

中文摘要

随着中国资本市场的迅速发展和居民收入水平的较大提高，越来越多的人参与到股票市场的投资中，希望能够实现财富的保值和增值。虽然股票是一个可能带来高回报的投资品种，但是同时它又是高风险的，其时刻变化不定的价格，让投资者感受到这个市场的复杂性，所以他们迫切需要一种理论来解释价格变动的原因，并需要一种科学的预测方法来指导投资，从而规避风险，获得较好的投资回报。股票价格的预测是一个世界性难题，但这项研究蕴含着巨大的潜在经济利益，所以吸引着全球许多的学者专家来对其进行研究。

Fama 在 20 世纪 70 年代提出了著名的有效市场假设，认为股票价格的波动是随机游走过程，将来的价格是不可预测的，人们不可能依靠分析历史价格信息和公开信息而获得超常收益。然而，近年来越来越多的实证研究表明，这种理论与事实不符合。为了更有效地解释证券市场的波动规律，Peters 在 20 世纪 90 年代初提出了分形市场假说，强调证券市场信息接受程度和投资时间尺度对投资者行为的影响，认为所有稳定的市场都存在分形结构，证券价格在一定程度上存在可预测性。

在承认股票价格存在可预测性的前提下，本文系统阐述了股票价格预测的各种方法，既包括传统的基本分析以及技术分析等方法，也包括基于时间序列分析的 ARIMA 和 GARCH 模型等现代预测方法，当然由于影响股票价格的因素众多，不可能用一种简单的模型或方法就能够完全预测其未来的涨跌情况。

本文的创新之处主要体现在所采用的预测方法上，文献中常见的是用 GARCH 模型来对收益率的波动率进行估计，本文尝试使用 GARCH 模型和 ARIMA 模型对股价波动趋势进行短期预测。首先利用 7 日移动平均方法对上证指数日收盘价数据进行了适当的降噪处理，然后对此趋势序列进行了 ARIMA 建模和 GARCH 建模分析，并利用所建立的模型进行了短期趋势预测，发现有较好的短期预测效果，对提醒投资者注意风险以及帮助投资者把握买卖时机上有较强的现实意义和一定的实用价值。

关键词：股票价格；预测；时间序列

ABSTRACT

With the rapid development of china's capital market and the great increment of people's income, more and more people put their money into the stock market hoping to keep and increase the value of their wealth. Although investing in stock market can bring you a great fortune, it is also very risky. The nonstop up and down movement of its price puzzles investors. They realize that the capital market is very difficult to understand. They desperately need a theory to explain the reason why the prices change. They want to understand the mechanism of the price movement. And they need scientific forecast methods which can guide their investment to avoid risk and get good return. Because of the great potential economic interest, the forecast of stock price which attracts the research of many scholars and professionals, in spite of its difficulty.

In 1970's, Fama proposed the famous Efficient Market Hypothesis which concluded that the movement of stock price is a random walk, the price of stock is impossible to predict. Nobody can get extra return by means of analyzing the history of price and public information. However, recent researches show that this theory contradicts the facts. To better explain the movement of stock price, Peters proposed the Fractal Market Hypothesis which emphasizes the influence of information acceptance and time scale. And this theory points out that fractal structure exists in all stable markets. The price of stock is predictable in some degree.

In the presupposition that stock price is predictable, the methods of stock price forecast are systematically introduced. They are traditional fundamental analysis and technical analysis, and modern forecast methods of ARIMA and GARCH model which are based on time series analysis are also introduced. Of course, because of the complexity of stock price movement, we can not predict the future price exactly with one single simple model.

In this dissertation, the innovation is the empirical analysis of trend of Shanghai stock index based on ARIMA model and GARCH model. We find that the constructed models have excellent short-term prediction ability. They are useful to note the investors the risk and the opportunity of buying and selling. They have certain practical values.

Key Words: stock price; forecast; time series.

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):赵国顺

2009 年 6 月 19 日

厦门大学学位论文著作权使用声明

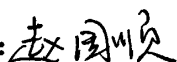
本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

2009 年 6 月 19 日

第一章 绪论

股票是由股份公司发行的，表示其股东按其持有的股份享受权益和承担义务的可转让的资产凭证。它是股份公司重要的融资工具，也是资本市场重要的投资工具，其最基本特性是收益性和风险性

股票市场可以分为两个层次，即一级市场和二级市场。一级市场是新证券的发行、包销及认购市场，二级市场是已发行证券的交易市场。本文所指的股票市场为二级市场，即股票的交易流通市场。高风险高回报是股票投资的重要特征，从股票市场诞生以来，投资者们就一直在进行探索、分析及研究其发展变化规律，希望能够预测其发展趋势，从而获得理想的投资回报。

第一节 研究的背景和意义

一、研究背景

股票市场是金融市场的重要组成部分，与国民经济发展密切相关，股票作为资本证券化的工具，一种虚拟资本的代表，可以方便地在股票市场融资。从表面上看，股票市场是缺乏秩序的，或者说缺乏通常意义下的规律性，似乎它的变动是难以预测的。许多经济学家用毕生精力，创造出各种数学模型来简化股票市场，试图了解资本的流动方向及其规律性。

人们对以股票价格为代表的各种金融资产价格的分析预测方法的研究，从股票市场和各种金融市场的出现开始，就没有停止过，涌现出了各种各样的分析方法体系。虽然这些方法派系众多，但从其出发点和分析对象来考虑，可以分为基本分析（fundamental analysis）和技术分析（technical analysis）两大类。基本分析是通过研究影响证券价格的基本因素，确定证券的内在价值，找出被市场错误定价的证券，买入并长期持有，直到证券的市场价格恢复到其内在价值为止。技术分析则是根据证券市场信息的历史数据，通过图表、技术指标等寻求股票价格变化的规律，来找到最好的买入点和卖出点。

随着数学理论研究的深入和各种数据分析工具开发的迅速发展，人们用各

种不同的方法和工具来分析金融时间序列，做出各种金融时间序列预测的模型，尤其是股票价格的预测模型。

时间序列分析方法是统计学研究的一个重要分支。它直接以事物在不同时刻的状态所形成的数据为研究对象，通过对时间序列数据的特征进行分析和研究，揭示事物的发展变化规律。经典的时间序列分析方法有图表法、指标法和模型法。模型法是目前对时间序列进行深层次分析和刻画的主要方法，它的分析思路是利用事物过去和现在的变化规律数据，构造依时间变化的序列模型，借助外推原则来推测未来。时间序列模型分析方法具有逻辑的美感，模型的形式通常是一个或者一组数学方程。这不但具有表达上的简洁性，而且可解释性和可理解性也比较好，为进一步的处理、推导和应用提供了较大的方便。

一些经典的时间序列分析模型如 ARMA, ARCH, GARCH 等已经被大量应用于金融时间序列预测中来，如美国经济学家 Engle 就因为他 1982 年针对金融时间序列所提出的 ARCH 模型而获得 2003 年度诺贝尔经济学奖。ARCH 模型很好地抓住了金融波动的变异性和丛集性特点，常被用来对金融时间序列进行分析、建模和预测。

二、研究意义

九十年代初，我国先后建立了上海和深圳两大证券交易所，截至 2009 年 4 月，我国股票账户数达到 12727.97 万，是目前社会参与程度最高的金融市场。通过建立合理的模型对股票价格趋势进行短期预测，同时不断利用新的数据更新模型，做出新的预测，我们就可以实现连续预测，这对投资者如何较好地把握买卖时机将起到重要的指导作用。

目前证券投资已经成为社会生活的一个重要部分，证券市场的运行状况对我国实体经济的影响逐渐加深，正在真正地发挥着国民经济“晴雨表”和“报警器”的作用。深刻理解股票市场的运行规律，预测分析股票价格的未来走势，无论是对广大投资者的降低投资风险还是宏观经济管理部门的宏观调控，保障我国证券市场的健康持续发展，都有着十分重要的意义。

第二节 本文主要内容及结构

股票价格的时间序列一直对预测者有着巨大的吸引力,是各种预测方法应用的热门领域。本文系统论述了股票价预测理论方法体系,研究了它们在证券市场的基本原理和实用价值。实证研究中,对上证指数短期趋势时间序列建立 ARIMA 模型和 AR-GARCH 模型,取得了较好的拟合效果和短期预测效果,可以帮助投资者规避风险,把握买卖时机。

结构:本文共分为六章。

第一章是绪论。主要对股票以及股票市场的概念做了说明,介绍了研究的背景和意义、论文的机构安排以及论文的创新点。第二章主要论述了自股票市场诞生以来,股价预测理论方法的发展历史。第三章论述了传统的基本分析和技术分析等传统的分析预测方法,对技术分析的理论知识做了比较详细的阐述。第四章主要论述了传统的基于线性范式下的资本市场理论所面临的挑战以及非线性市场理论的发展,并概念性地介绍了一些非线性的预测方法。第五章论述了基于时间序列分析的预测方法,着重阐述了 ARIMA 模型和 GARCH 模型的建模预测方法。第六章利用 ARIMA 模型和 GARCH 模型对代表上证指数波动趋势的 7 日移动均线进行了实证分析,取得比较好的拟合效果,并且做出了 6 日短期预测。

第三节 本文的创新与不足

本文的创新主要有:(1)本文所进行的对股票价格趋势的预测研究,不是对上海综合指数的整个历史数据区间建立模型,然后用这个模型预测将来某一天的价格。本文是基于局部数据进行建模,进行短期预测,如果及时对样本数据进行更新,使模型能够及时反映最新的信息,调整模型的结构,重估模型的参数,就可以实现动态化连续化预测。(2)一般的文献中多采用 GARCH 模型对收益率的波动率进行估计,本文在方法上也使用了 GARCH 模型,然而我们研究的对象是价格变化的趋势,同时使用的具体模型是在一般文献中不太常见的 AR-GARCH 模型,对上证指数趋势序列进行建模,并进行短期预测,取得

了不错的效果。

本文的不足主要有：（1）我们建立模型的对象是上证指数 7 日移动均线数据，而不是每天的收盘价。由于移动均线对收盘价存在一定的滞后效应，移动阶数越大，这种滞后效应越明显。所以如果用建立的模型来预测收盘价，则会存在一定的滞后效应。（2）预测对象是市场指数而非个股，对那些与指数波动性相似的股票的指导作用比较明显，但是对那些与指数波动规律差异较大的个股，指导作用就不那么强。

本文的结论主要有：（1）股价趋势预测的时间序列模型法有非常好的拟合效果，比一般的线性回归模型和多项式回归模型都要好；（2）ARIMA 模型和 GARCH 模型有着较好的短期预测效果，预测误差与预测步长同向变化；（3）我国股票市场价格波动规律具有时变性，不同样本时间窗内股价的前后相关性不同，建立的模型随着所依据的样本区间的不同而不同。实际预测中，应当进行动态化建模，从而实现连续预测。

第二章 股价预测理论与方法概述

全球股票市场从诞生至今有几百年的历史，但是，人们对它的理解和认识只是在近几十年才有了较大的发展。纽约大学马科维茨教授（90 年诺贝尔经济学奖获得者）在自传中提道“一次偶然的谈话提示我将数学方法应用于股票市场的可能性”。由此可以看出，人类用科学的方法对股票市场进行分析研究也只是从近几十年才开始的。

过去相当长的一段时间里，由于股票市场规模不大，信息交流速度和分析手段非常有限，人们通常只能凭借着个人的主观感受进行决策判断。随着近几十年证券市场的迅速发展，信息披露规范化和信息传输手段的多样化，人们逐渐认识到，需要一些行之有效的预测方法来对市场投资决策提供指导。

麦克尼斯发现自 20 世纪 70 年代前期开始，在所有重要转折关头，经济学家们做出的预测都是严重错误的，其中包括联邦储备委员会主席格林斯潘负责的研究所。麦克尼斯发现着这些转折关头，预测者往往作为一个群体发生偏离。即使预测正确，也往往只是在短期意义上的，“一个变量的很小变化往往产生比理论所预期的大得多的冲击”。尽管如此，人们还是以极大的热情投入市场预测工作，至少寄期望于短期预测的部分是准确的。大量研究人员通过对时间序列数据本身进行分析，从其自身的变化来寻找规律进而寻找新的方法来模拟和预测，但证券市场是一个复杂系统，它不但是一个有大量参与者的互动系统，而且还受很多市场以外的综合因素影响，它的客观表现确实难以预测和准确把握。然而由于对利益的追求和对未知规律的渴望，长期以来，人们从未停止过对股票市场的预测工作的研究和探索。

第一节 股票价格预测理论的发展史

格雷厄姆和多德在 1934 年《证券分析》一书中对 1929 年美国股票市场价格暴跌进行了深刻反思，认为股票价格的波动是建立在股票“内在价值”基础上的，股票价格会由于各种非理性原因偏离“内在价值”，但随着时间的推移这种

偏离会得到纠正而回到“内在价值”，因此，股票价格的未来表现可通过与“内在价值”的比较而加以判断。但“内在价值”取决于公司未来盈利能力，因此，对公司未来盈利能力及其现金流的准确把握成为预测未来股价的关键因素。此后，戈登在对“内在价值”进行深入的量化分析的基础上，提出了著名的股票定价的现金流量模型即“戈登模型”，然而，公司未来现金流是不确定的，为该模型的广泛应用带来麻烦，为此，关于股票定价的早期研究就集中在如何确定公司未来现金流。费雪(Fisher)教授认为未来资产收益的不确定性可以用概率分布来描述，马夏克、希克斯(Hicks)等学者经过一系列研究认为投资者的投资偏好可以看作是对投资于未来收益的概率分布矩的偏好，并可用均方差空间的无差异曲线来表示，同时，他们还发现“大数定律”在包含多种风险资产的投资中会发挥某种作用。戈登模型在股票价值分析中占有非常重要的地位，成为单只股票估价分析的基本方法，然而，该方法并没有解决股票投资风险与未来现金流折现率的关系，直到亨利·马科维茨教授的现代证券组合理论的建立才对这一基本问题有了明确的认识，从而一定程度上消除了该模型的致命缺陷。

