的改变而同时做出投资策略的调整与投资者模仿地做出相似的决策是有所区别的,只有后者能被认为是真实的羊群效应。这种情况下,CH、CCK 模型均能检测出市场中的羊群效应,然而这样的结论却并不完全正确。据此施东辉、孙培源同样在 CAPM 模型的基础上,提出以 β_i 来表示的市场收益分散度指标:

$$S_i = \frac{1}{N} \sum_{i=1}^N (\beta_{i,t} - 1)^2 \quad (3-14)$$

在 CAPM 模型下,若市场处于正常状态,由于不同股票 $\beta_{i,t}$ 差异显著, S_i 应显著不为 0; 若市场中存在羊群效应的假设下,个股的风险溢价 $r_{i,t}$ 将趋同于 $r_{m,t}$, 则此时对于绝大部分的 i , $E(\beta_{i,t})$ 都趋向于 1, 度量指标 S_i 应显著为 0。故通过对度量指标 S_i 作假设检验便可测出市场中羊群效应是否存在。

与 CH、CCK 模型相比，这样的测量方法好处在于：当大量的市场机构或投资者根据基本面的改变（如市场将出现某重大利好消息）进行投资策略调整时，绝大部分个股的 $\beta_{i,t}$ 将变小；而在 CAPM 模型下总有 $E(\beta_{i,t})=1$ 成立，那么绝大部分 $\beta_{i,t}$ 的变化只能是微小的，否则其期望值也将发生变化。因此，度量指标 S_i 若在某段时间显著为了 0，则可以认为这是真实羊群效应所引起的而并非由于基本面的变化或市场重大消息的公布发生剧烈的变化。

施东辉、孙培源(2004) 选取了 1992 到 2003 年之间我国上海、深圳 A 股市场的数据做实证研究。结果发现，我国股票市场在极端波动的情况时会出现羊群效应；进一步地，这种羊群效应在股市处于上涨的时期表现得更加明显。

第四章 基于聚类分析的羊群效应建模

在此前的研究当中,国内学者对我国市场整体羊群效应的实证研究主要基于上文中已经介绍的 CH 模型、CCK 模型等几个经典模型。其中,宋军和吴冲锋^[21] (2001)利用 CH 模型,对我国证券市场的羊群行为进行实证研究。发现在极端市场收益率的情况下,我国证券市场存在羊群效应。孙培源、施东晖^[20] (2002)则用 CCK 模型做实证研究,同样检测出我国市场存在明显的羊群效应。进一步地,孙培源、施东晖^[10] (2004)在 CAPM 模型的框架下,提出用包含 β_{it} 系数的指标代替 $CSSD_t$ 、 $CSAD_t$ 作为羊群行为度量指标,试图更有效地甄别出市场中的真羊群行为。

然而,上述提到的一些实证研究仅验证羊群行为的存在性,而且只能够探测出较强烈的羊群行为;而对于其形成原因、哪些时期羊群更容易形成等方面有着一定的局限性。而且由于股票市场的数据具有条件异方差的性质,将这些数据应用于回归分析之中似乎并不合适。

有鉴于此,本文提出引入聚类分析的方法对市场数据进行建模分析,根据羊群效应存在时的某些度量指标特征,通过聚类分析进行特征提取,尝试将羊群效应存在的时间点数据聚为一组,并尽量使组内的数据有相近似的方差,使得各组数据能够应用到回归分析之中。然后结合 CH、CCK 的回归模型对可能存在羊群效应的一组数据作进一步的分析,更有效地研究市场中的羊群效应。

4.1 聚类分析模型的概述

所谓的聚类分析通常是指将一个数据集用统计建模的方法划分成若干独立、不相交的组，每个数据属于唯一的一个组，并且组内的数据有类似的特征，而组间的数据则有相对较显著的差异。比如，根据年级、学系、生源地等属性将一群大学生划分成若干组就是一个简单的聚类。

由于聚类的目标在于找出有类似属性的数据并归为一类，如何度量数据之间的相似度是聚类分析的关键问题之一。聚类分析通常用各种距离来量化数据间的差异。由于相似度的度量并不唯一，那么该度量的选择对聚类的精确性和时效性有着至关重要的影响。下文将会对聚类分析的各种分类模型，使用到的各种度量进行简单的介绍。

概括来说，根据聚类分析的方法与目标的不同，最常用的聚类分析算法主要分为两大类：基于划分的聚类算法及层次聚类算法。

下面将简要说明它们的原理^{[18][19]}：

1. 基于划分的聚类算法

顾名思义，基于划分的聚类算法最主要的目标就是尝试找出一个最优的划分将数据集划分成若干个类，使得其中任意一点只属于某一类，并且它与所在类内其他数据之间的相似度较大，而与其他类的数据之间的差异显著。用数学的语言表示就是：

对数据集 $D = \{x_1, x_2, \dots, x_n\}$ 和设定的类集合 $C = \{C_1, C_2, \dots, C_k\}$ ，对于 D 中的每一个 x 都属于唯一的一个类 $C_i (i=1, \dots, k)$ ，记 C_i 的类中心为 Center_i ，则 $\forall x \in C_i$ 都有 $d(x, \text{Center}_i) \leq d(x, \text{Center}_j)$ ，其中 $d(\cdot, \cdot)$ 为表示距离的函数。

目前比较流行的基于划分的聚类算法包括 K 均值算法与 K 中心算法。其主要思想在于通过迭代的方式找出 K 个最优类中心，使得所有数据与相应类中心距离之和达到最小。

K 均值与 K 中心算法的优点在于其算法十分简单，并且具可解性，经过有限次迭代后，其类标序列是收敛的。但值得注意的是 K 均值算法是一种贪心算法，每次迭代能使所有点都分到与类中心距离最小的类之中，保证了解的局部最优性，但却不能保证所得到的解是全局最优的。

2. 层次聚类算法

层次聚类算法是通过对数据集的分裂或凝聚的方法进行聚类的。层次聚类算法与基于划分的聚类算法最大的区别在于，前者通过对给定的数据集进行层次的分解聚类，其最终分成的类数量是根据数据特征决定的；而后者则是预先设定聚类的个数然后进行划分的，是带有明确目标的聚类。现时主要的层次聚类算法有两种：凝聚层次聚类和分裂层次聚类。

凝聚层次聚类是一种自底向上的聚类方法。它首先将数据集中每一个数据作为一类，然后将具有最大相似度或最小距离的两类合并成新的类；然后重复类合并的过程，直至满足某一给定的条件为止，结束聚类过程，并保存聚类结果。绝大部分的层次聚类方法属于这一类，其中包括 BIRCH、ROCK 等方法。

分裂层次聚类与凝聚层次聚类刚好相反，是一种自顶而下的聚类方法。它开始时将整个数据集作为一个类，然后按照一些离散度准则进行分裂，将具有最大离散度的类分成两个新的类，从而达到缩小类内距离的目的；然后重复类分裂过程，直到满足终止条件为止，如：不能再进行分裂或达到某希望的类数量。由于分裂层次聚类的时间复杂度通常比凝聚方法高，应用并不广泛。

由于本文已明确了要将时间序列聚成 3 类，因此选用 $K=3$ 时的 K 均值算法来对数据集进行划分。下面将对基于聚类的羊群效应分析作详细的叙述。

4.2 本文基于聚类分析的羊群效应建模

4.2.1 CH、CCK 模型的局限性

尽管 CH、CCK 等研究市场整体羊群效应的数据分析模型在实证检验中有很好的效果，但这些模型却有一定的局限性：

1. 模型的异方差性。应用多元线性回归方法建模的基本假设为模型中的各个随机误差均服从同方差的正态分布。然而从我国股票市场近年大起大落的情况来看，市场受政府政策变动、金融危机、外围经济因素等影响较大，市场收益率波动明显存在异方差性。异方差的存在对回归模型的估计、检验的准确性都有很大的影响。因此，在此情况下应用回归分析检测羊群效应似乎并不合适。

2. 只反映出市场中羊群效应的存在性。CH、CCK 模型的方法只能给出羊群效应存在与否、是否强烈等简单答案，而对羊群效应存在的时间段、对前后一段时间股票市场造成的影响等问题却无能为力。

为了弥补这方面的不足，本文创新性地提出将聚类分析加入到研究羊群效应的建模方法之中。希望通过对整体收益率序列的特征提取，将显著存在羊群效应及整体收益率特征相似的数据聚为一组，并应用 CH、CCK 模型检验市场羊群效应。由于此时已将市场数据按时间分组，同一组的数据有较稳定的方差，这样可以避免异方差性对 CH、CCK 模型结果的影响。当某一组被检验出显著存在羊群效应时，就成功地确定了羊群效应存在的时间点，为风险评估、政策监管等应用方面提供了科学的依据。

4.2.2 基于聚类分析的羊群效应建模

设 $t=1,2,\dots,T$ 为考察周期, $i=1,2,\dots,N$ 为考虑的股票编号, 记由每只股票的日收益率得到的时间序列为 $X^{(i)}=\{R_{i,t}, t=1,2,\dots,T\}$, $i=1,2,\dots,N$ 。

本文希望研究在哪些时间点会出现羊群效应, 因此尝试引入聚类方法将时间从 $t=1$ 到 T 这个范围内的市场数据划分为三组, 即显著有羊群效应的、显著没有羊群效应的及中间情况 (分别记为 Group1、Group2、Group3)。

不妨设划分到 Group1 的时间点为 $T^{(1)}=\{t_1^{(1)}, t_2^{(1)}, \dots, t_{n_1}^{(1)}\}$, 划分到 Group2 的时间点为 $T^{(2)}=\{t_1^{(2)}, t_2^{(2)}, \dots, t_{n_2}^{(2)}\}$, 划分到 Group3 的时间点为 $T^{(3)}=\{t_1^{(3)}, t_2^{(3)}, \dots, t_{n_3}^{(3)}\}$; 其中 $T=n_1+n_2+n_3$ 。则相应的 Group1、Group2、Group3 中的收益率数据为:

$$\text{Group1}=\{R_{i,t} \mid i=1,2,\dots,N, t \in T^{(1)}\}$$

$$\text{Group2}=\{R_{i,t} \mid i=1,2,\dots,N, t \in T^{(2)}\}$$

$$\text{Group3}=\{R_{i,t} \mid i=1,2,\dots,N, t \in T^{(3)}\}$$

本文研究的重点是将显著有羊群效应和显著没有羊群效应的时间点找准, 直接利用时间序列数据来聚类是无效的。这时便需要找出能够刻画羊群效应的特征信息。只有通过提取有效的特征信息, 才能产生有效的分组, 达到找准显著有羊群效应和显著没有羊群效应时间点的目标。本文所提出的描述市场羊群效应的度量指标的设计将在下一节中介绍。下面先给出基于聚类分析进行羊群效应分析的详细过程:

1. 找出 t 时刻的市场信息的特征。即由 t 时刻的各股票的收益率数据 $\{R_{1,t}, R_{2,t}, \dots, R_{N,t}\}$ 产生的市场特征向量 $F_{(t)}=[f_{t,1}, f_{t,2}, \dots]$, 其中 $t=1,2,\dots,T$ 。

2. 对 $t=1,2,\dots,T$ 这 T 个时间点对应的特征向量利用 K 均值方法进行聚类, 划分为三组: C_1 、 C_2 、 C_3 .