在实际投资实践中，很少有投资者会将所有的投资集中在一只股票上。基于此，马科维茨教授于1938年提出了投资组合的概念，建立了现代证券组合理论，以统计学上的均值和方差等概念来衡量组合的收益和风险，给出了投资者如何根据自己的风险承受能力建立自己的最优组合以最大化其投资收益，并将风险分解为系统和非系统风险，从而，指导投资者最优化其投资行为。此后，其学生威廉·夏普(M·Sharpe)、林特纳等为强化该理论的应用，将其注意力从马科维茨的微观研究转向整个市场，将其复杂形态简化为以市场指数为基础的单因素关系，并发现在均衡市场条件下的资本资产的收益与风险遵循线性关系，即著名的以均值一方差模型为前提的资本资产定价模型(CAPM)。然而，由于CAPM所要求的前提过于严格，从而限制了其应用，许多经济学家试图研究在一定弱化条件下的定价理论，他们是迈耶斯的存在大量非市场化资产的投资定价理论、罗斯(Ross)的套利定价理论(APT)以及布里登(Breeden)资产收益率与平均消费增长率的线性关系模型(CCAPM)等等为数众多的数量化投资模型，为市场投资行为选择提供了一定决策依据。

Roberts 和 Osborne 在对股票市场价格的长期研究后，发现市场价格遵循“随

机漫步”或“随机游动”的规律，由此，以 Fama 教授为代表的经济学家提出了有效市场假说(EMH)理论，认为投资者对市场信息会作出合理的反应，将市场信息与股票价格相结合。有效市场假说(EMH)理论将有效市场分成三个层次：

一是强式有效市场(Strong-Form Efficiency)。在该市场中，现时的股票价格反映了所有信息，没有人能够利用包括内幕消息在内的任何信息获得超额收益。

二是半强式有效市场(Semi-Strong-Form Efficiency)。在该市场中，现时股票价格反映了所有公开的相关信息，如年报、简报、报纸专栏等等。人们不可能通过分析所有这些公开信息来获得超额收益。

三是弱式有效市场(Weak-Form Efficiency)。在该市场中，现时股票价格反映了所有过去的相关信息。人们不可能通过分析如过去的价格等历史信息来获得超额收益。

市场效率的决定因素是市场对不同信息的反应程度，包括历史信息、公开信息和内部信息。EMH 理论认为在弱式有效市场上，历史信息如股票过去的价格等对未来价格的影响已经体现在现有价格中，投资者一般不能简单依靠过去信息获取超额利润，过去的信息对未来的收益只有随机影响。

进入 1980 年代，在探寻一般均衡定价模型进展不大的情况下，将定价理论的研究方向转向注重市场信息的考察。经过实证检验，邦德特和塞勒(Bondt and Theler, 1985)发现股市存在投资者有时对某些消息反应过度(overreact)，而杰格蒂什(Jegadeesh, 1990)、莱曼(Lehmann, 1990)等则发现了股价短期滞后反应现象，由此，杰格蒂什和迪特曼(Titman, 1993)认为投资者对有关公司长远发展的消息往往有过度的反应，而对只影响短期收益的消息则反应不足，关于这一点仍然存在着争论，尽管如此，信息与股价之间应存在着某种关系得到了经济学家们的认同，并且，弗伦奇和罗尔(Roll)的实证研究证明了股价波动幅度与可获得信息量之间存在着良好的正相关关系。

然而，这些定价理论在现代经济金融学家的推动下得到巨大发展的同时也遇到了严峻的挑战，这种挑战表明了“对(股票、债券等)金融资产价格变动缺乏有效的解释手段反映了我们科学体系的不成熟”，股票市场普遍是混沌的。面对这一现实，金融学家们开始尝试利用非线性方法与混沌思想来理解股票市场

行为。

另外，人们为了从汪洋大海般的数据中将许多隐藏的有决策意义的信息挖掘出来，于 1989 年提出了数据挖掘的概念。它是从大量的、不完全的、有噪声的、模糊的、随机的数据中，提取隐含在其中的、人们事前不知道的、但有潜在有用的信息和知识的过程，又称为数据库中的知识发现。主要方法有灰色理论、模糊理论、混沌理论、神经网络和支持向量机等。金融时间序列是一类常见而重要的数据，对其进行数据挖掘分析也是目前的研究热点。

第二节 中国股票市场的可预测性

股市预测是经济预测的一个分支，指以准确的调查统计资料和股市信息为依据，从股票市场的历史、现状和规律性出发，运用科学的方法，对股票市场的未来走势做出判断。股市价格的未来走势是否具有预测性与市场的有效性密切相关。如果股票市场是符合有效市场假说的，也即股票价格服从随机游走，则股票价格的预测就毫无意义。

随着我国证券市场的发展，广大经济工作者对我国股票市场股票价格的预测也进行了深入的研究。先后有许多学者对我证券市场的有效性和可预测性进行了研究。从目前学者们对我国股票市场特征的研究来看，许多学者认为我国股票市场还没有达到弱势有效，也就是说用股票交易的历史信息在一定程度上可以预测其未来的走势。同时从笔者自身的投资经验来看，中国的股票市场在一定程度上，特别是在短期内，是具有一定的可预测性的。当然，我们说股价可以预测，并不是说可以 100%准确预见，而是说其短期趋势具备一定的可预测性。

近年来，我国的广大研究人员把随机过程和模糊数学以及信息、控制、人工智能、应用数学等专业大量应用于股票价格的研究，比如维纳过程、马尔科夫方法、“黑盒子”理论的灰色系统预测、系统动力学、蒙特卡罗模拟，甚至国际上前沿的混沌理论。也有一些学者从博弈的角度研究股票的价格波动，即庄家、散户、机构，甚至政府之间进行双方和多方的博弈。近年，博弈论经济学方面的经济学家接连获得诺贝尔经济学奖，人们也开始将博弈论广泛地应用在股票价格的模拟和说明上。

第三章 股价预测分析的传统方法

传统的预测分析方法主要是指基本分析方法和技术分析方法，这是目前仍然被广大投资者广泛使用的方法。

股票的市场价格由股票的价值决定，但是同时受到许多因素的影响。其中，供求关系是最直接的因素，其他因素都是通过作用于供求关系而影响股票价格的。由于影响股票价格的因素复杂多变，所以股票的市场价格呈现出高低起伏的波动性特征。这些影响因素基本上可分为以下三类：市场内部因素、基本面因素、政策因素。

一、市场内部因素

市场内部因素主要是指市场的供给和需求，即市场资金量和股票供给量，比如一定阶段的股市扩容节奏将对该因素产生重要影响。

二、基本面因素

基本面因素包括宏观经济因素和公司内部因素。宏观经济因素主要是指能影响市场中股票价格的因素，包括经济增长、经济景气循环、利率、财政收支、货币供应量、物价、国际收支等。公司内部因素主要是指公司的财务状况。

三、政策因素

政策因素指足以影响股票价格变动的国内外重大活动以及政府的政策、措施、法令等重大事件。政府的社会经济发展计划、经济政策的变化、新颁布法令和管理条例等均会影响到股价的变动。

第一节 基本分析方法

一、基本面分析的基本原理

基本面分析是指证券分析师根据经济学、金融学、财务管理学及投资学等基本原理，对决定证券价值及价格的基本要素，如宏观经济指标、经济政策走势、行业发展状况、产品市场状况、公司销售和财务状况等进行分析，评估证券的投资价值，判断证券的合理价位，提出相应的投资建议的一种分析方法。

基本面分析重点研究股票的内在价值—上市公司经营状况的真实反映。这种分析方法认为股票价格在市场上的频繁波动是受多种内外因素影响的，但股票的市场价格总是围绕着内在价值变化，市场价格和内在价值之间的差距最终会被市场纠正，因此市场价格低于(或高于)内在价值之日，便是买(卖)机会到来之时。影响股票市场价格波动的因素很多，主要的有三个方面的原因：一是宏观经济的繁荣或萧条，二是企业所处行业以及相关行业的发展状况，三是企业自生的经营状况，如经营得当，盈利能力比较强，则它的股票价值就高，价格就贵，反之，则价值就低，价格就贱。

二、基本分析应考虑主要因素：

(一) 经济环境

GDP、国民收入(涵盖个人收入和个人可支配收入)、经济周期分析

(二) 经济政策

货币政策：准备金制度、再贴现制度、公开市场业务

财政政策：税收、预算和平衡

经济学家普遍认为，在刺激经济增长时，财政政策比货币政策的效果更明显，而在对付通货膨胀方面，货币政策比财政政策更加直接和有效。分析政府的财政政策有助于掌握经济运行的走势，预测公司收益的变化和证券价格的变化趋势。

(三) 行业因素

产业发展周期：

初创阶段、成长阶段、成熟阶段、衰退阶段

经济运行周期：

周期性行业、稳定性行业、增长性行业

(四) 公司状况

竞争能力、盈利能力、财务状况、经营管理能力

第二节 技术分析方法

基本分析的目的在于确定证券的内在价值，然后将内在价值与市场价格进

行比较，若前者高于后者，说明证券价值被低估，则买进；反之则卖出。而技术分析则着眼于对证券价格走势和影响价格的供求状况的分析，它以市场价格、交易量这些历史信息为基础，凭借图表和各种指标来解释、预测市场的未来走势，强调心理因素对证券价格走势的影响。换言之，利用技术分析作买卖决策时，也许不用花费时间、精力去做缜密的宏观、行业和公司研究，甚至无须知道公司的名称，投资者只要坐在电脑前，研究该公司的证券价格走势及价格所表现的各种指标和图表，就可以从中寻找到买卖时机。因而有人称技术分析侧重于对投资时机的把握。

技术分析是当今证券投资中常用的一种重要分析方法，它依据过去与现在的统计资料数据，运用图表、形态、指标等分析手段，对证券价格的发展趋势进行各种有针对性的分析研究，从而作出对证券价格波动发展趋势的预测估计。由于技术分析方法运用了极为广泛的数据资料，并采用了各种不同的数据处理方法，因而受到了投资者的重视和青睐。

一、技术分析的理论基础

（一）市场行为说明一切

这一观点认为所有影响证券市场波动的因素，最终都必然体现在证券的价格波动之上，外在的、内在的、基本面的、政策面的、心理面等诸多影响市场走势因素必然会最终转化成为对于市场筹码的供给和需求，从而体现在价格走势之中。而投资者只需关心这些因素对于市场行为的影响效果，而无需关心具体导致这些变化的原因是什么。

这个假设有充分的合理性。任何一个因素对股票市场的影响最终都必然体现在股票价格的变动上。如果某一个消息公布以后，股票价格同以前一样而没有大的变动，说明这个消息不是影响股票市场的因素。如果有一天我们看到，价格向上跳空很多，成交量急剧增加，不用问，一定是出现了一个或者几个利多消息，具体是什么消息，完全没有必要过问，因为它已经体现在市场行为之中了。反之，如果某一天价格向下跳空很多，成交量大增，也是一定出现了什么利空消息，上述现象就是这个消息在股票市场行为中的反映。

（二）历史会重演

价格的波动表面上是受到证券的供求关系所决定的，但是就其深层次而言，

影响市场供求关系的是投资者预期变化而导致的对于筹码供需的变化。投资者往往会因为赢利面和赢利程度的不断扩大而提高对于未来市场走势的预期，这样当市场进入最高点的时候，投资者的预期值到达最大值，对于筹码的需求达到最大，而持有筹码的投资者则抛出筹码的预期到达最低值（极少部分先知先觉的投资者及好运的投资者出外），这样便自然形成市场价格的最高点，这时正好是人性的“贪婪”本性也到达了最高点；相反投资者同样也会因为亏损面和亏蚀程度的加剧而不断降低对于未来市场的预期，当市场进入最低点的时候，投资者的预期值到达最低值，对于筹码的需求到达最小，而持有筹码的投资者的抛出意愿达到最高值（极少部分先知先觉的投资者及好运的投资者除外），这样便自然形成了价格的最低点，而这时恰恰也是人性“恐惧”本性达到了最高值。因而对于“历史往往重复发生”的理解应该是人性的“贪婪”和“恐惧”本性永远在市场波动的高、低点过程中循环往复地同步发生着。

股票市场是一个买卖双方的市场，价格的变动每时每刻都受到供需关系的约束。股票价格上涨了肯定是需求大于供给。反之，股票价格下跌了，肯定是供给大于需求。在进行技术分析时，一旦遇到与过去某一个时期相同或者相似的情况，应该与过去的结果比较。过去的结果是已知的，这个已知的结果是现在对未来作预测的参考。任何有用的东西都是经验的结晶，是经过许多场合检验而总结出来的。这就是历史可重复性，股票市场的操作也不例外。

（三）价格沿趋势运动

承接“历史往往重复发生”是投资者预期由高至低或由低至高而不停的变化理解，这一句话背后的意义十分的清晰明了。当市场从高点回落时，投资者的预期也从“贪婪”的极点同步回落并影响了投资者对于证券的需求的降低和供给的扩大，因而价格在这一心理趋势变化的影响之下，自然便沿着下降的趋势而不断回落，直到达到预期的最低值；同理，当价格从最低点逐渐上升时，投资者的预期也逐步摆脱“恐惧”的阴影而不断上升，并影响了投资者对于证券需求的放大和供给的降低，因而价格同样在心理预期不断增强的影响下，自然沿着上升趋势不断创出新高，一直达到预期的最高值。

这一假设认为股票价格的变动是按照一定规律进行的，股票价格有保持原来方向的惯性。正是由于这一条，技术分析者才会花费大量心血，看图找出价

格变动的规律。一般来说，一段时间内股票价格一直是持续上涨或者下跌，那么，今后一段时间，没有出现意外，股票价格会按照这个方向继续上涨或者下跌，没有理由改变这一既定的运动方向。“顺势而为”是股票市场中的一句名言，如果没有调头的内部和外部因素，没有必要逆大势而为。

如果否认了这一假设，认为即使没有外部因素影响，股票价格也可以改变原来的变动方向，技术分析就没有了立根之本。试想，股票价格可以随意变化，没有任何规律可循，技术分析就如同赌博一样撞大运，技术分析还有存在的必要吗？只有股价的变动遵循一定规律，我们才一能运用技术分析这个工具找到这些规律，对于今后的买卖活动进行有效的指导。

二、技术分析的要素：量、价、时、空

（一）价和量是市场行为的最基本的表现

市场行为最基本的表现就是成交价和成交量。过去和现在的成交价、成交量涵盖了过去和现在的市场行为。技术分析就是利用过去和现在的成交量、成交价资料，以图形分析和指标分析工具来分析、预测未来的市场走势。这样，成交价和成交量就成为技术分析的基本要素。

（二）时间在进行行情判断时有着很重要的作用

一个已经形成的趋势在短时间内不会发生根本改变，中途出现的反方向波动，对原趋势不会产生大的影响。一个形成了的趋势又不可能永远不变，经过了一定时间又会有新的趋势出现。循环周期理论着重关心的就是时间因素，它强调了时间的重要性。

（三）空间指的是价格波动所能够达到的极限

从某种意义上讲，空间因素也可以认为是价格的一个方面。

三、技术分析的方法体系

在价、量历史资料基础上进行的统计、数学计算和绘制图表的方法是技术分析方法的主要手段。技术分析的方法繁多，形式多样，本章仅介绍一些比较常用和比较实用的技术分析方法。一般说来，可以将技术分析方法分为如下五类：指标类、切线类、形态类、K线类、波浪类。

（一）指标类

指标类要考虑市场行为的各个方面，建立一个数学模型，给出数学上的计

算公式，得到一个体现股票市场的某个方面内在实质的数字，这个数字叫指标值。指标值的具体数值和相互间关系，直接反映股票市场所处的状态，为我们的操作行为提供指导方向。指标所反映出来的许多信息是不能从行情报表信息中直接看出来的。

（二）切线类

切线类是按一定方法和原则在由股票价格的数据所绘制的图表中画出一些直线，然后根据这些直线的情况推测股票价格的未来趋势，这些直线就叫切线。切线的作用主要是起支撑和压力的作用。支撑线和压力线的往后延伸位置对价格趋势起一定的制约作用。

（三）形态类

形态类是根据价格图表中过去一段时间走过的轨迹形态来预测股票价格未来趋势的方法。根据技术分析的假设，市场行为包括了一切信息。价格走出的形态是市场分析的重要部分，是股票市场对各种信息进行消化后的具体表现。从而用价格的形态来推测股票价格的将来就十分有意义。

（四）K 线类

K 线类的研究手法是侧重若干天的 K 线组合情况，推测股票市场多空双方力量的对比，进而判断股票市场多空双方谁占优势，是暂时的，还是决定性的。K 线图是进行各种技术分析的最重要的图表。

（五）波浪类

波浪理论把股价的上下变动和不同时期的持续上涨、下跌看成是波浪的上下起伏。波浪的起伏遵循自然界的规律，股票的价格运动也就遵循波浪起伏的规律。波浪理论较之于别的技术分析流派，最大的区别就是能提前很长的时间预计到行情的底和顶，而别的流派往往要等到新的趋势已经确立之后才能看到。但是，波浪理论又是公认的较难掌握的技术分析方法。

四、技术分析的发展趋势

近十几年来，证券分析技术出现了一些新的发展趋势。计算机技术的进步使各种分析、预测理论得以和统计学、信息技术、优化算法紧密结合，使证券分析向数量化、技术化、复杂化方向发展。将随机过程，混沌理论、模糊数学等数学理论和神经网络、人工智能等工程学手段应用于股票价格预测逐渐成为

了现在股票价格预测分析的热门领域。

西方金融理论界对技术分析的研究主要集中于两个方面：寻求价格运动规律存在的统计支持；研究基于技术分析的交​​易规则，获得更直接的实证支持。比如在统计上对证券收益中厚尾现象（fat tail）的发现，使得一些研究者采用分形和混沌理论来解释证券的价格行为，并得到很好的效果。还有贝叶斯统计、神经网络、人工智能等方法，这些工程学的手段也逐步使用在各种投资软件和自动交易系统中。与简单的移动平均等方法相比，现代技术分析在预测技术方面达到了更高的境界。

五、技术分析的使用方法

技术分析的各个流派从不同的方面理解和考虑股票市场，有的有相当坚实的理论基础，有的就没有很明确的理论基础，很难说清楚为什么。但它们都有一个共同的特点，即都经过股票市场的实践考验。这五类技术分析方法尽管考虑的方式不同，但目的是相同的，彼此并不排斥，在使用上可相互借鉴

在股票投资中，投资者需要解决的两个问题，一是投资什么样的股票，二是何时买卖。这也正是基本面分析和技术面分析设法解决的问题。在现实的经济活动中，股票价格的起伏变动不仅是由上市公司的经营状况决定的，各种外部因素的变化也会直接或间接地影响股票的走势。基本分析帮助我们选择股票的种类，技术分析可以帮助我们选择合适的买卖时机。这技术与基本分析不是排斥的，两者应当结合使用。

第四章 股价预测的非线性理论与方法

第一节 线性资本市场理论的局限性

一、线性范式的发展与运用

长期以来，线性范式在现代资本市场理论中一直处于主导地位。线性范式指的就是以 EMH 为基础，以线性函数关系为模型的资本市场理论框架。马柯威茨提出的均值一方差模型和夏普等人建立的资本资产定价模型 CAPM 是传统线性范式下资本市场理论的核心部分。

EMH 在二十世纪 70—80 年代备受推崇，按照当时人们对 EMH 的理解，如果资本市场是有效的，那么：资产价格已经反映了所有可获得的或公开的信息，过去的信息已经被计算到证券价格里面了；投资者是理性的，是以线性的方式对信息做出反应，即他们在得到信息时就立即做出反应，而不是以累计的方式对一个事件做出反应；证券收益率则是相互独立的、呈正态分布的随机变量，且其波动服从随机游走的模型。

二、线性范式的缺陷

然而，近些年来线性范式受到了空前的挑战，这主要是因为有不少的实证研究表明有效市场假设与人们所观察到的事实并不相符。主要事实有：

（一）受益率分布的非正态性

国内外大量实证研究发现，证券的收益率并非相互独立，其变化并不服从随机游走模型，其概率分布亦非正态分布，而呈现一种“尖峰厚尾”（也可称为“尖峰胖尾”）的特性，即同正态分布相比，总体在众数周围的集中程度很高，且大幅度偏离均值的异常值较多。

（二）方差度量资本市场的风险失效

根据概率论知识，股票收益的厚尾特性可能导致方差不确定或无穷大，此时是不能用方差来度量资本市场风险程度的。因此，股票收益的厚尾特性会直接导致方差度量资本市场风险的失效。此外，对收益率不呈正态分布或对称分布的市场来说，即使方差是有限的也不能有效衡量风险的大小。这就直接导致

了在 EMH 基础之上发展起来的 CAPM 等模型不能使用。

（三）规模效应和季节效应问题

规模效应即股票的收益率无论是总收益率还是风险调节后的收益率，都与公司的规模大小负相关。季节效应是指股价指数一月份的收益率总明显高于其它 11 个月份的收益率。除一月效应以外，还有周末效应、节日效应等。在很多证券市场上，都被发现规模效应和季节效应的存在。

非线性范式真正引起人们的注意是在 20 世纪 80 年代后期，人们越来越多地发现，线性范式下的资本市场理论已无法解释股市的一些现象，最为典型的是黑色星期一问题。所谓黑色星期一，指的是在 1987 年 10 月 19 日星期一，美国纽约股票市场突然爆发了前所未有的狂跌风潮。依据传统的资本市场理论，在有效市场中，价格趋向合理，市场趋于均衡，除非外界发生诸如战争及其他重大政治事件，否则不会出现股市暴涨暴跌这样的巨大变动。由于当时根本没有这类外界事件的发生，所以传统的资本市场理论无法对这一现象给出令人满意的解释。黑色星期一事件加速了非线性资本市场理论的发展。

第二节 非线性资本市场理论的发展

一、分形市场理论(Fractal Market Theory)的产生为资本市场理论的发展开拓了新的领域。

早在 1972 年，Mandelbrot 就首次提出可以用分形理论中的 R/S 分析法即重标极差法(Rescaled Range Analysis)来分析金融时间序列，随后 Green 和 Fielitz 对股票收益数据作了 R/S 分析，得出结论认为股票价格时间序列的确存在分形特征。20 世纪 80 年代后，R/S 分析在金融理论界得到了充分的重视，不少学者在这方面作了实证研究，但是得出的结论却不尽相同。Aydogan 和 Booth.Lo 以及 Ambrose eatl 在对股票收益率作了 R/S 分析之后，认为没有充分证据表明股票价格序列存在分形特征；而其他入，比如 Peters 和 Goetzmann 则认为分形特征在股票价格序列中是存在的。从他们具体的研究过程来看，Peters 和 Goetzmann 的结论更令人信服。因为他们研究所使用的历史数据链在同类研究

中是最长的，而且在进行 R/S 分析之前对原始序列作了去除趋势处理，对原始数据的短期序列相关作了调整。

需要特别指出的是，从 Peters 在 1994 年的研究中可以很清楚地看到非线性范式的确能够揭示出股票市场的内在特性。他对 1928 年 1 月 2 日至 1990 年 7 月 5 日的股票收益率数据进行 R/S 分析后发现，在这 63 年里，尽管美国经历了二战、朝鲜战争、越南战争、经济大萧条、20 世纪 60 年代社会动乱、70 年代石油危机以及 80 年代大规模举债经营，但用来表征股票市场非线性特性的 Hurst 指数却显得相当稳定，一直保持在 0.57—0.62 的范围内。此外，他还针对某些个股的月收益率作了 R/S 分析，结果表明按产业分组的股票倾向于有相近的 Hurst 指数值和大体相当的循环长度。Peters 在通过大量证明 EMH 失效和市场具有分形特征后，提出了“分形市场假说” (Fractal Market Hypothesis, 简记为 FMH)。

作为非线性分析方法的分形理论在运用时由于没有独立、正态和随机游走的前提约束，并且能够更好地刻画资本市场的特征，因此能够成为资本市场理论的有效分析工具，从而为资本市场理论的发展开拓了新的领域。并且，这一理论得到了越来越多的人的认可。分形理论能更好地刻画资本市场的运行机制，有助于我们认识资本市场的真实本质。经典资本市场理论认为资本市场以线性方式对外界作用起反应，并且其长期变化趋势由资本市场的内在因素决定，其短期的不规则波动由于外在的随机因素影响，于是可以通过线性方程加上随机项的方法为资本市场建立数学模型。实际上，资本市场的 irregular 波动并非外随机性作用的结果，而是由资本市场内部的非线性关系相互作用产生的。这种非线性效应表现为股市股价的巨幅变动及价格时间序列的高度自相关等等。可见，线性范式下的传统资本市场理论无法很好地揭示资本市场的本质特征。分形理论则是解决资本市场中非线性问题的有力工具，它能有效地刻画出资本市场内部相互作用的非线性关系，从而更准确地揭示出资本市场的内在特征，深化人们对资本市场这一非线性系统的认识。分形理论是非线性分析方法中的一种，尽管问世才几十年，但是由于它能有效地解释自然和经济现象中许多极其复杂多变的问题，因此它已成为发展非线性资本市场理论框架的有力分析工具，目前学术界也已普遍意识到其重要性。

二、混沌理论与方法在资本市场理论中的应用显示出强大的生命力。

1989年,美国学者 Scheinkman 和 Lebaron 发现了股票日收益率序列和周收益率序列中存在着混沌现象。此后,对各种经济与金融时间序列数据的非线性检验引起了人们广泛的兴趣。

混沌是指确定性系统中出现的类似随机的过程。表现为外观的无秩序性和内在的规律性的结合,具有以下几个基本特性。一是蝴蝶效应,又称为对初始条件敏感的依赖性,也就是说在混沌出现的参数范围内,初始条件的一个微小差异在迭代过程中会不断地被放大,使迭代结果变得极为不同。二是内随机性,系统其紊乱的根源来自系统的内部而不是外部,是确定性系统内部随机性的反映,系统由完全确定性方程描述、无须附加任何随机因素。三是其吸引子具有分形特征。人们把因时间增加,来自某一区域的所有轨迹趋于耦合的相空间区域称为“吸引子”。混沌系统的吸引子具有分形结构,必须用分形维数来描述其特征,被称为“奇异吸引子”。

资本市场的形成和发展是一个动力学过程,具有混沌特性。这是因为:首先,资本市场复杂多变,其运动状态的复杂性是系统内部因素相互作用导致的,并不是来自外部环境的干扰。即使处于某些完全确定性的条件下,不附加任何随机因素就可以出现与股市涨落相近的行为,这就表明资本市场确实具有内在随机性;其次,资本市场任一时刻的状态都受到前一时刻状态的影响,状态演化对初始值细微变化的敏感依赖性,使我们不可能准确预测市场的长时间的演化行为,这就说明资本市场确实具有蝴蝶效应;第三,资本市场的混沌现象可能出现在简单的系统中,当参数超过了一定的临界值就会出现混沌,产生奇异吸引子,资本市场与奇异吸引子的基本特征有许多相似之处。

到目前为止,我们运用混沌和分形等非线性方法对资本市场的非线性特征及预测理论进行的探讨只是初步的,要揭示资本市场演化和发展的内在复杂性还有很多问题有待于进一步研究。但是,近几年的研究成果表明,将非线性科学的理论和方法运用于资本市场研究中具有巨大的优势和潜力,显示出强大的生命力。目前,我国已有很多学者开始从事这方面的研究,并取得了一定的成果。