3. 分别设:

$$C_1 = \{F_{(t)}, t \in T^{(1)} = \{t_1^{(1)}, t_2^{(1)}, \dots, t_{n_1}^{(1)}\}\}$$

$$C_2 = \{F_{(t)}, t \in T^{(2)} = \{t_1^{(2)}, t_2^{(2)}, \dots, t_{n_2}^{(2)}\}\}$$

$$C_3 = \{F_{(t)}, t \in T^{(3)} = \{t_1^{(3)}, t_2^{(3)}, \dots, t_{n_3}^{(3)}\}\}$$

则 C_1 组中的时间点对应的各股票的收益率数据为 Group1; C_2 组中的时间点对应的各股票的收益率数据为 Group2; C_3 组中的时间点对应的各股票的收益率数据为 Group3。

4. 分别对 Group1、Group2、Group3 中的数据应用 CH、CCK 模型检验市场的羊群效应, 其中包括模型参数的求解和显著性检验。

为了使聚类产生的结果有效, 即 Group1 中显著存在羊群效应、Group2 中显著不存在羊群效应, 市场特征向量 $F_{(t)}, t=1,2,\dots,T$ 的求解是很关键的。下一节将给出本文的设计方法及相关分析。

下面给出基于聚类分析的羊群效应建模的流程图, 见图(4-1):

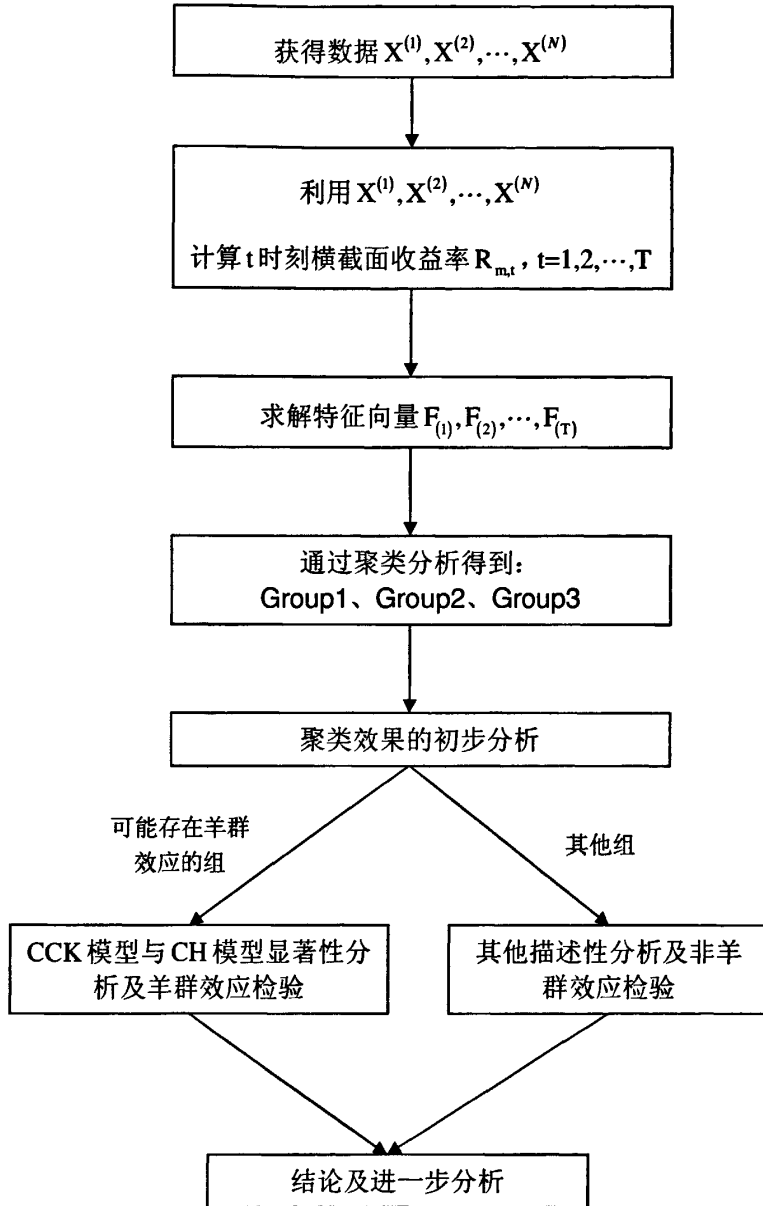


图 4-1：基于聚类分析的羊群效应建模流程图

4.3 本文提出描述羊群效应的度量指标

由于聚类分析过程希望通过聚类的方式提取可能出现羊群效应的时间点，这样就需要有能够描述羊群效应特征的度量指标作为聚类的属性。本文创新地引入了与变异系数有关的度量指标，希望能反映出 $CSSD_t$ 及 $CSAD_t$ 的特性用于描述羊群效应。

我们首先回顾一下度量指标 $CSSD_t$ 与 $CSAD_t$ 是如何定义的：

$$CSSD_t = \sqrt{\frac{\sum_{i=1}^N (R_{i,t} - R_{m,t})^2}{N-1}}, \quad t=1,2,\dots,T$$

$$CSAD_t = \frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_{m,t}|, \quad t=1,2,\dots,T$$

度量指标 $CSSD_t$ 用相当于标准差的形式定义了 t 时刻个股收益率与整体收益率的差异程度，而 $CSAD_t$ 系数则用两者之差的绝对值来定义。本文提出了包含 t 时刻的横截面变异系数 $C_t = \sigma_t / \mu_t$ 的度量指标 H_t ：

$$H_t = \sum_{i=1}^N |R_{i,t} - R_{m,t}| \cdot |C_t| \quad (4-1)$$

$$\text{其中 } \mu_t = E(R_{i,t}) = R_{m,t}, \quad \sigma_t^2 = \frac{\sum_{i=1}^N (R_{i,t} - R_{m,t})^2}{N}, \quad t=1,2,\dots,T.$$

由于羊群效应显著出现在市场绝对收益率 $|R_{m,t}|$ 较大的时候，值较小的 $|R_{m,t}|$ 会令 H_t 变大；值较大的 $|R_{m,t}|$ ，相应的 H_t 值则较小。另一方面，羊群效应是否存在还取决于个股收益率的趋同性。CH、CCK 模型中使用到的度量指标 $CSSD_t$ 与 $CSAD_t$ 能很好地区分个股收益率的趋同程度。若个股收益率有一致的趋同性，

则 $CSSD_t$ 与 $CSAD_t$ 的值均较小。度量指标 H_t 包含了 $CSSD_t$ 与 $CSAD_t$ 的乘积，加上较大的 $|R_{m,t}|$ 值影响，羊群效应显著存在时其值应该较小；相反，若 H_t 比较大，则羊群效应并不显著。此外，将表达式(4-1)稍微变形便得到：

$$H_t = \sum_{i=1}^N \left| \frac{R_{i,t} - R_{m,t}}{R_{m,t}} \right| \cdot \sigma_t = \sum_{i=1}^N \left| 1 - \frac{R_{i,t}}{R_{m,t}} \right| \cdot \sigma_t \quad (4-2)$$

在上一章的介绍中提到了施东辉、孙培源(2004)在 CAPM 模型的基础下推导出羊群效应存在时 β_i 系数趋于 1 的结论。而 β_i 系数用个股与市场风险溢价之比来定义，当无风险收益率较小时， β_i 近似于个股收益与市场收益的比值。度量指标 H_t 在(4-2)式中的表达形式正好能够反映 β_i 是否趋于 1 特点，从而甄别羊群性的真伪。当 H_t 特别小时，则可以认为羊群是真实羊群。

这样，便得到了上一节所涉及的市场特征向量：

$$F_{(t)} = \left[H_t, |R_{m,t}| \right], \quad t=1,2,\dots,T$$

其中， H_t 是本文所提出的能刻画羊群效应的度量指标，其定义见(4-1)； $|R_{m,t}|$ 是 t 时刻市场整体的收益率。

第五章 中国股票市场羊群效应的实证分析

进入二十一世纪以来,随着中国经济的高速发展,金融市场也逐渐庞大起来。然而,我国证券市场在制度建设与完善方面等却不能完全跟上金融市场发展的步伐,股票市场经常受到政策变动的影响。另一方面,各种各样的投资者纷纷涌入迅速膨胀的证券市场,由于投资者素质的参差,市场经常处于大起大落的波动状态之中。

正是由于市场管理、政策干预频繁、投资者素质等因素的影响,我国的股票市场容易形成泡沫,投资者对市场过度乐观等极端情绪。这样的投资环境中更容易产生羊群效应。

而且,我国股票市场的投资者以散户为主,市场中大型机构对市场的炒作现象也十分严重。一旦大规模的经济危机再次出现,大部分跟风投资的“羊群”将损失惨重,尚未成熟的投资市场也可能瞬间崩毁。

因此,研究我国股票市场整体的羊群效应、深入了解羊群现象的形成将有利于逐渐巩固、完善投资市场。下面将应用第四章讨论的基于聚类分析建模方法对我国股票市场做实证分析。

5.1 数据的选取与处理

为了研究中国股票市场中羊群效应的存在性及其存在的时间段,本文选取上海证券交易所中计算入上证指数的 200 只股票(具体见附录)进行实证分析,样本选取的时间为 2006 年 7 月 31 日到 2009 年 10 月 13 日,共 780 个交易日。为了保持数据的准确性、有效性,本文尽量剔除异常样本。

利用股票的日收盘价格数据,计算下列信息:

①每个股市交易日 t 的个股收益率 R_{it} , $t=1, \dots, T$; $i=1, \dots, N$

② 整体收益率 $R_{m,t}$, $t=1, \dots, T$;

③ 个股收益的横截面标准差 $CSSD_t$, $t=1, \dots, T$;

④ 个股收益的横截面绝对偏离差 $CSAD_t$, $t=1, \dots, T$;

⑤ 本文提出的度量指标 H_t , $t=1, \dots, T$

⑥ 最后得到描述 t 时刻市场特征的特征向量 $F_{(t)} = [H_t, R_{m,t}]$, $t=1, \dots, T$

此外, 为了对数据做标准化处理, 这里还计算了上述统计量的均值和标准差。

其中涉及的计算公式如下: ($i=1, \dots, N$, $t=1, \dots, T$)

$$R_{i,t} = \ln \left(\frac{P_{t,i}}{P_{t-1,i}} \right) \quad (5-1)$$

$$R_{m,t} = \frac{1}{N} \sum_{i=1}^N R_{i,t} \quad (5-2)$$

$$CSAD_t = \frac{1}{N} \sum_{i=1}^N |R_{i,t} - R_{m,t}| \quad (5-3)$$

$$CSSD_t = \sqrt{\frac{\sum_{i=1}^N (R_{i,t} - R_{m,t})^2}{N - 1}} \quad (5-4)$$

$$H_t = \sum_{i=1}^N |R_{i,t} - R_{m,t}| \cdot |C_t| \quad (5-5)$$

$$C_t = \sigma_t / \mu_t \quad (5-6)$$

(5-1)中 $P_{t,i}$ 为交易日 t 个股 i 的价格; (5-2)中 N 为投资组合中股票的数量; (5-6)

式中 σ_t 、 μ_t 表达式分别为 $\mu_t = E(R_{i,t}) = R_{m,t}$, $\sigma_t^2 = \frac{\sum_{i=1}^N (R_{i,t} - R_{m,t})^2}{N}$

5.2 K 均值聚类方法的原理和实现

根据第四章的讨论, 本文将应用 K 均值算法对所有的市场数据作聚类, 希望将样本数据聚成若干有相似属性特征的数据集, 通过对可能存在羊群效应的组, 即其中度量指标 $CSSD_t$ 、 $CSAD_t$ 及 H_t 值较小而 $R_{m,t}$ 偏大的 Group1 数据, 应用 CCK 模型进行回归分析并检验相应的回归系数显著程度, 以便有效地检测中国股票市场中是否存在羊群行为。

K 均值聚类算法的基本流程是:

1. 先在所有样本中随机地找出 k 个中心, k 表示类的数量, 它是一个预设值;
2. 对非中心的样本依次计算与 k 个中心的距离, 并将该样本归到中心与其距离最短的类之中, 从而获得类标;
3. 所有点都被分配到唯一的类后, 用同一类内所有点的均值向量作为中心, 并重复步骤 2-3 直至类标序列收敛

K 均值算法的核心是距离度量的选取。假设我们有统计数据 $x(x_1, x_2, \dots, x_n)$ 和 $y(y_1, y_2, \dots, y_n)$, 其中 x_i 和 y_i ($i=1, \dots, n$) 为数据 x, y 的 n 个属性, 那么度量他们之间差异的最简单直接的量就是欧氏距离:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

其他度量方法还包括: Minkowski 距离、最大值度量及 Pearson 相关系数等。本文中选择的常用于 K 均值算法的欧氏度量。一旦定义了距离的度量就可以描述数据之间的差异程度, 从而确定某数据点与类中心的差异, 并找到其所属的类。

假设有 m 个样本， K 均值算法实现的伪码具体如下：

```
{
    随机选取  $k$  个类中心  $Center_i, i=1, \dots, k$ 

    Do while 类标序列不收敛
        For  $j = 1$  to  $m$ 
            For  $i = 1$  to  $k$ 
                计算样本  $j$  与第  $i$  个类中心的距离  $d(i, j)$ 

            Next
            样本  $j$  的类标为  $c_j = \{l | d(j, l) = \min d(i, j), i = 1, 2, \dots, k\}$ 

        Next
        For  $i = 1$  to  $k$ 
            用各类的均值向量作为新的类中心  $Center_i$ 

        Next
    Loop
}
```

将 K 均值算法应用于羊群效应分析时，有几方面是值得注意的：

1. 聚类个数的选取

由于在 K 均值聚类分析的算法中，要将数据分为 K 个类，而这个 K 值是预先选取的， K 值的大小将对聚类以致羊群行为检测的效果有重要的影响。考虑到股票市场的特点，本文的实证研究中选取 $K = 3$ ，即尝试将样本数据聚为 3 大类：

A. 收益率正常、不存在羊群效应的稳态市场。这类数据应该符合传统金融、统计学的理论，即股市的总体的对数收益率服从正态分布，而且个股收益率分散，并没有显著的趋同性。

B. 存在羊群效应的样本数据类。此类数据的整体绝对收益率 $|R_{m,t}|$ 较大，而个股收益率有趋同性，即度量指标 $CSSD_t$ 、 $CSAD_t$ 较小。

C. 整体收益率偏大但不存在羊群效应。整体绝对收益率 $|R_{m,t}|$ 与 B 类数据类似,不同的是个股收益明显分散,即度量指标 $CSSD_t$ 、 $CSAD_t$ 较大。

2. 样本数据属性的选取

由于聚类的效果还取决于用于聚类的属性是否有效地区分不同的类,鉴于羊群效应存在时的数据特点,这里选取本文提出的度量指标 H_t 及市场绝对收益率 $|R_{m,t}|$ 作为样本属性,即对 5.1 中求解出的特征向量 $F_{(t)} = [H_t, |R_{m,t}|]$, $t=1, \dots, T$ 进行聚类。

3. 数据标准化

由于不同类型的数据在量纲方面可能有显著差异,因此在用 K 均值数据作聚类分析之前,应对数据作简单的标准化处理。

5.3 实证分析中涉及其它统计方法的简介

本节主要对在结果分析中应用到的多元线性回归方法、正态性检验中涉及的方法及统计量作简单的介绍^{[23] [24] [25] [26]}。

5.3.1 多元线性回归

本文在探讨羊群效应时用到多元回归的方法,其中便包括对 CH 模型(5-7)、CCK 模型(5-8)中系数的估计及其显著性的分析,因此有必要在本节对多元线性回归作一些简单介绍。

$$CSSD_t = \alpha + \beta_1 D_t^H + \beta_2 D_t^L + \varepsilon_t \quad (5-7)$$

$$CSAD_t = \alpha + \beta_1 |R_{m,t}| + \beta_2 R_{m,t}^2 + \varepsilon_t \quad (5-8)$$

可以看出 CH 模型、CCK 模型的求解是多元回归方程的求解过程，下面简述一般多元回归建模的要点。

一般设因变量 y 与自变量 x_1, x_2, \dots, x_k 之间有关系式：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (5-9)$$

抽样得到 n 组样本数据：

$$(y_1; x_{11}, x_{21}, \dots, x_{k1})$$

$$(y_2; x_{12}, x_{22}, \dots, x_{k2})$$

.....

$$(y_n; x_{1n}, x_{2n}, \dots, x_{kn})$$

其中 x_{ij} 是自变量 x_i 的第 j 个观测值, y_j 是因变量 y 的第 j 个值, 代入模型(5-9)得到求解回归系统的下列方程组：

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \dots + \beta_k x_{k1} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \dots + \beta_k x_{k2} + \varepsilon_2 \\ \dots \quad \dots \quad \dots \\ y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_k x_{kn} + \varepsilon_n \\ \varepsilon_1, \varepsilon_2 \dots \varepsilon_n \sim N(0, \sigma^2) \end{cases} \quad (5-10)$$

模型(5-10)称为 k 元正态线性回归模型, 其中 $\beta_0, \beta_1, \dots, \beta_k$ 及 σ^2 都是未知待估计的参数。

下面将从三个方面来讨论回归方程的求解和有效性检验问题。

1. 参数估计问题

对于 k 元正态线性回归模型(5-10)一般采用最小二乘法估计回归系数 $\beta_0, \beta_1, \dots, \beta_k$, 称使得表达式(5-11):

$$Q(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})]^2 \quad (5-11)$$

达到最小的 $\beta'_0, \beta'_1, \dots, \beta'_k$ 为参数 $(\beta_0, \beta_1, \dots, \beta_k)$ 的最小二乘估计, 最小二乘估计就是如下方程组的解:

$$\begin{cases} l_{11}\beta_1 + l_{12}\beta_2 + \dots + l_{1k}\beta_k = L_{1y} \\ l_{21}\beta_1 + l_{22}\beta_2 + \dots + l_{2k}\beta_k = L_{2y} \\ \dots \quad \dots \quad \dots \quad \dots \\ l_{k1}\beta_1 + l_{k2}\beta_2 + \dots + l_{kk}\beta_k = L_{ky} \\ \beta_0 = \bar{y} - (\beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \dots + \beta_k \bar{x}_k) \end{cases} \quad (5-12)$$

其中:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x}_i = \frac{1}{n} \sum_{i=1}^n x_{ii}, \quad (i=1, 2, \dots, k)$$

$$l_{ij} = \frac{1}{n} \sum_{i=1}^n (x_{ii} - \bar{x}_i)(x_{ji} - \bar{x}_j) = l_{ji}, \quad (i, j=1, 2, \dots, k)$$

$$L_{iy} = \frac{1}{n} \sum_{i=1}^n (x_{ii} - \bar{x}_i)(y_i - \bar{y}), \quad (i=1, 2, \dots, k)$$

一般称方程组(5-12)为正规方程组, 其中前 k 个方程的系数矩阵记为 $L^* = (l_{ij})_{k \times k}$, 则当 L^* 可逆时, 上述正规方程组有解, 便可得到 $\beta_0, \beta_1, \dots, \beta_k$ 的最小二乘估计 $\beta'_0, \beta'_1, \dots, \beta'_k$:

$$\begin{pmatrix} \beta'_1 \\ \vdots \\ \beta'_k \end{pmatrix} = (L^*)^{-1} \begin{pmatrix} L_{1y} \\ \vdots \\ L_{ky} \end{pmatrix} \quad (5-13)$$

$$\beta'_0 = \bar{y} - (\beta'_1 \bar{x}_1 + \dots + \beta'_k \bar{x}_k) \quad (5-14)$$

2. 回归方程的显著性检验问题

多元回归分析是在 y 与 x_1, \dots, x_k 之间呈现线性相关的前提下进行的, 所

求得的回归方程是否有显著意义,还需要对 y 与各个 x_i 间是否存在线性相关关系作显著性假设检验。对 $y' = \beta'_0 + \beta'_1 x_1 + \dots + \beta'_k x_k$ 是否有显著意义问题,可以通过检验下列原假设 $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 来实现,拒绝原假设则说明回归显著。