三、其他非线性理论与方法在资本市场研究中的应用。

除了用分形与混沌等非线性理论来研究资本市场之外，还有一些其他的范式试图解释传统资本市场理论不能解释的一些现象，以揭示资本市场演化发展的内在复杂性。下面介绍的几种理论相对较新，它们都与分形、混沌等非线性理论有着紧密的联系，可视为用非线性理论来研究资本市场的几个发展方向：

（一）突变理论

突变理论 (Catastrophe Theory) 的创始人是法国数学家雷内·托姆 (Rene Thom)，他于 1972 年出版的《结构稳定性和形态发生学》一书阐述了突变理论，荣获国际数学界的最高奖——菲尔兹奖章。突变论的出现引起各方面的重视，被称之为“是牛顿和莱布尼茨发明微积分三百年以来数学上最大的革命”。

40 多年来，突变理论已在物理学、化学、生物学和工程学等领域取得了广泛的应用性成果，在社会科学方面也有许多应用，在对经济危机、通货膨胀、股票市场等非连续变化的质变现象的研究方面取得了多项成果。与以往微积分解释连续而光滑变化过程不同，突变理论描述的是自然界中大量存在的不连续的、突然变化的过程。突变理论以拓扑学为基础，研究梯度系统在平衡状态下的临界点的性态，描述由逐渐变化的力量或运动而导致突然变化的现象，这是微分方程描述方法所不能胜任的。相对于其他理论，突变理论可以更好地刻画资本市场运动的某些规律，为解释资本市场中价格的变化提供新的手段。资本市场是一个十分复杂的系统，除了连续的渐变的过程外，大量存在的是不连续的、突变的过程。可以说，突变是资本市场的一个基本特性，一个重要的非线性特征。这是因为从众心理所产生的群体行为可能导致股票价格猛涨和猛跌，对这种股票市场剧烈变化的行情，无法完全归结于噪声（即随机信号）的干扰，而是属于典型的突变现象。因此，用突变理论研究资本市场的非线性特征及预测理论显然有其用武之地。

（二）协同市场假说

瓦加 (Vaga) 在 EMH 基础上加进了非线性因素提出了协同市场假说 (Coherent Market Hypothesis, 简记为 CMH)。CMH 是一个资本市场的非线性统计模型，它认为资本市场出现的许多异象或非线性现象与 EMH 并不矛盾，所谓的异象或非线性也好，市场有效也好，它们都只不过是资本市场这一非线性系统在不同条件下因主导因素组合发生变化而产生的不同状态而已。CMH 认

为,市场的概率分布主要取决于两种主导因素的不同组合,一是基本经济环境,二是市场中存在的投资者情绪偏倚量或“集体思维”的水平。这两种主导因素组合在时间上的变化,就引起了市场状态的变化,市场的概率密度函数的形状也随之发生变化。瓦加认为,市场随主导因素组合的变化可出现下列四种状态:一是随机游走。在投资者彼此完全独立作出投资决定的条件下,所有信息会很快反映到当前价格之中,价格的变动就服从随机游走模型,这时资本市场就是EMH下的情形;二是过渡市场。随着投资者“集体思维”形成和程度的不断提高,投资者情绪中的偏倚可以使得信息的影响延展至一个很短的时间区间,市场不再有效;三是混沌市场。投资者情绪对于集体思维有很强的传导力,但基本状况是中性的或不确定的。集体情绪可能出现大的摇摆,致使市场出现骤变;四是协同市场。强有力的正面(负面)基本状况与强烈的投资者情绪相结合,就可能导致协同市场,这时趋势可能是强有力正面(负面)的,而风险很低(高)。

由此可见,CMH给EMH及众多争议之间的矛盾作出了一个很好的解释,从另一个角度看可视为EMH加入非线性因素后的版本,虽尚待作更多的实证研究,但CMH所提出的框架应是资本市场非线性研究的一个发展方向。

(三) 小波理论(Wavelet Theory)

小波理论是谱分析的一个推广,小波变换被誉为“数学显微镜”。谱分析是时间序列分析的一个重要工具,依赖于Fourier变换,谱分析可把一个信号分解成一个个正弦波序列,这些正弦波序列加在一起就可以重现原来的信号。但谱分析要求系统有一个特征标度,即每一个更小的增量按照一个固定的倍数缩放。作为谱分析的一个推广,小波理论是当前数学中一个迅速发展的新领域,它具有理论深刻和应用广泛的双重意义。与Fourier变换、窗口Fourier变换(Gabor变换)相比,小波变换是一个时间和频率的局域变换,因而能有效地从信号中提取信息,通过伸缩和平移等运算功能对函数或信号进行多尺度细化分析(Multiscale Analysis),解决了Fourier变换不能解决的许多困难问题。对于资本市场价格波动这样一个具有混沌或分形特征的时间序列,由于无特征标度,谱分析无法对之进行有效分析,而小波理论可以处理具有多重标度的信号,因此可以用来分析分形和混沌时间序列,这可能是未来资本市场研究的一个重要的领域。

到目前研究为止，资本市场的非线性理论在很多方面、以多种形式得到了发展，但从根本上讲，只是发现了资本市场具有一些非线性特征的证据，对某些问题的研究有了一定的设想，对经典资本市场理论难以解释的某些现象给出了一些较为合理的解释，要建立起一套考虑所有非线性效应的资本市场的描述和预测模型，能够对资本市场演化规律做出科学合理的描述和预测，还有很长的路要走。

第三节 股价的非线性预测方法

资本市场是一个复杂的非线性系统，市场中证券价格的波动是由影响市场的内部、外部等多种因素之间复杂的关系相互作用所决定的。外部因素主要有国家的经济政策、政治方针、上市公司经营行为、经营业绩等；内部因素主要有投资者的心理预期和操作策略等，它们的变化直接导致证券价格的波动。不管是内部因素还是外部因素，都不能独立地影响这个市场，而是相互渗透、相互制约、共同作用的。这种影响的复杂性，使得资本市场价格的波动具有不确定性。正是在这种情况下，一些新的建模和预测技术发展开来。主要包括以下几个方向。

一、动力学方法

Newton 和 Laplace 都曾经断言，只要建立起系统的动力学方程，就可以根据初始条件和边界条件准确地确定系统任何时刻以后的运动。在 Newton 所创立的经典力学体系里，任意一个确定的系统在确定性的激励下，其响应必然也是确定的，即因果之间是一一对应的，这就是所谓的确定性论。通过动力学途径进行预测的方法其理论基础就是这种确定性论，它是建立在人们已经熟悉的物理规律上，如运动的连续方程、各种守恒方程等，通过对系统的简化用这些模型表示待预测系统，并通过求数值解达到预测的目的。由于一般系统运动过程本身的错综复杂性，使得我们很难对其物理机制进行充分了解，特别是简化过程中的假定是否符合实际也难以作出判断，因此这类方法虽然有一定的科学依据，但其可靠性也常常受到人们的怀疑。

二、混沌学方法

动力学预测方法的理论基础是确定性论,认为系统是完全确定的。统计学预测方法的理论基础是概率论,认为系统是随机的。随着混沌现象的发现,使人们认识到确定性动力系统的行为不仅仅是定常、周期和拟周期的,确实还有大量随机性的,而这种随机性不同于概率论中的随机性,是一种内在随机性。因此,动力学方法和统计学方法对于具有内在随机性的所谓混沌系统的预测都会失效。混沌学预测方法的理论基础是混沌理论,它是依据系统的混沌特征(如内在随机性、对初值的敏感依赖性等)对系统的未来行为作出预测。

需要指出的是,混沌理论认为:即使是完全确定的模型,经充分精确的数值求解,所获得的长时间演化结果也可能是类随机的,且系统对初始条件具有敏感依赖性,因而使得长时间预测是不可能的。但是,在短时期内,系统运动轨道发散应该较小,从而利用观测资料即时间序列进行短期预测是可行的。因此,混沌学预测方法都是针对短期预测的。

三、灰色理论方法

灰色系统理论(Grey System Theory)是控制论的观点和方法延伸的产物,也是自动控制与运筹学的数学方法结合的结果。20世纪80年代初由华中理工大学邓聚龙教授提出并进一步发展了灰色系统理论。灰色预测法是一种对含有不确定性因素的系统进行预测的方法,已经广泛地应用于农业、气象、生态、水利和生物等领域,在经济和管理决策中也取得了显著的成果。

灰色理论运用颜色的深浅来表示系统信息的完备程度,把一个系统的内部特征是完全已知的,即系统的信息是完全已知的信息系统白色系统;一个内部对外界来说是一无所知的,只能通过它与外部的联系来加以观测研究的信息系统称为黑色系统;灰色系统是介于白色系统和黑色系统之间的一种系统,即系统内的一部分信息是已知的,系统内各因素间具有不确定性的关系。

灰色预测通过对系统因素之间发展趋势的相异程度进行分析,并对原始数据进行生成处理来寻找系统变动的规律,生成有较强规律性的序列,然后建立相应的微分方程模型,从而预测实物未来发展趋势的状况。灰色时间序列预测是用观察到的反应预测对象特征的时间序列来构造灰色预测模型,预测未来某一时刻的特征量,或达到某一特征量的时间。

四、组合理论方法

随着现代预测理论与方法的发展，各种方法应运而生，预测者很难给出最佳选择。但是组合预测无论在理论上，还是在时间中都是一种值得重视的预测方法，许多研究都表明了这种方法具有比单一方法更为有效的预测能力，因此，组合预测颇受青睐并获得成功。

面对复杂的社会经济系统，单准则预测以及单方法预测难以解决决策者所提出的问题。针对某一具体问题，系统分析者或者决策者往往面对的是来自不同专家、不同预测模型所提供的预测与分析结果。从某种意义上讲，每一种方法，不同专家的预测均基于不同的信息集并在一定的假设条件下给出，它们各自的环境及性能指标下是最优的，但都不可避免地具有各自的局限性。因此，如何有效地处理来自不同领域的信息，有效地支持决策就是广大系统分析者所面临的实际课题。研究表明，每一种预测方法在数据处理及不同准则方面均有其独到之处，都能从不同的角度为最终预测提供有效信息，很难建立一套系统准则来评价预测的好坏，因此预测者趋向于不只是简单地选择最佳模型而抛弃余下模型，因为单个模型毕竟没有充分利用所有有效的信息。这导致了组合预测的研究，主要包括非线性组合预测、混合预测等方法。

第五章 股价预测的时间序列模型方法

时间序列预测是 20 世纪以来学术界的研究热点,尤其是随着证券市场的迅速发展,金融时间序列成为众多学者关注的焦点。2003 年诺贝尔经济学奖得主美国经济学家罗伯特·恩格尔和英国经济学家克莱夫·格兰杰,他们的重要贡献就是将统计学方法用于研究经济时间序列,从而引发了对经济时间序列的研究热潮。

第一节 时间序列分析方法概述

一、时间序列分析方法的发展简史

所谓时间序列就是按照时间的顺序记录的一系列有序数据。对时间序列进行观察、研究,找寻它的变化发展规律,预测它将来的走势就是时间序列分析。作为数理统计学的一个专业分支,时间序列分析遵循数理统计学的基本原理,都是利用观察信息估计总体的性质。但是时间序列的不可重复性,使得我们在任意一个时刻只能获得唯一的序列观察值,这种特殊的数据结构导致时间序列分析有它非常特殊的、自成体系的一套分析技术。

时间序列预测技术源远流长,一般分为确定性时间序列方法和随机性时间序列方法。确定性时间序列预测应用很早,随机性时间序列预测则出现在上世纪二战以后,一方面是由于计算机技术的飞速发展成为随机性时间序列分析的建模和预测提供了强有力的工具;另一方面则要归功于美国统计学家 *George E.P.Box* 和英国统计学家 *Gwilym M.Jenkins* 的工作。他们于 1968 年提出了一套较为完善的随机时间序列分析和建模理论,并于 1970 年出版专著《时间序列分析—预测与控制》。以后这种技术便迅速发展,在经济预测、气象预报和过程控制等方面得到了广泛应用。为了纪念他们二人的杰出工作,人们常称这类模型为 *B-J* 模型。

时间序列预测的基本思想是将预测对象随时间推移而形成的数据序列视为一个随机序列,即除去个别的因偶然原因引起的观测值外,时间序列是一组依

赖于时间的随机变量。这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性，而这种自相关性一旦被相应的数学模型描述出来，就可以从时间序列的过去值及现在值来预测其未来值。这种方法是基于预测对象的变化仅与时间有关的假设前提下进行的，它简化了外部因素对预测对象的影响，使得预测更为直接简便。

时间序列预测模型还可分为传统时间序列预测模型与现代时间序列预测模型。传统时间序列预测模型有移动平均与分解模型、指数平滑模型等；现代时间序列预测模型则包括 *ARIMA* 模型族、*ARFIMA* 模型、*ARCH / GARCH* 族、*SV* 模型等。这类方法在中国股票市场预测问题研究中的应用也比较多。

二、时间序列的定义

在统计研究中，常用按时间顺序排列的一组随机变量

$$\cdots, X_1, X_2, \cdots, X_t, \cdots$$

来表示一个随机事件的时间序列，简记为 $\{X_t, t \in T\}$ 或 $\{X_t\}$ 。用

$$x_1, x_2, \cdots, x_n$$

或

$$\{x_t, t = 1, 2, \cdots, n\}$$

式 (1.2) 表示该随机序列的 n 个有序观察值，称之为序列长度为 n 的观察值序列，有时也称式 (1.2) 为式 (1.1) 的一个实现。我们研究的目的是揭示随机时序 $\{X_t\}$ 的性质，而要实现这个目标就是通过分析它的观察值序列 $\{x_t\}$ 的性质，由观察值序列的性质来推断随机时序 $\{X_t\}$ 的性质。