假设检验 H_0 的检验统计量可以通过对总偏差平方和 L 得到:

$$\begin{aligned} L &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - y'_i + y'_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - y'_i)^2 + \sum_{i=1}^n (y'_i - \bar{y})^2 = Q_e + U \end{aligned}$$

其中 $y'_i = \beta'_0 + \beta'_1 x_{1i} + \dots + \beta'_k x_{ki}$; 而 $U = \sum_i (y'_i - \bar{y})^2$, $Q_e = \sum_i (y_i - y'_i)^2$ 分别称为回归平方和、残差平方和。

在 H_0 成立下, 应有: $\frac{U}{\sigma^2} \sim \chi^2(k)$, $\frac{Q_e}{\sigma^2} \sim \chi^2(n-k-1)$, 而且由 Q_e 与 U 相互独立, 因此得到统计量 F :

$$F = \frac{U/k}{Q_e/(n-k-1)} \sim F(k, n-k-1) \quad (5-15)$$

而相应原假设的拒绝域为:

$$\{F > F_{1-\alpha}(k, n-k-1)\}$$

通过 F 检验得到回归方程有显著意义, 只能说明 y 与 x_1, x_2, \dots, x_k 之间存在显著的线性相关关系, 而衡量经验回归方程与观测值之间拟合好坏程度的常用统计量有复相关系数 R 及拟合优度系数 R^2 。

分别定义复相关系数 R 及拟合优度系数 R^2 为:

$$R^2 = \frac{U}{L} = 1 - \frac{Q_e}{L}, \quad |R| = \sqrt{1 - \frac{Q_e}{L}}$$

R 是观测值 y_1, y_2, \dots, y_k 与回归值 y'_1, y'_2, \dots, y'_k 的相关系数。在实际应用中, 为

了消除自由度的影响，可以定义修正的拟合优度系数：

$$\bar{R}^2 = 1 - \frac{Q_e / (n - k - 1)}{L / (n - 1)} \quad (5-16)$$

由于总的偏差平方和 L 是固定值，回归平方和 Q_e 越小，则 \bar{R}^2 越接近 1，说明拟合度越高；而 \bar{R}^2 越接近 0 则拟合度越低。需要注意的是，修正的拟合优度系数 \bar{R}^2 可能为负，此时取 $\bar{R}^2 = 0$ 。

3. 回归系数的显著性检验问题

对于多元线性回归方程，还需要对每个变量 x_j 做显著性检验。下面考虑用原假设 $H_{0j}: \beta_j = 0$ ，并构造显著性统计量 t ，对回归系数进行显著性 t 检验。

由于回归模型(5-9)假设了误差项 $\varepsilon_j \sim N(0, \sigma^2)$ ，而回归方程中 β'_j 与 ε_j 呈线性相关关系，故 β'_j 也服从正态分布，并且有 $\beta'_j \sim N(\beta_j, \sigma^2 c_{jj})$ ，利用方程组 5-12 中定义的系数矩阵 $L^* = (l_{ij})_{k \times k}$ 可以得到 $\sigma^2 c_{jj}$ 的无偏估计 $S^2(\beta'_j)$ ：

$$S^2(\beta'_j) = \frac{Q_e}{n - k - 1} c_{jj} \quad (5-17)$$

其中 c_{jj} 为矩阵 $((L^*)^T L^*)^{-1}$ 中对角线上的第 j 个元素。由此构造 t 统计量：

$$t = \frac{\beta'_j}{S(\beta'_j)} \sim t(n - k - 1) \quad (5-18)$$

相应原假设的拒绝域为：

$$\{|t| > t_{1-\alpha/2}(n - k - 1)\}$$

检验过程中，对每个变量 x_j 做显著性 t 检验，若拒绝原假设则说明该变量显著不为 0；否则该变量显著为 0，即对因变量 y 影响并不显著。

5.3.2 样本数据的正态性检验

本文提出的研究羊群效应的模型在回归前加入聚类分析对数据进行过滤，试图有效地将“羊群数据”和“非羊群数据”有效地聚成不同的类，并对不同的类作针对性的分析。由于传统的金融理论往往认为有效市场中的资产收益率服从对数正态分布。因而对于“非羊群数据”对数收益率的正态性检验将有助于检验这一理论，同时在一定程度上反映出羊群效应对市场的负面影响。本小节将介绍检验样本数据正态性的统计方法。

检验样本数据是否具有正态性的初步方法是 Q-Q 图(Quantile-Quantile Plot)，它是将样本的分位数与拟合分布的分位数画在同一图中，由此判断该组样本数据是否符合某种分布。就正态分布而言，若 Q-Q 图近似一条直线，则样本服从正态分布；若 Q-Q 图是凸的，样本为分布薄尾；若 Q-Q 图是凹的，则样本为分布厚尾。

然而 Q-Q 仅仅给出直观的、初步的判断，数理统计中检验样本数据正态性主要利用样本数据与拟合分布之间的一些差异构造统计量做假设检验。以下是一些著名的用于检验样本分布的统计量：

$$D = \sup_x |F_n(x) - F_0(x)| \quad (5-19)$$

$$A^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 [F_0(x)(1 - F_0(x))]^{-1} dF_0(x) \quad (5-20)$$

$$W^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 dF_0(x) \quad (5-21)$$

统计量表达式(5-19)、(5-20)、(5-21)分别称为 Kolmogorov-Smirnov 统计量、Anderson-Darling 统计量和 Cramer-von Mises 统计量，其中 $F_n(x)$ 、 $F_0(x)$ 分别表示样本的累计分布和拟合模型的分布。检验的原假设为样本服从正态分布，样本均值和方差分别等于相应正态分布的均值和方差。假设检验同样根据显著性概率 p 作为依据判断是否拒绝原假设。检验中， p 值大于显著水平 α ，表示没有理由否定正态性假设，即样本符合正态分布。

5.4 实验结果与分析

5.4.1 聚类分析的结果

实现聚类分析前，先对原始的收益率数据作剔除异常值的处理，并根据表达式(5-1)、(5-2)计算各时刻的个股对数收益率 R_{it} 及整体收益率 R_{mt} ，其中 $t=1,2,\dots,T$ ，并对得到的数据进行标准化处理。接着利用式(5-5)、(5-6)计算并标准化度量指标 H_t ，并利用得到的度量指标 H_t 及整体收益率 R_{mt} 实现 K 均值聚类。

K 均值聚类将实验数据初步分为 3 大组，分别记为相应的组为 Group1、Group2、Goup3。其中各组的样本数量分别为 190、360、230，经过计算得到相应的组内度量指标 $CSSD_t$ 、 $CSAD_t$ 、 H_t 及绝对收益率 $|R_{mt}|$ 均值、标准差的值，见表(5-1)。

表 5-1：各度量指标的组内均值、标准差

	Mean			Standard Deviation		
	Group1	Group2	Group3	Group1	Group2	Group3
R	2.69%	0.88%	2.27%	0.0174	0.0070	0.0165
CSAD	0.0155	0.0183	0.0264	0.0029	0.0028	0.0045
CSSD	0.0215	0.0249	0.0389	0.0042	0.0035	0.0101
H	0.0244	0.2159	0.1692	0.0514	1.437	0.8438

注：表中 R 表示的为市场整体的绝对对数收益率，即 $|R_{mt}|$

由表(5-1)中的描述性统计量可以看出，K 均值聚类分析的结果符合预期，数据被大致分为 3 组，其中 1、3 两组数据的绝对收益率明显偏大，而 Group1 的度量指标 $CSSD_t$ 、 $CSAD_t$ 及 H_t 相对较小，这组数据很可能显著地反映了市场中存在羊群效应。而对于 Group2 的数据，其收益率均值较小，而且 $CSAD_t$ 等度量指标的值较为均衡，个股收益率并不明显趋同于市场收益率，可以认为是显著并不存在羊群效应的较稳定市场情况；而 Group3 则平均收益率较大但其他度量指标

亦很大，市场绝对收益较大，但应该不存在羊群效应。据此，下面的研究将分为两部分：

1. 对于 Group1 的数据，应用 CH、CCK 等计量模型进行分析，根据回归系数的显著程度判断是否存在羊群效应。
2. 分别对于 Group2、Group3 的数据，做正态性分析看是否符合稳定市场的相关性质特征。

5.4.2 分组实现 CH、CCK 模型及显著性检验

在上述聚类分析结果的基础上，为了检测 Group1 数据中是否显著存在羊群效应及其他组数据的羊群程度，我们根据 CH、CCK 模型应用统计软件分别对 3 类数据及整体的样本数据分别做了回归分析。其中，CH 模型(5-7)、CCK 模型(5-8)具体的实现方法已在第三章 3.3 节中有详细的论述，模型中应用到的回归表达式为：(t=1,2,...,T)

$$\text{CCK 模型: } \text{CSAD}_t = \alpha + \beta_1 |R_{m,t}| + \beta_2 R_{m,t}^2 + \varepsilon_t$$

$$\text{CH 模型: } \text{CSSD}_t = \alpha + \beta_1 D_t^H + \beta_2 D_t^L + \varepsilon_t$$

模型具体的求解步骤如下：首先，利用表达式(5-3)、(5-4)计算各时刻度量指标 CSSD_t 、 CSAD_t 的值。利用聚类分析中得到的各组数据应用 CH、CCK 模型进行分析。最后，根据估计的回归方程中的系数及相应的显著性统计量的值判断是否存在羊群效应。

实验结果发现，在经过聚类分析的数据 Group1 中，拟合模型的修正拟合优度系数值为 0.8132，模型对数据的拟合程度较高；而各个回归参数显著，并且二次项系数 β_2 显著为负，根据 CCK 模型的理论，这类数据中明显存在羊群行为；而对包含所有时间点的样本数据应用 CCK 模型做回归分析，二次项系数 β_2 同样显著为负表明市场中存在羊群效应，而相应的 t 值则比 Group1 的结果小得多，

说明做了聚类分析后的数据能更有效地检测出市场的羊群效应。而且 R^2 值仅为 0.5021, 说明模型对数据的拟合程度较低。而对于 Group2、3 的数据, 回归系数 β_1 、 β_2 的显著性 t 统计量 p 值均颇大, 说明符合回归系数可能为 0。具体数据见表(5-2)。

表 5-2: CCK 模型在各组的回归参数及显著性

	Alpha	Beta1	Beta2
Group1	0.0098(22.42)*	0.00823(13.43)*	-0.00198(-11.57)*
All	0.01797(48.30)*	0.18221(5.82)*	-2.08153(-4.34)*
Group2	0.01726(56.33)*	0.11063(1.77)	0.79636(0.32)
Group3	0.02639(37.26)*	-0.00086(-0.02)	0.01224(0.02)

注: 括号内为个系数的 t 值, 并用*表示相应的显著性概率 $p < 0.0001$, 即相应回归系数显著不为 0。

另一方面, 应用 CH 模型的分析则显示无论用包含所有时间点的数据或 Group1 的数据做回归分析, 其结果差异并不大。在下跌的市场中, 羊群效应存在但其程度并不强烈, 而在上升的市场中则并不存在, 这说明国内投资者“杀跌”倾向大于“追涨”, 与此前相关研究的一些实证分析结果是一致的。CH 模型的具体回归数据见表(5-3)。

表 5-3: CH 模型在各组的回归参数及显著性

	Alpha	Beta1	Beta2
Group1	0.0217(64.86)*	0.00140(1.62)	-0.00544(-3.46)*
Group3	0.03908(55.19)*	0.00163(0.28)	0.0219(0.99)

注: 1. 括号内为个系数的 t 值, 并用*表示相应的显著性概率 $p < 0.0001$, 即相应回归系数显著不为 0。

2. 收益率最大(最小)的 30 个样本中 D_t^H (D_t^L) 取 1, 其余为 0。由于 Group2 数据中 D_t^H 和 D_t^L 值全为 0, 此处不作回归分析。

5.4.3 分组进行正态性分析的结果

由 CH、CCK 模型的分析结果发现, Group2、Group3 的数据并不存在羊群效应。而根据传统的金融理论显示, 正常的市场中对数收益率应服从正态分布。因此, 针对 Group2、3 的样本数据还做了正态性检验。

图(5-1)为 Group3 的 Q-Q 图, 由图可以看出数据有一定的厚尾性, 因此可以认为 Group3 的数据并不服从正态分布; 而仅用 Group2 的数据则近似直线, 显示出可能服从正态分布, 见图(5-2)。对 Group2 的数据作进一步的正态性分析, 其中包括计算样本的偏度、峰度及利用表达式(5-19)、(5-20)和(5-21)分别计算 Kolmogorov-Smirnov 等统计量。计算结果如表(5-4)所示, 该样本的偏度和峰度均接近于 0, 并且 Shapiro-Wilk 的 W 统计量为 0.993539, p 值为 0.127; 而 Kolmogorov-Smirnov、Cramer-von Mises 和 Anderson-Darling 等统计量的 p 值均大于 0.05, 并没有足够的证据否定数据的正态性。

正态性分析的结果显示, Group2 的数据符合正态分布, 根据传统的金融理论, 可以认为是来自稳定市场的数据。而 Group3 的数据则属于介乎稳定市场与羊群市场之间的情况。

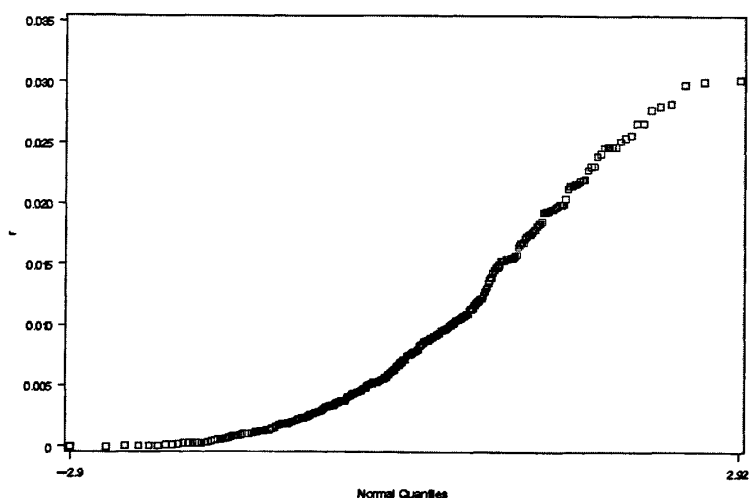


图 5-1 Group3 数据的 Q-Q 图

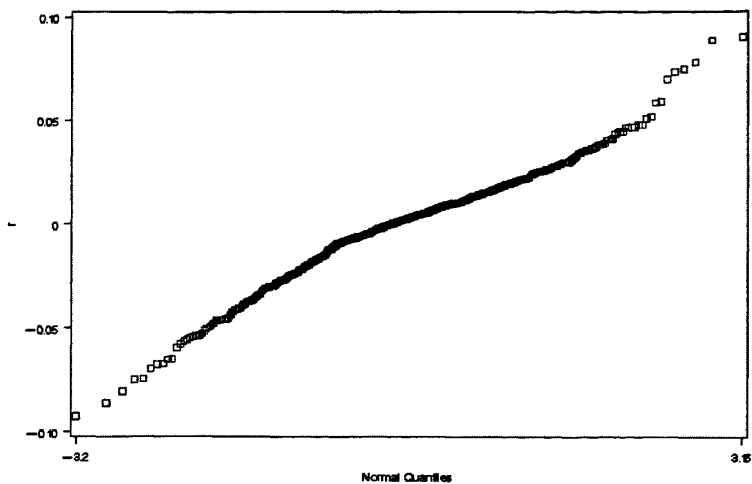


图 5-2 Group2 数据的 Q-Q 图

表 5-4: Group2 数据的正态性检验结果

The UNIVARIATE Procedure			
Variable: r			
Moments			
N	360	Sum Weights	360
Mean	0.00197197	Sum Observations	0.70991088
Std Deviation	0.01105105	Variance	0.00012213
Skewness	-0.1180872	Kurtosis	0.12824617
Uncorrected SS	0.04524304	Corrected SS	0.04384312
Coeff Variation	560.405207	Std Error Mean	0.00058244
Basic Statistical Measures			
Location		Variability	
Mean	0.001972	Std Deviation	0.01105
Median	0.002057	Variance	0.0001221
Mode	.	Range	0.06032
		Interquartile Range	0.01395
Tests for Location: Mu0=0			
Test	-Statistic-	-----p Value-----	
Student's t	t 3.385705	Pr > t	0.0008
Sign	M 35	Pr >= M	0.0003
Signed Rank	S 7108	Pr >= S	0.0003
Tests for Normality			
Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.999539	Pr < W	0.1270
Kolmogorov-Smirnov	D 0.032881	Pr > D	>0.1500
Cramer-von Mises	W-Sq 0.091775	Pr > W-Sq	0.1485
Anderson-Darling	A-Sq 0.637702	Pr > A-Sq	0.0967
Quantiles (Definition 5)			

5.4.4 模型结果的综合分析

上文的分析结果显示, Group1 的数据通过了 CH、CCK 模型的显著性检验, 说明在该组数据对应的时间点羊群效应显著存在; 而用 CH、CCK 模型对 Group2、Group3 的数据进行检验, 则并没有发现显著存在羊群效应。进一步地, Group2 的数据通过了正态性检验, 说明该组数据服从正态分布, 可以认为是来自稳定市场的数据。Group3 的数据则既不显著存在羊群性, 也不服从正态分布; 根据该组数据的整体收益率 $R_{m,t}$ 及度量指标 $CSSD_t$ 、 $CSAD_t$ 均较大的特征, 认为该组数据属于介乎稳定市场与羊群市场之间的情况。

因此, 初步的 K 均值聚类分析将数据有效分为 3 个特征显著的组: 存在羊群效应组数据(Group1)、市场收益率符合正态分布的组数据(Group2)、中间状态的组数据(Group3)。据此, 可以将我国的股票市场大致分为: 稳定市场和羊群市场。与许多经典的金融理论一致的是: 市场处于稳定的状态下, 其对数收益率符合正态分布, 而这一类数据主要集中在 2006 年 11 月至 2007 年 10 月的大牛市期间。而 CCK 模型的分析显示, 我国的市场在一定时期内存在羊群效应, 这导致市场整体的收益率偏离正常水平。

值得注意的是, 根据对存在羊群效应的 Group1 类数据的分析还发现: 在上证指数从 2000 点左右暴涨至 6000 多点的大牛市期间(2006 年 11 月至 2007 年 10 月)的数据仅有 28 条, 占该类数据数量约 15%; 而在全球金融海啸爆发后(2007 年 11 月以后)的数据有 138 条, 占该类数据数量超过 70%。这在一定程度说明了金融危机是股票市场羊群效应出现的原因之一, 而且国内市场的羊群效应在下跌时明显比上涨时强烈。

5.5 模型的评价及本章小结

前文在已有的检测羊群效应模型的基础上提出了加入聚类分析的建模思想对数据进行特征提取,并运用我国股票市场 2006 年到 2009 年的数据进行了实证分析,尝试检验市场中羊群效应的存在性及其存在的时间点。该模型及实验分析得到了下列一些结论:

1. 聚类分析的处理有效地对数据进行特征提取,通过聚类得到的各组数据均有显著的组特征。

2. 利用 CH、CCK 模型对经过聚类的各组数据进行分析,发现了我国股票市场中确实存在显著的羊群效应,而且羊群效应程度在金融风暴的下跌市场中尤为剧烈。

3. 与直接运用原始数据回归的模型相比,加入聚类分析对数据进行特征提取的模型检测羊群效应的效率更高。

4. 经过聚类分析得到的两组数据分别显示出羊群性和正态性,据此可以认为市场基本由稳定市场和羊群市场两种状态组成。

5. 本文提出的模型很好地区别符合传统股票价格随机波动理论的数据和存在羊群性的数据。而且通过羊群性数据可以找出羊群效应显著存在的时间段。这为下文关于股票价格波动状态的分析提供了基础。

总的来说,基于 K 均值聚类分析的羊群模型对市场的样本数据有效地进行了分组。这使得我们应用在险价值(Value at Risk, VaR)对风险资产作风险管理时,能更有效地针对那些处于分布尾部的异常波动情况,从而作出更准确的风险防范和规避措施。与此同时,也说明了股市中的羊群效应确实会对股票市场造成异常的波动,增加了市场的风险,也影响市场正常、稳定的发展。在本模型分析结果的基础上,针对相应组的数据作进一步研究分析可以帮助完成市场政策完善、风险管理、理解市场特征等不同的目标。

当然,基于 K 均值聚类分析的羊群模型也有一定的不足之处。由于 K 均值聚类算法的 K 是设定值, K 的选取将对聚类的结果有一定的影响。在不同的市

场中 K 可能取不同的值，如何根据市场的特征选取最合适的 K 值将是今后进一步研究的重点之一。另一方面， K 均值聚类算法属于贪心算法，得出的解并非全局最优的。但笔者认为由于羊群效应的研究对象是整体的，只要能保证聚类结果中每一组数据都有一定的数量，那么解的最优性对回归分析的结果影响应该不大。