三、时间序列的分析方法概述

(一) 描述性时序分析

早期的时间序列分析都是通过直观的数据比较或者绘图观察，寻找序列中蕴含的发展规律，这种分析方法就被称为描述性时序分析。古代埃及人发现尼罗河泛滥的规律就是依靠这种分析方法。而在天文、物理、海洋学等自然科学领域，这种简单的描述性时序分析也常常能使人发现意想不到的规律。描述性时序分析方法具有操作简单、直观有效的特点，从史前直到现在一直被人们广

泛使用，它通常是人们进行统计时序分析的第一步。

（二）统计时序分析

随着研究领域的不断拓广，人们发现单纯的描述性时序分析有很大的局限性。在金融、保险、法律、人口和心理学等社会科学研究领域，随机变量的发展通常会呈现出非常强的随机性，想通过对序列简单的观察和描述，总结出随机变量的发展变化规律，并准确预测出它们将来的走势通常是非常困难的。为了更准确地估计随机序列发展变化的规律，从 20 世纪 20 年代开始，学术界利用数理统计学原理分析时间序列。研究的重心从表面现象的总结转移到分析序列值内在的相关关系上，由此开辟了一门应用统计学学科——时间序列分析。

纵观时间序列分析方法的发展历史可以将时间序列分析方法分为两类。

1. 频域分析方法

频域（*frequency domain*）分析方法也称为“频谱分析”或“谱分析”（*spectral analysis*）方法。

早期的频谱分析方法假设任何一种无趋势的时间序列都可以分解为若干种不同频率的周期波动，借助傅立叶分析从频率的角度来揭示时间序列的规律，后来又借助了傅立叶变换，用正弦、余弦项之和来逼近某个函数。20 世纪 60 年代，Burg 在分析地震信号时提出了最大熵谱估计理论，该理论克服了传统分析固有的分辨率不高和频率泄漏等缺点，使谱分析进入一个新的阶段，我们称之为现代谱分析阶段。

目前谱分析方法主要应用于电力工程、信息工程、物理学、天文学、海洋学和气象科学等领域，它是一种非常有用的纵向数据分析方法。但是由于谱分析过程一般都比较复杂，研究人员通常要具有很强的数学基础才能熟练使用它，同时它的分析结果也比较抽象，不易于进行直观解释，导致谱分析方法的使用具有很大的局限性。

2. 时域分析方法

时域（*time domain*）分析方法主要是从序列自相关的角度揭示时间序列的发展规律。相对于谱分析方法，它具有理论知识扎实、操作步骤规范、分析结果易于解释的优点。目前它已广泛应用于自然科学和社会科学的各个领域，

成为时间序列分析的主流方法。本文所使用的方法就是时域分析方法。

时域分析方法的基本思想是事件的发展通常都具有一定的惯性，这种惯性用统计的语言来描述就是序列值之间存在一定的相关关系，而且这种相关关系具有某种统计规律性。我们分析的重点就是寻找这种规律，并拟合出适当的数学模型来描述这种规律，进而利用这个拟合的模型来预测序列未来的走势。

时域分析方法具有相对固定的分析套路，通常都遵循如下分析步骤：

- (1) 考察观察值序列的特征
- (2) 根据序列的特征选择适当的拟合模型
- (3) 根据序列的观察数据确定模型的口径
- (4) 检验模型，优化模型
- (5) 利用拟合好的模型来推断序列其他统计性质或预测序列将来的发展

时域分析方法最早可以追溯到 1927 年，英国统计学家 *GU.Yule* 提出自回归模型。不久之后，英国数学家、天文学家 *G.T.Walker* 在分析印度大气规律时使用了移动平均（*Moving Average*）模型和自回归移动平均模型（*Auto Regression Moving Average*）模型（1931 年）。这些模型奠定了时间序列时域分析方法的基础。

1970 年，美国统计学家 *G.E.P.Box* 和英国统计学家 *G.M.Jenkins* 联合出版了 *Time Series Analysis Forecasting and Control* 一书。在书中，*Box* 和 *Jenkins* 在总结前人研究的基础上系统地阐述了对求和自回归移动平均（*Autoregressive Integrated Moving Average*，*ARIMA*）模型的识别、估计、检验及预测的原理及方法。这些知识现在被称为经典时间序列分析方法，是时域分析方法的核心内容。为了纪念 *Box* 和 *Jenkins* 对时间序列发展的特殊贡献，现在人们也常把 *ARIMA* 模型称为 *Box-Jenkins* 模型。

Box-Jenkins 模型实际上主要应用于单变量、同方差场合的线性模型。随着人们对各种领域时间序列分析的深入研究，发现该经典模型在理论上和应用上都还存在着许多局限性。所以近 20 年来，统计学家纷纷转向多变量场合、异方差场合和非线性场合的时间序列分析方法的研究，并取得了突破性的进展。

在异方差场合,美国统计学家、计量经济学家 *Robert F. Engle* 在 1982 年提出了自回归条件异方差 (*ARCH*) 模型,用以研究英国通货膨胀的建模问题。为了进一步放宽 *ARCH* 模型的约束条件, *Bollerslev* 在 1985 年提出了广义自回归条件异方差 (*GARCH*) 模型。随后 *Nelson* 等人又提出了指数广义自回归条件异方差模型 (*EGARCH*)、方差无穷广义自回归条件异方差模型 (*IGARCH*) 和依均值广义自回归条件异方差模型 (*GARCH-M*) 等限制条件更为宽松的异方差模型。这些异方差模型是对经典的 *ARIMA* 模型的很好的补充。它们比传统的方差齐性模型更能准确地刻画了金融市场风险的变化过程,因此 *ARCH* 模型及其衍生出的一系列拓展模型在计量经济学领域有着广泛的应用。*Engle* 也因此获得了 2003 年的诺贝尔经济学奖。

在多变量场合, *Box* 和 *Jenkins* 在 *Time Series Analysis Forecasting and Control* 一书中研究过平稳多变量序列的建模, *Box* 和 *Tiao* 在 1970 年前后讨论过带干扰变量的时间序列分析。这些研究实际上是把对随机事件的横向研究和纵向研究有机地结合在了一起,提高了对随机事件分析和预测的精度。1987 年,英国统计学家、计量经济学家 *C. Granger* 提出了协整 (*Cointegration*) 理论,进一步为多变量时间序列建模松绑。有了协整的概念之后,在多变量时间序列建模过程中“变量是平稳的”不再是必要条件了,只要它们的某种线性组合平稳就可以了。协整概念的提出极大地促进了多变量时间序列分析方法的发展, *Granger* 因此与 *Engle* 一起获得了 2003 年诺贝尔经济学奖。

非线性时间序列分析也有重大发展,汤家豪教授等在 1980 年前后提出了利用分段线性化构造的门限自回归也成为目前分析时间序列的经典模型。

第二节 时间序列数据预处理

拿到一个观察值序列之后,首先要对它的“平稳性”和“纯随机性”进行检验,这两个重要的检验称为序列的预处理。

一、平稳性

根据限制条件的严格程度，可分为严平稳和宽平稳。

严平稳是一种条件比较苛刻的平稳性定义，它认为只有当序列所有的统计性质都不会随着时间的推移而发生变化时，该序列才能被认为平稳。因为随机变量族的统计性质完全由它们的联合概率分布族决定。所以严平稳的定义如下：设 $\{X_t\}$ 为一时间序列，对任意正整数 m ，任取 $t_1, t_2, \dots, t_m \in T$ ，对任意整数 τ ，有

$$F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m) = F_{t_1 + \tau, t_2 + \tau, \dots, t_m + \tau}(x_1, x_2, \dots, x_m)$$

则称时间序列 $\{X_t\}$ 严平稳。

宽平稳是使用时间序列的特征统计量来定义的一种平稳性。认为序列的统计性质主要由序列的低阶矩决定，所以只要保证序列的低阶矩平稳，就能保证序列的主要性质近似平稳。

如果 $\{X_t\}$ 满足如下三个条件：

- (1) 任取 $t \in T$ ，有 $EX_t^2 < \infty$ ；
- (2) 任取 $t \in T$ ，有 $EX_t = \mu$ ， μ 为常数；
- (3) 任取 $t, s, k \in T$ ，且 $k + s - t \in T$ ，有 $\gamma(t, s) = \gamma(k, k + s - t)$ ；

则称 $\{X_t\}$ 为宽平稳时间序列。

对序列平稳性的检验方法主要有两种检验方法，一种是根据时序图和自相关图显示的特征做出的图检验方法；另一种是构造检验统计量进行假设检验的方法。图检验方法是一种操作简便、运用广泛的平稳性判别方法，在实践中被广泛采用。

(一) 时序图检验

根据平稳时间序列的均值、方差为常数的性质，平稳序列的时序图应该显示出该序列始终在一个常数均值附近波动，而且波动范围有限的特点。如果观察序列的时序图显示出该序列有明显的趋势性或周期性，那它通常不是平稳序列。

(二) 自相关图检验

平稳序列通常具有短期相关性。该性质用自相关系数来描述就是随着延迟

期数 k 的增加, 平稳序列的自相关系数 ρ_k 会很快衰减向零。反之, 非平稳序列的自相关系数 ρ_k 衰减向零的速度通常比较慢, 这就是我们利用自相关图进行平稳性判断的标准。

二、纯随机性

如果序列经过平稳性判断之后被认为是平稳的, 那么我们有一套比较成熟的平稳序列分析方法对其建模, 然而并不是所有的平稳序列都值得建模。只有那些序列值之间具有密切相关关系, 历史数据对未来发展有一定影响的序列, 才值得我们花时间去挖掘历史数据中的有效信息, 用来预测序列未来的发展。为了确定平稳时间序列是否值得分析下去, 我们需要对平稳序列进行纯随机性检验。

Barlett 证明, 如果一个时间序列是纯随机的, 得到一个观察期数为 n 的观察值序列 $\{x_t, t=1, 2, \dots, n\}$, 那么该序列的延迟非零期的样本自相关系数将近似服从均值为零、方差为序列观察期数倒数的正态分布, 即

$$\hat{\rho}_k \sim N(0, \frac{1}{n}), \forall k \neq 0$$

式中, n 为观察期数。

根据 *Barlett* 定理, 我们可以构造检验统计量来检验序列的纯随机性。

原假设和备择假设分别为:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 1$$

$$H_1: \text{至少存在某个 } \rho_k \neq 0, \forall m \geq 1, k \leq m$$

检验统计量为 *LB(Ljung - Box)* 检验统计量:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2(m), \forall m > 0$$

当 *LB* 统计量大于 $\chi^2_{1-\alpha}(m)$ 分位点, 或该统计量的 p 值小于 α 时, 则可以以 $1-\alpha$ 的置信水平拒绝原假设, 认为该序列为非白噪声序列; 否则, 接受原假设, 认为该序列为纯随机序列。

一个序列经过预处理识别为“平稳非白噪声”序列, 那就说明该序列是一个

个蕴含着相关信息的平稳序列。在统计上，我们通常是建立一个线性模型来拟合该序列的发展，借此提供该序列中的有用信息。

第三节 ARMA 模型

ARMA 模型的全称是自回归移动平均（Auto Regression Moving Average）模型，它是目前最常用的拟合平稳序列的模型。

一、定义

把具有如下结构的模型称为自回归移动平均模型，简记为 $ARMA(p, q)$ ：

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

若 $\phi_0 = 0$ ，该模型为中心化 $ARMA(p, q)$ 模型。

中心化 $ARMA(p, q)$ 模型可以简写为：

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

引进延迟算子， $ARMA(p, q)$ 模型简记为：

$$\Phi(B)x_t = \Theta(B)\varepsilon_t$$

式中

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \text{ 为 } p \text{ 阶自回归系数多项式}$$

$$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q, \text{ 为 } q \text{ 阶移动平均系数多项式}$$

显然，

当 $p = 0$ 时， $ARMA(p, q)$ 模型就退化成了 $MA(q)$ 模型

当 $q = 0$ 时， $ARMA(p, q)$ 模型就退化成了 $AR(p)$ 模型

所以， $AR(p)$ 模型和 $MA(q)$ 模型实际上就是 $ARMA(p, q)$ 模型的特例，他们

都统称为 $ARMA$ 模型。而 $ARMA(p, q)$ 模型的统计性质也正是 $AR(p)$ 模型和 $MA(q)$ 模型统计性质的有机结合。

二、 $ARMA(p, q)$ 模型的统计性质

(一) 均值

对于一个非中心化平稳可逆的 $ARMA(p, q)$ 模型

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

两边同求均值，有

$$Ex_t = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

(二) 自协方差函数

$$\begin{aligned} \gamma(k) &= E(x_t x_{t+k}) \\ &= E\left[\left(\sum_{i=0}^{\infty} G_i \varepsilon_{t-i}\right)\left(\sum_{j=0}^{\infty} G_j \varepsilon_{t+k-j}\right)\right] \\ &= E\left[\sum_{i=0}^{\infty} G_i \sum_{j=0}^{\infty} G_j \varepsilon_{t-i} \varepsilon_{t+k-j}\right] \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} G_i G_{i+k} \end{aligned}$$

(三) 自相关系数

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\sum_{j=0}^{\infty} G_j G_{j+k}}{\sum_{j=0}^{\infty} G_j^2}$$

根据自相关系数的表达式很容易判断 $ARMA(p, q)$ 模型的自相关系数不截尾。这和 $ARMA(p, q)$ 模型可以转化为无穷阶移动平均模型的性质是一致的。同时，根据 $ARMA(p, q)$ 模型可以转化为无穷阶自回归模型，可以判断它的偏相关系数也不截尾。

综合考察 $AR(p)$ 模型， $MA(q)$ 模型和 $ARMA(p, q)$ 模型的自相关系数和偏自

相关系数的性质，我们可以总结出如下规律： $AR(p)$ 模型的自相关系数是拖尾的，而偏自相关系数是 p 步截尾的。 $MA(q)$ 模型的自相关系数是 q 步截尾的，而偏自相关系数具有拖尾性。 $ARMA(p,q)$ 模型的自相关系数和偏自相关系数都是拖尾的。