第六章 羊群市场有限波动性模型

长期以来，GARCH 模型在金融市场的实证分析中能很好地描述市场收益的波动性。GARCH 模型认为市场收益的波动性与前面一段时期的波动性相关，较大的波动率的不断积累将使市场长期处于大幅波动的状态中。

对于欧美等发展较为成熟的市场，由于部分投资者能够灵敏地捕捉偏离正常水平的资产价格，并进行套利操作。这样的操作往往能起到稳定市场，降低市场收益率波动的作用。因此，成熟市场中收益率波动性从异常回归平稳的过程可以认为主要是由市场自身调节所引起的。

而像中国这样并不成熟的市场，一方面大部分投资者在大幅波动的市场当中更倾向于非理性的羊群投资策略；另一方面，诸多交易限制亦使得套利无法实现。上述各种因素使得市场无法实现自我调节，而只能由政府、金融监管机构等出台措施控制。因此，当我国股票市场长期处于大幅波动并达到某一临界点时，政府或金融监管机构将利用调控政策、措施来扭转这一趋势，试图恢复股票市场的稳定。

综上所述，可以认为股票市场收益的波动率应该处于一个有限的范围内。而前文的分析将不同时期的市场分为两种状态，自然地可以认为不同状态下的市场应该有来自不同总体的波动率，并且该波动率会在这两种市场中不断变化转换。因此，本章根据前文对我国股票市场的羊群聚类分析，尝试提出基于稳定市场和羊群市场波动率的有限波动模型，从羊群效应的角度出发解释我国股票市场的波动情况。

6.1 市场波动率的 GARCH 模型拟合

6.1.1 传统的时间序列模型

传统的时间序列模型的研究主要是通过建立随机动态模型研究不同时间点的的数据之间的相互依赖关系及随机项对后续序列的影响。传统时间序列研究的核心部分是研究方差的时变性，而异常的方差能影响到往后在一定时间内方差大小。正是由于时间序列模型的这一重要特征，使得它能很好地描述金融市场中风险的动态性、聚集性变化。因而，各类时间序列模型在金融工程学的实证分析中有着广泛的应用。其中最经典、应用最广泛的包括 Engle(1982)提出的 ARCH 模型、Bollerslev(1986)在 ARCH 模型基础上改进的 GARCH 模型等。

Engle^[27] (1982)提出用 ARCH 模型来研究通货膨胀率的时间序列，开创了时间序列条件异方差研究的先河，为此后金融时间序列的研究奠定了基础。对于一切的正整数 t ，ARCH 模型的结构为：

$$y_t = g(x_{t-1}; c) + \varepsilon_t \quad (6-1)$$

$$\varepsilon_t | \phi_{t-1} \sim N(0, \sigma_t^2) \quad (6-2)$$

$$\sigma_t^2 = \text{Var}(\varepsilon_t | \phi_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \quad (6-3)$$

其中(6-1)式为回归方程， $g(x_{t-1}; c)$ 为含前 $t-1$ 时刻信息的向量 x_{t-1} 及参数 c 的函数， ε_t 为误差项；(6-2)式表示在前 $t-1$ 时期的条件下，误差项 ε_t 服从均值为 0，方差为 σ_t^2 的正态分布； σ_t^2 的具体表达式由(6-3)给出，它表示 t 时期的随机误差由前 p 期误差的影响产生的，对应的系数 $\alpha_i > 0$ 反映的是前 p 期误差对 t 时期随机误差影响的大小。

由(6-3)式可以看出，若 t 时刻前的一段时期内 ε_{t-i} 很大，即出现大幅度的随机扰动，那么 ε_t 也将很大。假如回归方程的因变量是资产的收益率，那么相应时期的大幅波动的收益率就能从模型中反映出来。同时金融市场中收益率波动的聚

集性也能被 ARCH 模型检测出来。

尽管 ARCH 模型在提出之初主要被用于研究通货膨胀率,但它对资产价格聚集性的精确刻画很快被扩展应用到金融市场之中。Bollerslev^[28] (1986)建立的 GARCH 模型就是一个很好的例子。Bollerslev 推广 Engle 的 ARCH 模型,提出了条件一般化的自回归异方差模型(Generalized Autoregression Conditional Heteroscedasticity, GARCH)。GARCH 模型能够很好地描述金融时间序列收益率方差的时变性及自相关性。

一般来说, GARCH(p, q)模型由条件均值方程(6-1)与条件方差方程(6-4)组成,由于 GARCH 模型主要关注时间序列的波动性,因而它主要关注的是条件方差方程(6-4): ($t=1,2,\dots$)

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (6-4)$$

条件方差方程(6-4)中, ε_{t-i} 为 $t-i$ 时刻的随机收益率, $\sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$ 称为 ARCH 项, 而 $\sum_{j=1}^q \beta_j \sigma_{t-j}^2$ 则称为 GARCH 项。

ARCH 项描述的是 t 时刻的条件异方差 σ_t^2 与前面各期随机收益率 ε_{t-j} 平方的线性相关关系, 而系数 α_i 则反映了 $t-i$ 时刻的随机收益率对后续时刻波动性的影响程度。若前 p 期内出现较大的随机收益率 ε_{t-i} , 而相应的 α_i 又不显著为 0, 则该随机扰动对 t 时刻的波动性仍有较大影响。

而模型的 GARCH 项正是 Bollerslev 对 ARCH 模型的推广,它反映的是条件异方差序列的自相关性及聚集性。事实上, GARCH(p, q)模型等价于无穷阶的 ARCH 模型, 而且 GARCH 模型仅包含有限个参数。正是由于模型中的待估参数大大少于 ARCH 模型, GARCH 模型在描述金融风险中得到了广泛的应用。尤其是 GARCH(1,1)模型, 其形式十分简洁, 但却能很好地描述金融回报率波动性。

6.1.2 羊群市场波动率的 GARCH(1,1)模型拟合

本文尝试利用 GARCH(1,1)模型对市场对数收益率波动性进行拟合。GARCH 模型的拟合利用市场收益率 $R_{m,t}$ 的时间序列作为输入的数据，并通过最大化对数似然函数进行参数估计。在随机扰动 ε_t 服从高斯分布的假设下，对数似然函数为：

$$L(\theta) = \sum_{t=1}^T l_t(\theta) = \sum_{t=1}^T \left(-\frac{1}{2} \left(\ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2} \right) + \ln \frac{1}{\sqrt{2\pi}} \right) \quad (6-5)$$

利用统计软件实现拟合的 GARCH(1,1)模型为：(t=1,2,...)

$$\sigma_t^2 = 8.3097 \times 10^{-6} + 0.05179 \varepsilon_{t-1}^2 + 0.92865 \sigma_{t-1}^2 \quad (6-6)$$

另外，表 6-1 为 GARCH(1,1)模型参数的估计及 t 统计量的结果，利用模型的 t 统计量可以知道在 5%(t 值为 1.960)显著水平下，模型系数是显著的。此外图 6-2 则显示了市场收益率（图 6-2 的上图）及相应条件异方差的波动情况（图 6-2 的下图）。

如图 6-2 的下图所示，我国市场收益率的波动是有界的，大致在 0.01 到 0.04 之间，这与本章开头有限波动的分析相吻合。

此外，如 GARCH 模型(6-6)及图 6-2 所示，模型中较大的 GARCH 项意味着某时刻的大幅波动对市场波动性的影响将长期存在。存在羊群效应的中国金融市场将长期处于大幅度的波动状态，这样的市场是不成熟的，金融危机的冲击将长期破坏市场的稳定性。

表 6-1: GARCH(1,1)模型参数的求解结果

Parameter	Value	Standard Error	T Statistic
C	0.004964	0.0007773	6.3862
K	8.3097e-006	4.0948e-006	2.0293
GARCH(1)	0.92865	0.021961	42.2855
ARCH(1)	0.051791	0.01506	3.4390

Log Likelihood Value: 1383.47

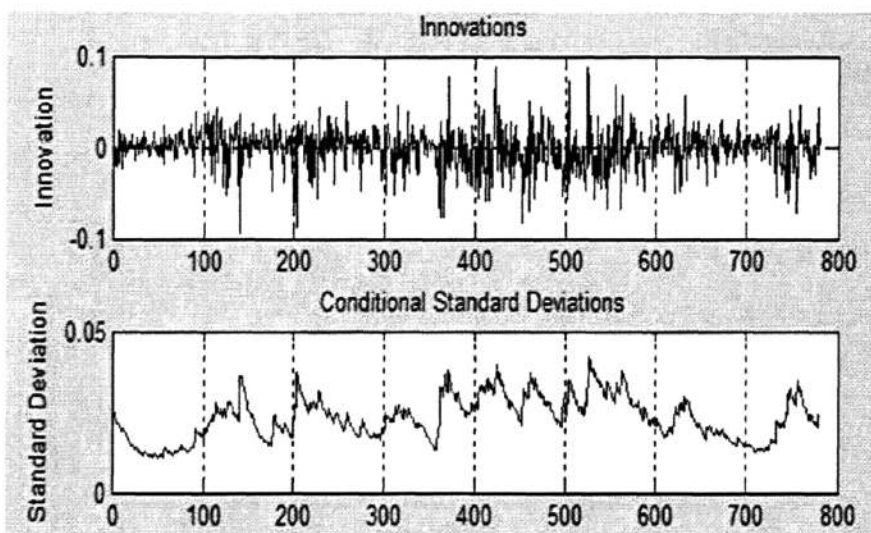


图 6-2: GARCH(1,1)模型拟合效果图

6.2 本文提出的羊群市场有限波动性模型

由于市场中羊群效应的出现往往伴随着高额的市场波动率,并将影响往后一段的时期市场波动。根据前文的分析,我国股票市场很可能由两种市场状态组成:稳定市场和羊群市场。针对这两种市场分析其相应的波动性特征将有效地了解我国股票市场收益率的波动状态。因此,本文尝试从羊群市场的角度出发,提出收益率的有限波动性模型来解释股票市场的波动状况。

6.2.1 羊群市场有限波动性模型

前文的研究中发现, 我国的股票市场主要呈现出两种情形: 在稳定市场中, 市场整体的收益率符合对数正态分布; 而在羊群市场中, 市场收益率波动性增大, 并且市场收益率会出现一定的聚集性。因此, 针对羊群市场数据研究其市场收益的波动率可以理解为两种状态的不同的波动性的组合, 其表达式为: $(t=1,2,\dots,T)$

$$\sigma_{m,t}^2 = \sigma_n^2 + \mu(x_t)(\sigma_h^2 - \sigma_n^2) \quad (6-7)$$

其中, σ_m^2 为市场整体收益波动率; σ_n^2 为稳定市场下波动率, 即高斯分布的方差; σ_h^2 为羊群市场下的波动率; 待定函数 $\mu(x_t)$ 为模型系数, 其取值一般为 0 到 1 之间, 取值接近 1 时模型表现出羊群性, 波动率接近羊群市场波动率, 而接近 0 时则为稳定市场状态; x_t 为描述市场羊群性的时间序列。值得注意的是, x_t 的大小将直接反映羊群程度, 故 $\mu(x_t)$ 一般应取为单调函数。

6.2.2 有限波动性模型对市场数据的拟合

这里尝试选取本文提出的度量指标 H_t 作为描述羊群效应的时间序列变量, 研究市场波动率。模型中的系数函数取为: $(t=1,2,\dots,T)$

$$\mu(H_t) = \exp(-H_t) \quad (6-8)$$

由度量指标 H_t 的定义可知, H_t 反映了市场的羊群程度, 羊群程度随着 H_t 值的增加而降低, 并且其值域为 $[0, +\infty)$; 而 $\mu(H_t)$ 为 H_t 的单调递减函数, 其值域为 $[0, 1]$ 。因此, H_t 越小, 羊群程度越强烈; 则相应地 $\mu(H_t)$ 值越大、波动率越大。

σ_n^2 、 σ_h^2 可以利用 Group1、2 数据中的市场对数收益率 $R_{m,t}$ 作矩估计，其值取分别为 0.000911、0.000122(相应的标准差为 0.0302、0.011)，即模型(6-7)中 $\sigma_n^2=0.000122$ ， $\sigma_h^2=0.000911$ 。据此，得到在羊群效应意义下，我国市场收益率的波动性拟合的表达式为 (6-9)：(t=1,2,...,T)

$$\sigma_{m,t}^2 = 0.000122 + 0.000789\mu(H_t) \quad (6-9)$$

利用度量指标 H_t 代入(6-8)、(6-9)式中便可计算相应的市场收益率的波动性 $\sigma_{m,t}^2$ 。图 6-3 为本文提出的有限波动性模型拟合效果图；图 6-4 则是有限波动模型拟合市场波动率的总体趋势图，图 6-2 中 GARCH 模型的拟合效果图有相似的总体趋势。

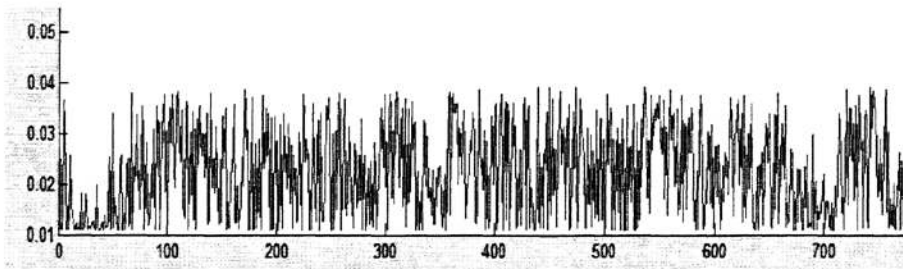


图 6-3：有限波动性模型拟合效果图

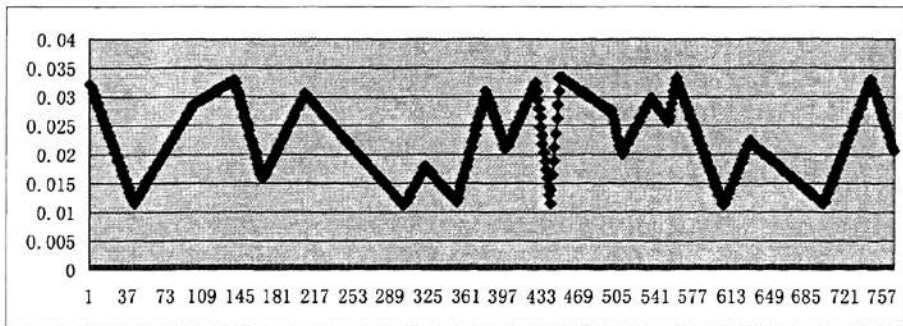


图 6-4：有限波动性总体趋势图

6.3 模型拟合的结果对比及分析

本文提出有限波动模型主要的目的有两个：一方面由 σ_n^2 、 σ_h^2 作为上下界反映了市场的有限波动状况；另一方面，尝试通过时间序列系数 $\mu(x_t)$ 反映收益率波动的性质。

根据本文的有限波动性模型与 GARCH(1,1)模型拟合结果的对比发现，有限波动性模型对市场收益率波动总体趋势的拟合有较好的效果：由图 6-2 的下图（GARCH 模型的拟合结果）与图 6-4（本文的有限波动模型的结果）可以看出两者总体的波形相当吻合，能反映一段较长时期内的波动性。

但与 GARCH(1,1)模型拟合的结果相比，有限波动性模型拟合的波动率振动较为强烈，如图 6-3 所示。这方面的结果与 GARCH 模型波动率具有聚集性的结论并不一致。

由于 GARCH(1,1)模型假定 t 时刻波动率与 $t-1$ 时刻的波动率相关，而模型拟合的 GARCH 项很大，这使波动率有聚集性。

另一方面，结合我国股票市场的实际情况，可以从以下几个方面来解析有限波动模型得到市场波动较频繁的结果：

1. 羊群效应并不长期连续出现

由我国股票市场近年来的数据来看，市场在经过数个交易日的大幅波动后往往会出现短期的调整。羊群投资者将观望市场的总体走势再作出投资决定。因此，羊群效应并不连续存在，这使得有限波动模型拟合的市场波动可能会在一段时期内表现出回复到较低水平的性质。

2. 套利投资者的反向操作

在市场出现羊群效应的情况下，部分股票的价格会偏离正常水平，形成套利的机会。套利投资者与羊群投资者相反方向的操作使得个股收益率的趋同性降低，这使度量指标 H_t 增大。而套利机会消失后，市场从新回复原有状态。因此，

市场波动较为频繁。

3. 投资者对信息极其敏感

我国股票市场的散户投资者对市场信息极其敏感的。在市场信息不对称的情况下,由于这部分投资者对市场信息真伪的辨别能力不高,使他们可能作出错误的决策,最终导致信息瀑布流,序列羊群产生。因此,即使在较为稳定的市场状态下,短期仍也有可能出现轻微的羊群效应,造成较大的波动。

总的来说,本文的有限波动模型在整体走势方面则能与 GARCH 模型大致相似,而且更能反映我国存在羊群效应的市场波动状况。

6.4 本章小结

根据市场羊群性的分析,本章尝试提出有限波动模型,从羊群效应的角度出发理解市场波动性状况。该模型主要通过本文提出的度量羊群效应的度量指标 H 作为区分羊群市场与稳定市场的指标,并由此拟合市场收益率的波动情况。尽管有限波动模型与著名的 GARCH 模型相比还比较粗糙,但该模型发现的一些结果还是值得注意的:

1. GARCH 模型与有限波动模型的拟合结果均反映了我国股票市场的波动率在一定的有限幅度内变化的性质。

2. 有限波动模型在拟合市场整体波动率的长期趋势时与 GARCH 模型的拟合结果大致相同。但有限波动模型发现市场波动可能存在短期频繁振动的现象,这与 GARCH 模型得到的波动率聚集性结果并不一致。

第七章 总结

7.1 全文总结

本文对现有各类关于投资市场中的羊群效应的理论及实证分析模型作了系统的介绍和分析。针对研究股票市场整体的 CH、CCK 模型并不能检测出羊群效应存在的时间段的问题,本文结合数据挖掘的相关理论,提出了基于 K 均值聚类分析的羊群效应建模思路。通过对我国市场数据的实证研究,有效地检验羊群效应的存在性及其存在的时间段;而对存在羊群效应的相关数据分析还发现金融危机是羊群效应产生的原因之一。另一方面,对市场数据的聚类分析能有效地区分稳定市场及存在羊群效应的市场;针对稳定市场的数据研究验证了市场对数收益率服从正态分布的假设。最后根据市场的分类状况,尝试在羊群效应的意义下提出研究市场波动性的模型,将市场波动率解析为稳定市场波动率与羊群市场波动率的线性组合,并且针对我国市场数据结合 GARCH 模型对比分析了存在羊群效应的市场收益率波动性状况。

由于各种因素的限制,本文还有一定的不足之处,其中包括:仅用了 K 均值聚类方法进行聚类分析;并没有用不同类型的股票市场验证聚类分析模型的效果等。后续的研究可适当地尝试结合其他有效的数据挖掘模型对数据进行特征提取;另一方面,利用不同市场(如:代表成熟市场的美国、香港等)的数据进行分析和对比,既能验证模型的效率也能对羊群效应的形成、防范等方面得出更深入的了解。另外,本文创新地提出的有限波动性模型还比较粗糙,系数 $\mu(x_i)$ 应如何选取才能达到较好的模型效果、如何对时间序列数据进行建模反映羊群效应等方面也是今后需要进一步重点研究的。

参考文献

- [1] Mandelbrot, B.. The Variation of Certain Speculative Prices. The Journal of Business of the University of Chicago 36, 1963: 394-419
- [2] Abhijit V. Banerjee. A Simple Model of Herd Behavior. The Quarterly Journal of Economics, 1992, Vol. 107:797-817
- [3] Kindleberger, C. Manias, Panics and Crashes: A History of Financial Crises. New York: Basic Books, 1989
- [4] Chang, Eric, Joseph Cheng, Ajay Khurana. An Examination of Herd Behavior in Equity Markets: An Empirical Perspective. Georgia Institute of Technology, 1998
- [5] Scharfstein, David and Jeremy Stein. Herd Behavior and Investment. American Economic Review, 1990, Vol.80:465-479
- [6] Graham, John R. Herding among Investment Newsletters: Theory and Evidence. Journal of Finance, 1999:237-268
- [7] Lakonishok, Josef, Andrei Shleifer and Robert W. Vishny. The Impact of Institutional Trading on Stock Prices. Journal of Financial Economics, 1992, Vol. 32:498-500
- [8] Christie, William G, Roger D, Huang. Following the Pied Piper: Do Individual Returns Herd Around the Market?. Financial Analysts Journal, 1995, July-August:31-37
- [9] 宋军、吴冲锋. 基于分散度的金融市场中的羊群行为研究. 经济研究, 2001-11
- [10] 孙培源、施东辉. 中国证券市场羊群行为实证研究. 经济研究, 2004-11
- [11] 陈浩. 中国股票市场机构投资者羊群行为实证研究. 南开经济研究, 2004-2
- [12] Bikhchandani, Sushil and Sharma, Sunil. Herd Behavior in Financial Market: A Review. IMF working paper, 2001.
- [13] Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades. Journal of Political Economy, 1992, Vol. 100:992-1026
- [14] Keynes, John Maynard. The General Theory of Employment, Interest and Money. New York: Harcourt Brace and Co., 1936:176-190
- [15] Avery, Christopher and Peter Zemsky. Multidimensional Uncertainty and Herd Behavior in Financial Markets. American Economic Review, 1998, Vol.88:724-748