（四）建立模型的步骤

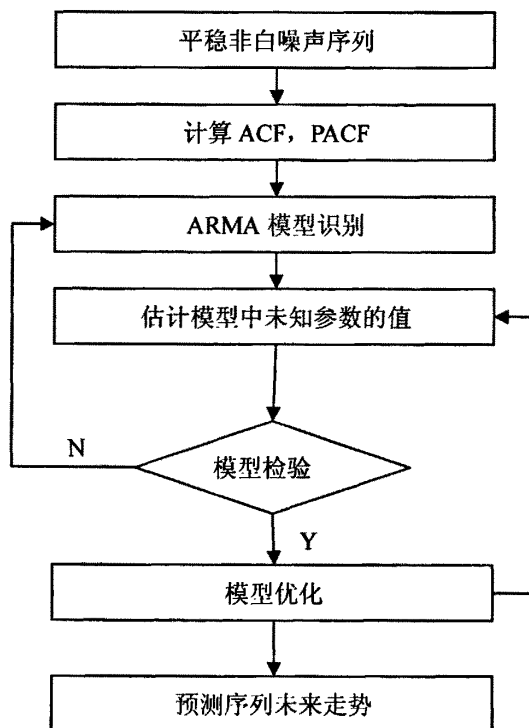


图 5.1 ARMA 模型建模流程图

（五）模型以及参数的检验

1. 模型的显著性检验主要是检验模型的有效性

一个模型是否显著有效主要看它提取的信息是否充分。一个好的拟合模型应该能够提取观察值序列中几乎所有的样本相关信息，换言之，拟合残差中将

不再蕴含任何相关信息,即残差序列应该为白噪声序列。这样的模型称为显著有效模型。反之,如果残差序列为非白噪声序列,那就意味着残差序列中还留有相关性信息,说明拟合模型不够有效,通常需要选择其他模型,重新拟合。

2. 参数的显著性检验

参数的显著性检验就是要检验每一个未知参数是否显著为非零。这个检验的目的是为了使模型最精简。

如果某个参数不显著,即表示该参数所对应的那个自变量对因变量的影响不明显,该变量就可以从拟合模型中删除。最终模型将由一系列参数显著性非零的自变量表示。

检验假设

$$H_0: \beta_j = 0 \leftrightarrow H_1: \beta_j \neq 0, \forall 1 \leq j \leq m$$

用于检验未知参数显著性的 t 检验统计量

$$T = \sqrt{n-m} \frac{\hat{\beta}_j - \beta_j}{\sqrt{a_{jj} Q(\hat{\beta})}} \sim t(n-m)$$

当该检验统计量的绝对值大于自由度为 $n-m$ 的 t 分布的 $1-\frac{\alpha}{2}$ 分位点,即

$$|T| \geq t_{1-\alpha/2}(n-m)$$

或者该检验统计量的 P 值小于 $\frac{\alpha}{2}$ 或大于 $1-\frac{\alpha}{2}$ 时,拒绝原假设,认为该参数显著。否则,认为该参数不显著。这时,应该删除不显著参数所对应的自变量重新拟合模型,构造出新的、结构更精炼的拟合模型。

第四节 ARIMA 模型

ARMA 模型主要是针对平稳时间序列的分析方法。实际上,在现实中绝大部分序列都是非平稳的,因而对非平稳序列的分析更普遍、更重要。对非平稳时间序列的分析方法可以分为确定性时序分析和随机性时序分析两大类。

非平稳序列的确定性因素分解方法具有具体原理简单、操作简单、易于理解等优点,在宏观经济管理与预测领域有着广泛的应用。但是随着研究方法的

深入

和研究领域的拓宽，人们发现确定性因素分解方法还存在一些问题，主要有：一是确定性因素分解方法只能提取强劲的确定性信息，对随机性信息浪费严重。二是确定性因素分解方法把所有序列的变化都归结为四大因素的综合影响，始终无法提供明确、有效的方法判断各大因素之间确切的作用关系。

这些问题导致确定性因素分解方法不能充分提取观察值序列中的有效信息，导致模型拟合精度不够理想。随机时序分析方法的发展就是为了弥补确定性因素分解方法的不足，为人们提供了更加丰富的、更加准确的时序分析工具。

一、确定性信息的提取

拿到观察值之后，无论是采用确定性时序分析方法还是随机时序分析方法，分析的第一步都是要通过有效的手段提取序列中所蕴含的确定性信息。确定性信息的提取方法有很多，比如构造季节指数模型、拟合长期趋势模型、移动平均、指数平滑等诸多方法都是确定性信息的提取方法。但是它们对确定性信息的提取都不够充分。Box 和 Jenkins 在 *Time Series Analysis Forecasting and Control* 一书中特别强调差分方法的使用，他们使用大量的案例分析证明差分方法是一种非常简便、有效的确定性信息提取方法。而 Cramer 分解定理则在理论上保证了适当阶数的差分一定可以充分提取确定性信息。根据 Cramer 分解定理，方差齐性非平稳序列都可以分解为如下形式：

$$x_t = \sum_{j=0}^d \beta_j \cdot t^j + \Psi(B)a_t$$

式中， $\{a_t\}$ 为白噪声序列。

离散序列的 d 阶差分就相当于连续变量的 d 阶求导，显然，在 Cramer 分解定理的保证下， d 阶差分就可以将 $\{x_t\}$ 中蕴含的确定性信息充分提取。

实践中，我们会根据序列不同的特点选择合适的差分方法，常见情况有以下三种。

1. 序列蕴含着显著的线性趋势，1 阶差分就可以实现趋势平稳

2. 序列蕴含着曲线趋势，通常低阶（2 阶或 3 阶）差分就可以提取出曲线趋势的影响

3. 序列蕴含着固定周期，通常进行步长为周期长度的差分运算，通常可以较好地提取周期信息。

差分运算有着强大的确定性信息提取能力，许多非平稳序列差分后会显示出平稳序列的性质，这时我们称这个非平稳序列为差分平稳序列。对差分平稳序列可以使用 *ARIMA* 模型进行拟合。

二、ARIMA 模型定义

具有如下结构的模型称为求和自回归移动平均模型 (*Autoregressive Integrated Moving Average*)，简记为 *ARIMA*(p, d, q) 模型：

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, \forall s < t \end{cases} \quad (1)$$

式中，

$$\nabla^d = (1 - B)^d;$$

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

为平稳可逆 *ARMA*(p, q) 模型的自回归系数多项式；

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

为平稳可逆 *ARMA*(p, q) 模型的移动平滑系数多项式。

式 (1) 可以简记为：

$$\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \quad (2)$$

式中， $\{\varepsilon_t\}$ 为零均值白噪声序列。

由式 (2) 可以看出，*ARIMA* 模型的实质就是差分运算与 *ARMA* 模型的组合。这说明任何非平稳序列只要通过适当阶数的差分实现差分平稳，就可以对差分后序列进行 *ARMA* 模型拟合了。而 *ARMA* 模型的分析方法非常成熟，这意味着对差分平稳序列的分析也将是简单可靠的。特别地，

当 $d=0$ 时, $ARIMA(p,d,q)$ 模型实际上就是 $ARMA(p,q)$ 模型

当 $p=0$ 时, $ARIMA(p,d,q)$ 模型可以简记为 $IMA(d,q)$ 模型

当 $q=0$ 时, $ARIMA(p,d,q)$ 模型可以简记为 $ARI(p,d)$ 模型

三、建模步骤

在掌握了 $ARIMA$ 模型建模方法之后, 尝试使用 $ARIMA$ 模型对观察值序列建模是一件比较简单的事情。它遵循如下流程:

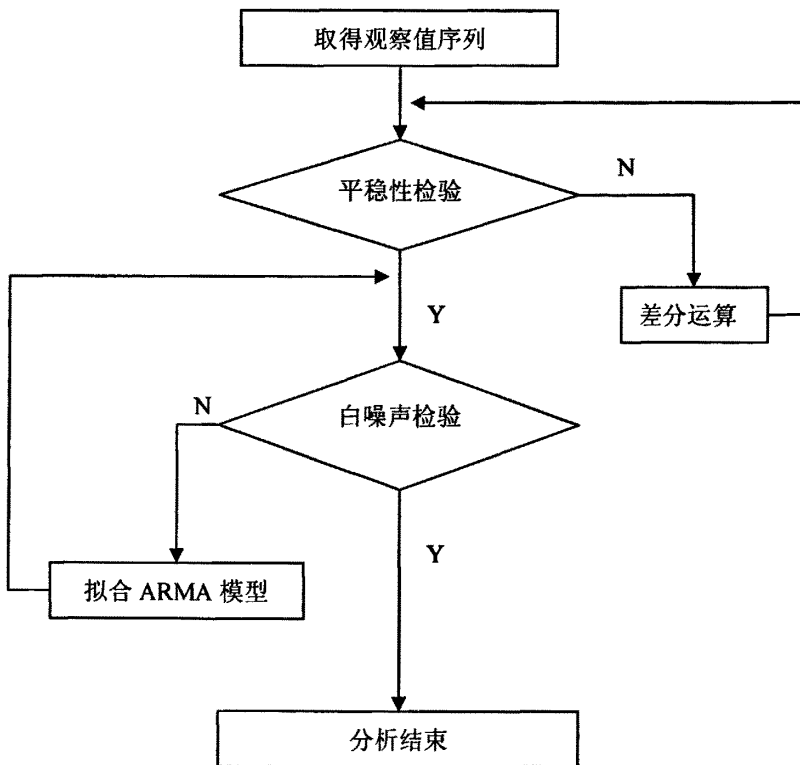


图 5.2 $ARIMA$ 模型建模流程图

第五节 $ARCH$ 与 $GARCH$ 模型

一、 $ARCH$ 模型的提出

1982 年 Engle 在分析英国通货膨胀率序列时, 发现经典的 $ARIMA$ 模型始终

无法取得理想的拟合效果。经过对残差序列的仔细分析研究，他发现问题出在残差序列具有异方差性。

异方差即随机误差项的方差不再是常数了，它会随着时间的变化而变化，它可以表现为时间的某个函数： $Var(\varepsilon_t) = h(t)$ 。这种情况被称为异方差。忽视异方差的存在会导致残差的方差被严重低估，继而参数显著性检验容易犯纳伪错误，这使得参数的显著性检验失去意义，最终导致模型的拟合精度受到影响。当残差序列存在异方差时，我们需要对它做进一步的处理，处理思路有两种：一是假如已知异方差函数具体形式，进行方差齐性变化；二是假如不知道异方差函数的具体形式，拟合条件异方差模型

方差齐性变换为异方差序列的拟合提供了一个很好的方法，但是这个方法更多的是具有理论上的意义，在实践中的可操作性并不强。因为要使用方差齐性变换必须得事先知道异方差函数的形式，而这通常是不可能的。

实践中，我们只能根据残差图及残差平方图所显示出来的特点，使用一些常用的函数形式来估计异方差模型。这体现在进行金融时间序列分析时，由于金融时间序列的标准差和水平之间通常具有某种正相关关系，所以异方差函数经常被假设为： $h(\mu_t) = \mu_t^2$ 。继而导致对数变换在进行金融时间序列分析时被普遍采用。但大量的实践证明这是一种过于简单化的方法，对数变换通常只能使绝大多数金融时间序列的异方差程度得到改善，但无法真正地实现方差齐性。

为了更精确地估计异方差函数，Engle 于 1982 年提出了条件异方差模型

(*Autoregressive Conditional Heteroskedastic*)，简记为 $ARCH(q)$ 。

它的完整结构为：

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 \\ e_t \stackrel{i.i.d}{\sim} N(0,1) \end{cases}$$

$ARCH$ 模型的实质是使用误差平方序列的 q 阶移动平均拟合当期异方差函

数值。由于移动平均模型具有自相关系数 q 阶截尾性，所以 $ARCH$ 模型实际上只适用于异方差函数短期自相关过程。

二、 $GARCH$ 模型的提出

在实践中，有些残差序列的异方差是具有长期自相关性的，这时如果用 $ARCH$ 模型拟合异方差函数，将会产生很高的移动平均阶数，这会增加参数估计的难度并最终影响 $ARCH$ 模型的拟合精度。

为了修正这个问题，*Bollerslev*在1986年提出了广义自回归条件异方差（*Generalized Autoregressive Conditional Heteroskedastic*）模型，简记为 $GARCH(p,q)$ ，它的结构如下：

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 + \sum_{i=1}^p \eta_i h_{t-i} \\ e_t \stackrel{i.i.d}{\sim} N(0,1) \end{cases}$$

$GARCH$ 模型实际上就是在 $ARCH$ 模型的基础上，增加考虑了异方差函数的 p 阶自相关性。它可以有效地拟合具有长期记忆性的异方差函数。显然 $ARCH$ 模型是 $GARCH$ 模型的一种特例， $ARCH(q)$ 模型就是 $p=0$ 的模型 $GARCH(p,q)$ 。

$GARCH$ 模型是至今为止最常用的、最可行的异方差拟合模型。但它有两个约束条件

（一）参数非负 $\omega > 0, \eta_i \geq 0, \lambda_j \geq 0$

（二）参数有界 $\sum_{i=1}^p \eta_i + \sum_{j=1}^q \lambda_j < 1$

这两个约束条件限制了 $GARCH$ 模型的使用范围。为了拓宽 $GARCH$ 模型的使用，许多统计学家从不同角度出發，构造了多个 $GARCH$ 模型的变体。主要有：指数 $GARCH$ 模型，简记为 $EGARCH$ ；方差无穷 $GARCH$ 模型，简记为 $IGARCH$ ；依均值 $GARCH$ 模型，简记为 $GARCH-M$ 。