- [16] Zwiebel, Jeffrey. Corporate Conservatism and Relative Compensation. *Journal of Political Economy*, 1995, Vol. 103, No.1.:1-25
- [17] Grinblatt, Mark, Sheridan Titman and Russ Wermers. Momentum Investment Strategies, Portfolio Performance and Herding: A Study of Mutual Fund Behavior. *American Economic Review*, 1995, Vol. 85 No.5.:1088-1105
- [18] Rui Xu and Donald Wunch II. Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 2005, Vol. 16 No.3:645-678
- [19] David Hand, Heikki Mannila, Padhraic Smyth 著; 张银奎 廖丽 宋俊等译. 数据挖掘原理. 北京: 机械工业出版社, 2003. pp.173-206
- [20] 孙培源、施东晖. 基于 CAPM 模型的中国股市羊群行为实证研究. *经济研究*, 2002
- [21] 宋军、吴冲锋. 中国股评家的羊群行为研究. *管理科学学报*, 2003
- [22] 史莱佛著; 赵英军等译. 并非有效市场: 行为金融学导论. 北京: 中国人民大学出版社, 2003. pp. 21-26
- [23] 王稳. 行为金融学. 北京: 对外经济贸易大学出版社, 2004. pp. 208-228
- [24] R. Lyman Ott, Michael Longnecker 著; 张忠占等译. 统计方法与数据分析引论. 北京: 科学出版社, 2003. pp. 583-678
- [25] 茆诗松, 王静龙, 濮晓龙. 高等数理统计. 北京: 高等教育出版社, 2006
- [26] 朱平芳. 现代计量经济学. 上海: 上海财经大学出版社, 2004. pp. 13-25
- [27] Engle, Robert F.. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrics*, 1982:987-1007
- [28] Bollerslev, Tim.. Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 1986, Vol.31:307-327
- [29] 王振龙. 时间序列分析. 北京: 中国统计出版社, 2000
- [30] 高惠旋. 实用统计方法与 SAS 系统. 北京: 清华大学出版社, 2001

附录

下面列出实证分析中涉及到的个股名称及代码：

浦发银行(600000)、白云机场(600004)、武钢股份(600005)、东风汽车(600006)
中国国贸(600007)、首钢股份(600008)、上海机场(600009)、包钢股份(600010)
华能国际(600011)、皖通高速(600012)、华夏银行(600015)、民生银行(600016)
日照港(600017)、上港集团(600018)、宝钢股份(600019)、中原高速(600020)
上海电力(600021)、济南钢铁(600022)、华电国际(600027)、中国石化(600028)
南方航空(600029)、中信证券(600030)、三一重工(600031)、福建高速(600033)
招商银行(600036)、歌华有线(600037)、保利地产(600048)、中国联通(600050)
五矿发展(600058)、双鹤药业(600062)、宇通客车(600066)、葛洲坝(600068)
同仁堂(600085)、长航油运(600087)、特变电工(600089)、云天化(600096)
广州控股(600098)、同方股份(600100)、莱钢股份(600102)、上海汽车(600104)
亚盛集团(600108)、国金证券(600109)、中科英华(600110)、包钢稀土(600111)
浙江东日(600113)、西宁特钢(600117)、中国卫星(600118)、兰花科创(600123)
铁龙物流(600125)、重庆啤酒(600132)、金发科技(600143)、中国船舶(600150)
航天机电(600151)、建发股份(600153)、中体产业(600158)、上海建工(600170)
雅戈尔(600177)、生益科技(600183)、兖州煤业(600188)、复星医药(600196)
新潮中宝(600208)、紫江企业(600210)、南山铝业(600219)、江苏阳光(600220)
海南航空(600221)、桂冠电力(600236)、广汇股份(600256)、北京城建(600266)
海正药业(600267)、外运发展(600270)、航天信息(600271)、南钢股份(600282)
蓝星新材(600299)、酒钢宏兴(600307)、华泰股份(600308)、烟台万华(600309)
洪都航空(600316)、振华重工(600320)、华发股份(600325)、宏达股份(600331)
国阳新能(600348)、山东高速(600350)、国旅联合(600358)、江西铜业(600362)
首开股份(600376)、健康元(600380)、金地集团(600383)、小商品城(600415)
江淮汽车(600418)、华鲁恒升(600426)、中远航运(600428)、吉恩镍业(600432)
宝钛股份(600456)、中金黄金(600489)、驰宏锌锗(600497)、中化国际(600500)

上海能源(600508)、贵州茅台(600519)、中铁二局(600528)、山东黄金(600547)
深高速(600548)、厦门钨业(600549)、天威保变(600550)、安阳钢铁(600569)
海油工程(600583)、海螺水泥(600585)、用友软件(600588)、龙溪股份(600591)
中孚实业(600595)、新安股份(600596)、光明乳业(600597)、北大荒(600598)
青岛啤酒(600600)、方正科技(600601)、大众交通(600611)、金枫酒业(600616)
百联股份(600631)、大众公用(600635)、新黄浦(600638)、浦东金桥(600639)
申能股份(600642)、爱建股份(600643)、城投控股(600649)、申华控股(600653)
豫园商城(600655)、福耀玻璃(600660)、陆家嘴(600663)、哈药股份(600664)
川投能源(600674)、中华企业(600675)、广船国际(600685)、上石化(600688)
青岛海尔(600690)、大商股份(600694)、天津港(600717)、中粮屯河(600737)
辽宁成大(600739)、华域汽车(600741)、大连控股(600747)、上实发展(600748)
安徽合力(600761)、综艺股份(600770)、水井坊(600779)、通宝能源(600780)
国电电力(600795)、鹏博士(600804)、马钢股份(600808)、山西汾酒(600809)
东方集团(600811)、华北制药(600812)、香溢融通(600830)、东方明珠(600832)
上海机电(600835)、海通证券(600837)、四川长虹(600839)、王府井(600859)
内蒙华电(600863)、创业环保(600874)、东方电气(600875)、航天电子(600879)
亚泰集团(600881)、国投电力(600886)、南京化纤(600889)、张江高科(600895)
长江电力(600900)、株冶集团(600961)、宜华木业(600978)、开滦股份(600997)
大同煤业(601001)、重庆钢铁(601005)、大秦铁路(601006)、南京银行(601009)
中国神华(601088)、中国国航(601111)、兴业银行(601166)、西部矿业(601168)
北京银行(601169)、中国钢铁(601186)、中国平安(601318)、交通银行(601328)
广深铁路(601333)、中国中铁(601390)、工商银行(601398)、北辰实业(601588)
中国铝业(601600)、中国太保(601601)、中国人寿(601628)、平煤股份(601666)
潞安环能(601699)、中海油服(601808)、中国石油(601857)、中海集运(601866)
招商轮船(601872)、中煤能源(601898)、中国远洋(601919)、建设银行(601939)
金钼股份(601958)、中国银行(601988)、大唐发电(601991)、中信银行(601998)

由股票数据软件“同花顺”导出的原始数据如下表所示。由于篇幅限制，这里仅取浦发银行(600000)的部分数据作为例子展示。

附表 1：浦发银行(600000)

时间	开盘	收盘	振幅	总手	换手%
2009-10-13, 二	20.57	21.13	3.47%	55,918,333	0.71
2009-10-12, 一	21.18	20.74	4.34%	95,848,640	1.21
2009-10-09, 五	20.16	20.99	5.55%	113,179,038	1.43
2009-09-30, 三	19.81	19.65	2.29%	97,700,149	1.23
2009-09-29, 二	19.16	19.67	4.14%	89,826,229	1.13
2009-09-28, 一	19.58	19.1	3.94%	58,480,607	0.74
2009-09-25, 五	19.78	19.54	4.04%	80,558,935	1.02
2009-09-24, 四	19.08	20.06	7.61%	170,619,700	2.15
2009-09-23, 三	19.12	19.19	3.20%	102,362,940	1.29
2009-09-22, 二	19.6	19.06	4.48%	118,315,677	1.49
2009-09-21, 一	20	19.85	4.79%	140,756,440	1.78
2009-09-18, 五	21.56	20.46	6.00%	134,064,181	1.69
2009-09-17, 四	21.4	21.5	1.99%	102,535,289	1.29
2009-09-16, 三	21.53	21.13	3.44%	116,819,818	1.47
2009-09-15, 二	21.65	21.79	4.95%	128,355,482	1.62
2009-09-14, 一	21.45	21.62	2.93%	139,787,310	1.76
2009-09-11, 五	19.7	21.53	9.98%	205,714,850	2.6
2009-09-10, 四	20.09	19.74	2.44%	73,020,666	0.92
2009-09-09, 三	20.44	20.07	3.21%	104,995,077	1.32
2009-09-08, 二	19.79	20.22	4.26%	101,138,469	1.28
2009-09-07, 一	20.32	19.94	4.51%	93,595,476	1.18
2009-09-04, 五	20.2	20.17	3.02%	111,260,023	1.4
2009-09-03, 四	19.31	20.23	6.75%	154,497,090	1.95
2009-09-02, 三	18.26	19.27	6.53%	140,914,800	1.78
2009-09-01, 二	17.76	18.54	7.29%	119,499,919	1.51

致 谢

感谢数计学院的所有老师。感谢本科和研究生学习阶段给予我指导的所有老师。特别感谢导师张磊老师对我孜孜不倦的教导和关怀。从论文的开题到撰写完成，恩师都不厌其烦地给予我各种宝贵的意见；感谢她在论文写作、实验、细节修改、评阅、组织答辩等各方面给予我的帮助。感谢百忙中评阅论文和参加答辩的各位老师。

感谢一起参加讨论班学习、研究的所有同学。特别感谢钟陈练、黎俊彬、温德美、郭兆聪、汪智斌等多年的好友在学习和生活中给予我的极大的支持和鼓励。

最后，感谢父母多年来对我细心的关怀和照顾。