对残差序列 $\{\varepsilon_t\}$ 拟合 $GARCH$ 模型有一个基本要求: $\{\varepsilon_t\}$ 为零均值、纯随机异方差序列。如果回归函数 $f(t, x_{t-1}, x_{t-2}, \dots)$ 不能充分提取原序列 $\{x_t\}$ 中的相关信息, 则残差序列可能具有自相关性, 而不是纯随机性的。这时需要对 $\{\varepsilon_t\}$ 先拟合自回归模型, 再考察自回归残差序列 $\{v_t\}$ 的方差齐性, 如果 $\{v_t\}$ 异方差, 对它拟合 $GARCH$ 模型。这样构造的模型称为 $AR(m)-GARCH(p, q)$ 模型, 它的模型结构是:

$$\left\{ \begin{array}{l} x_t = f(t, x_{t-1}, x_{t-2}, \dots) + \varepsilon_t \\ \varepsilon_t = \sum_{k=1}^m \beta_k \varepsilon_{t-k} + v_t \\ v_t = \sqrt{h_t} e_t \\ h_t = \omega + \sum_{j=1}^q \lambda_j \varepsilon_{t-j}^2 + \sum_{i=1}^p \eta_i h_{t-i} \\ e_t \stackrel{i.i.d}{\sim} N(0, 1) \end{array} \right.$$

第六章 上证指数波动趋势短期预测

本章所研究的对象是上证综合指数。该指数较好地代表了市场整体的价格水平，能准确及时地反映我国股市的行情动态，具有较高的预测价值和较好的可预测性。

分析的数据是上证指数自 2008 年 10 月 29 日到 2009 年 4 月 22 日的 7 日移动均线数数据，研究的目的是希望能够对其运行趋势在短期内做出一个判断。我们希望预测的是一种短期趋势而不是精确的未来每一天的股票收盘价，所以首先对采用简单移动平均的方法对收盘价数据进行了一定程度的降噪处理。

数据来源于 CSMAR 经济、金融、证券研究数据库。使用的分析软件是 SAS / ETS 软件。

第一节 ARIMA 建模分析

差分运算有很好的确定性信息提取能力，许多非平稳序列差分后会显示出平稳序列的性质，这时我们称这个非平稳序列为差分平稳序列。对差分平稳序列一般使用 ARIMA 模型进行拟合。

在我们所选择研究的样本区间，上证指数表现出明显的上升趋势，是典型的非平稳序列，不能直接使用 ARMA 模型进行拟合，所以考虑使用 ARIMA 模型。

下面我们按照 ARIMA 建模流程展开研究。

一、获得数据

数据来源于 CSMAR 经济、金融、证券研究数据库。研究对象是上证指数从 2008 年 10 月 29 日到 2009 年 4 月 21 日以来的 7 日移动均线序列数据。我们已经建立了数据集 *resdat.ma7*，其中变量 *date* 是日期，变量 *ma7* 是上证指数的 7 日移动平均线序列。

二、判断序列的平稳性

首先我们通过对其做时序图来直观地观察序列的平稳性。

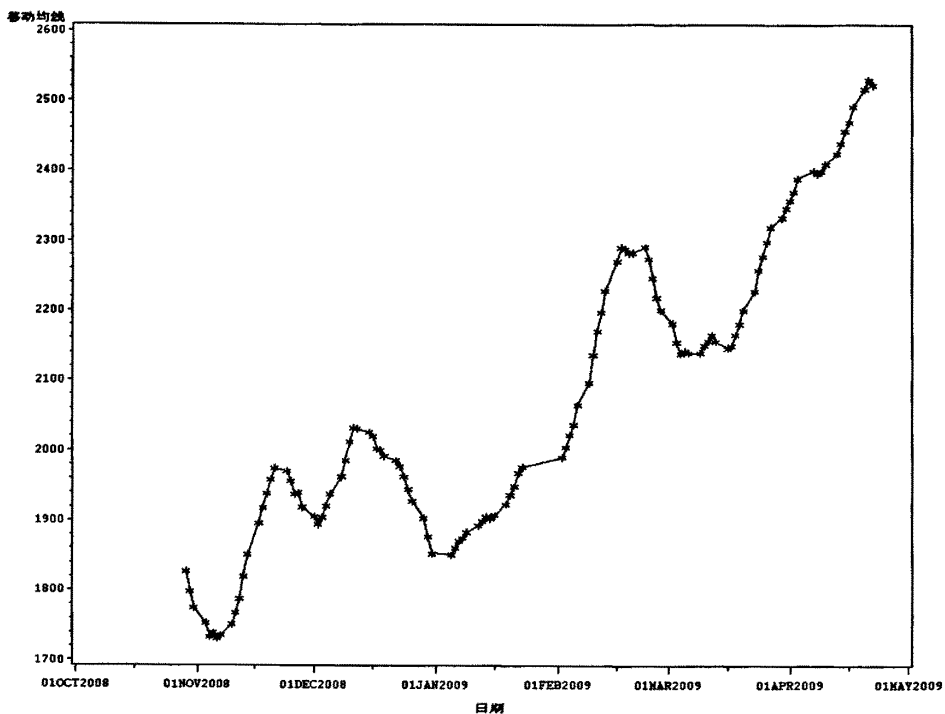


图 6.1: 上证指数 7 日移动均线时序图

由图 6.1 可以看出,有着非常明显的上升趋势,显然存在 $E(X_t) \neq \mu$, 这是典型的非平稳时间序列。为了消除其趋势因素,我们可以采用差分的方法。

三、对序列进行差分平稳化处理

因为原序列有近似的线性趋势,为消除线性趋势,所以选择进行 1 阶差分运算, $dma7 = (1 - B)ma7$, B 是滞后算子。

一阶差分序列 $dma7$ 时序图如下:

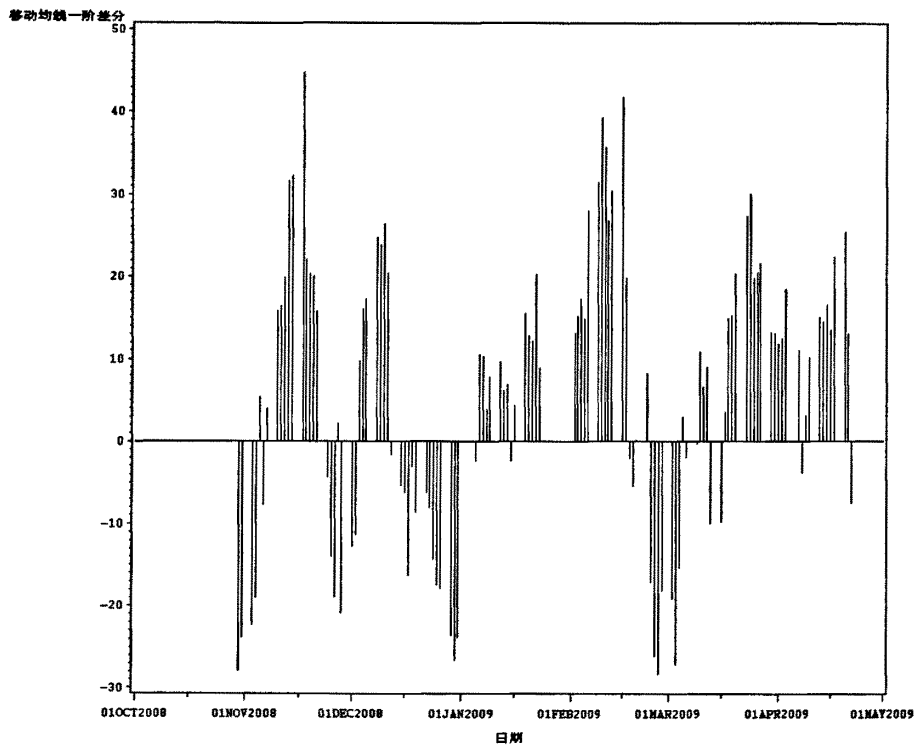


图 6.2: 移动均线一阶差分时序图

差分序列图显示出其在均值零附近比较稳定地波动，初步可以判断其为平稳序列，但前后有比较明显的相关关系，还有值得提取的信息。为了进一步地确定其平稳性，考察差分序列的自相关图。

所得差分序列的自相关图，如下：

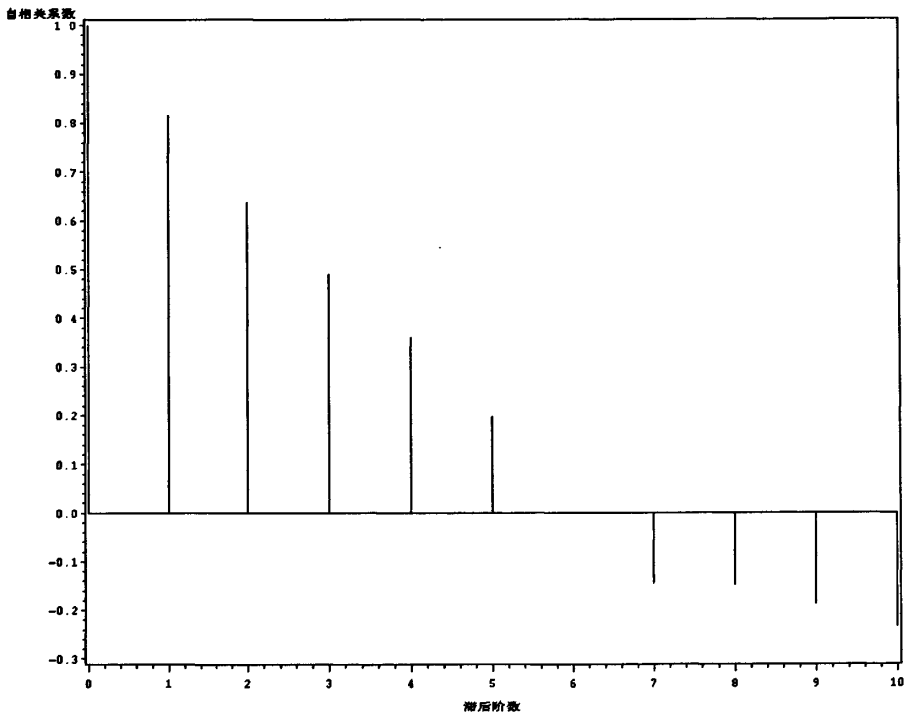


图 6.3: 一阶差分序列自相关系数图

可以看出自相关系数很快衰减到零, 这是平稳序列的典型特征, 所以可以认为一阶差分后的序列具备了平稳性。

四、对平稳的一阶差分序列进行白噪声检验

在检验的显著性水平为 0.05 的条件下, 由于延迟 6 阶的 χ^2 检验的统计量为 182.2, 其 P 值小于 0.0001, 远远小于 0.05, 所以该差分序列不能视为白噪声序列, 即差分后还蕴含着不容忽视的相关性信息可供提取。

五、对平稳非白噪声差分序列拟合 $ARMA$ 模型

一阶差分后序列的自相关图已经显示出该序列自相关系数有一定的拖尾性。再考察其偏自相关系数的性质。如下:

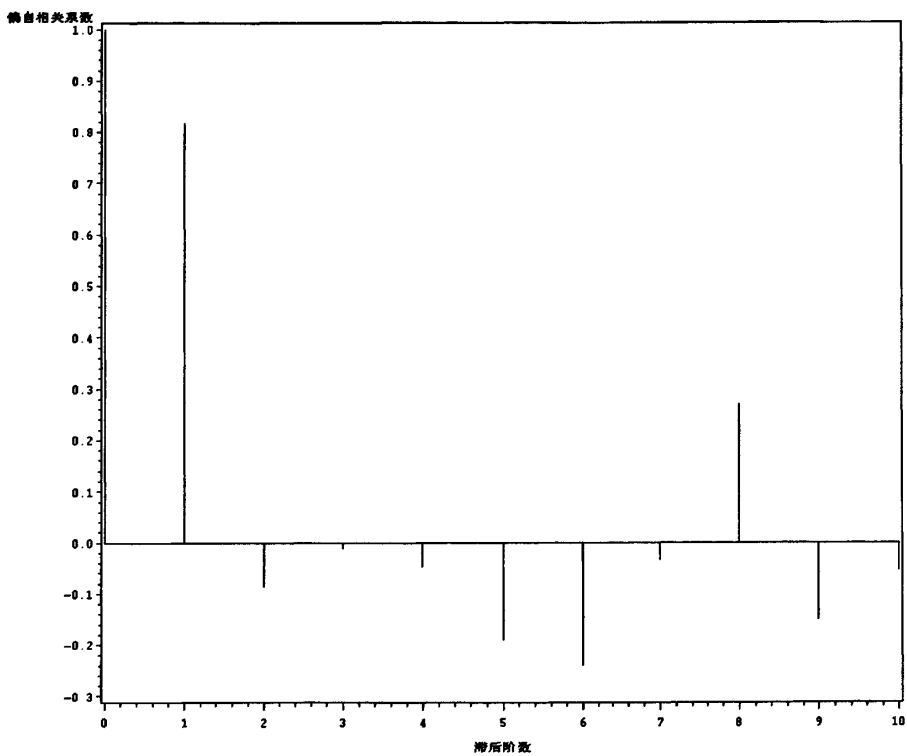


图 6.4: 一阶差分序列偏自相关图

可以看出偏自相关系数也具有一定的拖尾性，所以考虑用 $ARMA(1,1)$ 模型拟合一阶差分后的序列。

由 SAS / ETS 软件计算结果可得最终的估计模型是：

$$(1-B)ma7_t = \frac{(1+0.11283B)}{(1-0.81123B)} \varepsilon_t$$

在模型的结构和参数确定之后，就可以利用这个模型对上证指数 7 日均线做出短期预测了。

六、拟合与预测

我们的目标是进行短期预测，所以预测步长设定为 10，即预测未来 10 日内 7 日移动均线的走势。

所得到的拟合与预测效果图如下：

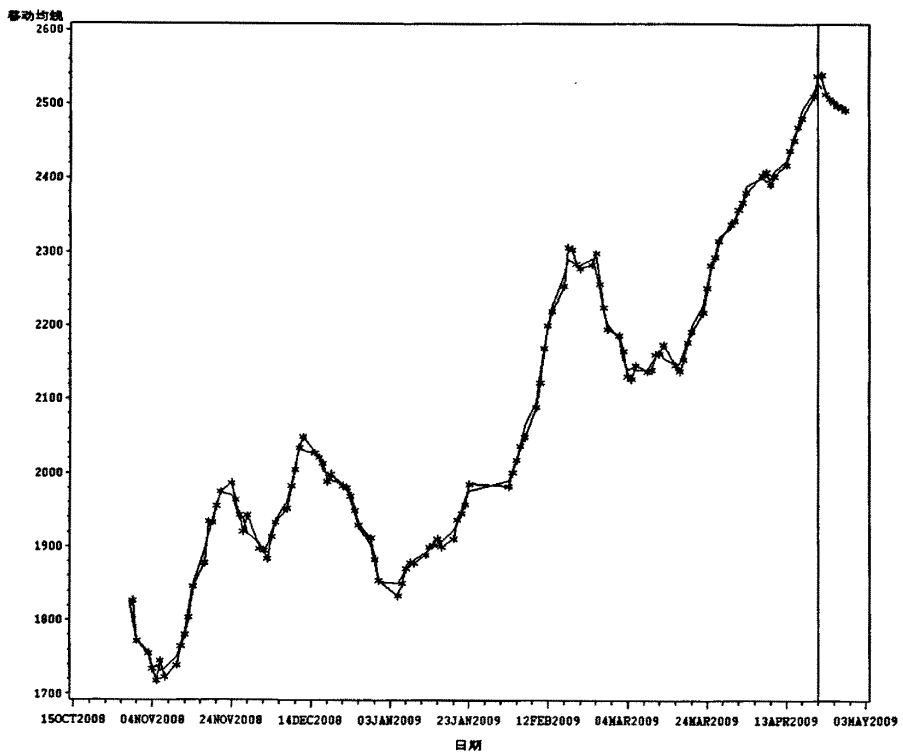


图6.5: 拟合与预测效果图

以上主要是应用 *ARIMA* 模型来进行建模分析, 由最终的拟合与预测效果图来看, 拟合的效果是相当好的。而且可以看出, 未来 7 日均线有下降的趋势, 这就提示我们注意风险。下面我们使用 *GARCH* 模型来进行建模。

第二节 $AR(m)-GARCH(p,q)$ 建模分析

由于股票价格的波动性往往呈现出异方差性和波动聚集性，而且异方差的函数形式通常是未知的， $GARCH$ 模型族提供了一种估计和校正异方差的方法，它假定误差项尽管不相关但是不独立，并且把条件异方差做为序列过去值的函数，采用 $GARCH$ 模型族对指数短期趋势序列进行建模，可以有效地处理异方差性，从而取得比较好的模拟预测效果。

具体建模分析步骤如下：

一、直观查看其时序图并以时间 t 为自变量对其做线性回归图
所得图形如下：

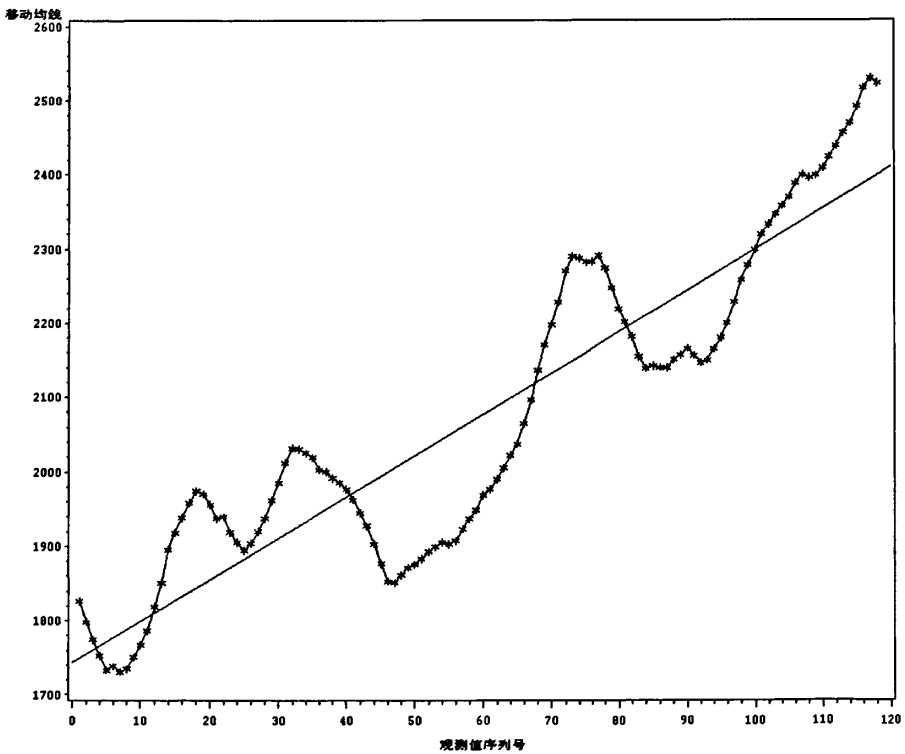


图 6.6：上证指数 7 日移动均线走势与线性回归直线图

可以看出时序图显示序列有明显的上升趋势，并且序列倾向于保持一段时

间内在趋势直线的上方或者下方，这是典型自相关的特征。下面我们用 *Durbin-Watson* 检验来对残差序列进行自相关性检验，结果如下：

表 6.1 DW 检验统计量

阶数	DW 值	P 值
1	0.0386	<0.001
2	0.1389	<0.001
3	0.2854	<0.001
4	0.4647	<0.001
5	0.6654	<0.001

可以看出各阶 *Durbin-Waston* 检验极其显著，其值<0.001，说明残差序列存在明显的自相关性，因此自相关性的校正是必要的。

二、对残差序列拟合合适的 AR 模型

确定需要做自相关性校正了，那么就需要选用合适的误差自回归模型的阶了。常用的方法是逐步自回归。逐步自回归方法最初拟合一个长阶自回归模型，然后逐步去掉不显著的自回归参数，直到所有剩下的自回归参数都有显著的 *t* 检验为止。我们选择初始阶长为 5。

自回归项向后消除的报告显示在时间间隔3、4的自回归参数不显著且被消除，如下图所示：

表 6.2 向后消除自回归项

滞后阶数	参数估计值	T 值	P> T
3	0.069017	0.54	0.5906
4	-0.029065	-0.18	0.8551

逐步自回归方法丢弃了时间间隔 3，4，保留了 1，2，5。这样的模型称为子集模型。其结构如下：

$$\begin{cases} y_t = a + bt + \varepsilon_t \\ \varepsilon_t = -\varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \varphi_3 v_{t-3} + v_t \end{cases}$$

由此产生了新的残差序列 $\{v_t\}$

三、检验异方差性

利用 AR 模型提取了残差 $\{\varepsilon_t\}$ 中的自相关信息，新的残差序列 $\{v_t\}$ 实现了零均值、纯随机的性质，现在需要考察它的异方差性，需要进行异方差性检验。

最常用的两个异方差检验是：*Portmantea Q* 检验和 *LM* 检验。

(一) *Portmantea Q* 检验

1、假设条件

H_0 : 残差平方序列纯随机

H_1 : 残差平方序列具有自相关性

用自相关系数表示，该假设条件等价于

$$H_0: \rho_1 = \rho_2 = \cdots \rho_q = 0$$

$$H_1: \rho_1, \rho_2, \cdots \rho_q \text{ 不全为零}$$

2、检验统计量

$$Q(q) = n(n+2) \sum_{i=1}^q \frac{\rho_i^2}{n-i}$$

式中， n 为序列观测长度； ρ_i 为残差序列延迟 i 期自相关系数，记 $\hat{\sigma}^2 = \frac{\sum_{i=1}^n \varepsilon_i^2}{n}$ 。

$$\rho_i = \frac{\sum_{t=i+1}^n (\varepsilon_t^2 - \hat{\sigma}^2)(\varepsilon_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^n (\varepsilon_t^2 - \hat{\sigma}^2)^2}$$

Portmantea Q 统计量近似服从 $\chi^2(q-1)$ ，在显著性水平为 α 的情况下，如果 $Q(q) \geq \chi_{1-\alpha}^2(q-1)$ ，即 P 值小于 α ，就拒绝原假设，反之则接受原假设。

(二) *LM* 检验

1、假设条件

H_0 : 残差平方序列纯随机

H_1 : 残差平方序列具有自相关性

用自相关系数表示, 该假设条件等价于

H_0 : $\rho_1 = \rho_2 = \cdots \rho_q = 0$

H_1 : $\rho_1, \rho_2, \cdots \rho_q$ 不全为零

2、检验统计量

$$LM(q) = W'W$$

式中,

$$W = \left(\frac{\rho_1^2}{\hat{\sigma}^2}, \frac{\rho_2^2}{\hat{\sigma}^2}, \cdots, \frac{\rho_q^2}{\hat{\sigma}^2} \right)$$

$$\rho_i = \frac{\sum_{t=i+1}^n (\varepsilon_t^2 - \hat{\sigma}^2)(\varepsilon_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^n (\varepsilon_t^2 - \hat{\sigma}^2)^2}$$

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n \varepsilon_t^2}{n}$$

$LM(q)$ 统计量近似服从 $\chi^2(q-1)$, 在显著性水平为 α 的情况下, 如果

$LM(q) \geq \chi_{1-\alpha}^2(q-1)$, 即 P 值小于 α , 就拒绝原假设, 反之则接受原假设。

对残差序列 $\{\varepsilon_t\}$ 进行异方差检验, 检验结果如下:

表 6.3 异方差性检验的 LM 统计量和 Q 统计量

阶数	Q 统计量	P>Q	LM 统计量	P>LM
1	105.1033	<0.001	103.0265	<0.001
2	178.4640	<0.001	109.4789	<0.001
3	220.8105	<0.001	109.5028	<0.001

4	240.4171	<0.001	109.5446	<0.001
5	246.5654	<0.001	109.5795	<0.001
6	247.2904	<0.001	109.7254	<0.001
7	247.3819	<0.001	109.7292	<0.001
8	248.3372	<0.001	109.7416	<0.001
9	249.9529	<0.001	109.7419	<0.001
10	251.6605	<0.001	109.7468	<0.001
11	253.0442	<0.001	109.7570	<0.001
12	253.8544	<0.001	109.8162	<0.001

从上图可以看出, Q 统计量利用从 1 到 12 的时滞窗检验了方差随时间的变化。这些检验强烈地指出有异方差性的存在, 对于所有的时滞窗其 p 值都小于 0.0001。拉格朗日乘子检验也指出有异方差性的存在。

四、异方差的处理

如上一章所述, 如果异方差的形式已知, 那么加权回归是一种很好的方法。然而通常情况下, 误差方差未知, 只能由数据来估计, 所以需要对变化误差方差建模。广义自回归条件异方差模型 ($GARCH$) 模型是一种普遍采用的为带异方差的时间序列建模的方法。

目前在未知异方差函数的情况下, 我们需要用 $GARCH$ 来进行建模。拉格朗日乘子检验可以帮助确定对数据使用的 $ARCH$ 模型的阶。此检验直到 12 阶都是显著的, 这说明需要非常高阶的 $ARCH$ 模型。

我们知道 $ARCH$ 模型是短记忆过程, 它仅用最近的 q 个残差平方来估计变化方差。而 $GARCH$ 模型为一长记忆过程, 它使用过去的所有残差平方和来估计当前的方差。上图的检验结果建议我们使用 $GARCH$ 模型而不是 $ARCH$ 模型。

检验一个模型有效性的工具是正态性检验, 正态性检验的原理是基于 $GARCH$ 模型的基本假定: 当异方差模型拟合好后, 如果拟合的准确, 对 $AR-GARCH$ 模型, 函数 $u_t = v_t / \sqrt{h_t}$ 会呈现出显著的标准正态分布。

(一) 假设条件

H_0 : u_t 服从标准正态分布

H_1 : u_t 不服从标准正态分布

(二) 检验统计量

1982 年 Bera 和 Jarque 基于峰度和偏度, 给出了正态检验统计量:

$$T_n = \frac{n}{6}b_1^2 + \frac{n}{24}(b_2 - 3)^2$$

式中, b_1 为偏度函数; b_2 为峰度函数。 $T_n \sim \chi^2(2)$, 在显著性水平为 α 的条件下, 如果 $T_n \geq \chi_{1-\alpha}^2(2)$, 即 P 小于 α , 就拒绝 H_0 , 反之则接受 H_0 。

本文中我们建立的模型 $AR(1,2,5)-GARCH(1,1)$ 的正态性检验结果是 T_n 的值为 3.4432, P 值为 0.1788, 在 0.05 的显著性水平下, 接受原假设, 这与 $\varepsilon_t / \sqrt{h_t}$ 服从正态分布相吻合。而且整个模型的 R^2 高达 0.9981。所以可以认为该模型拟合成功。由计算结果知, 最终建立的 $AR-GARCH$ 模型为:

$$\begin{cases} ma7_t = 1838 + u_t \\ u_t = 1.8347u_{t-1} - 0.8636u_{t-2} + 0.000995u_{t-3} + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t}a_t, a_t \stackrel{i.i.d}{\sim} N(0, 84.14131) \\ h_t = 77.0793 + 0.0858\varepsilon_{t-1}^2 + 5.658 \cdot 10^{-19}h_{t-1} \end{cases}$$

五、拟合与预测

当模型建立好之后, 我们就可以用估计好的模型来进行拟合和预测了。

拟合与预测效果如下图所示:

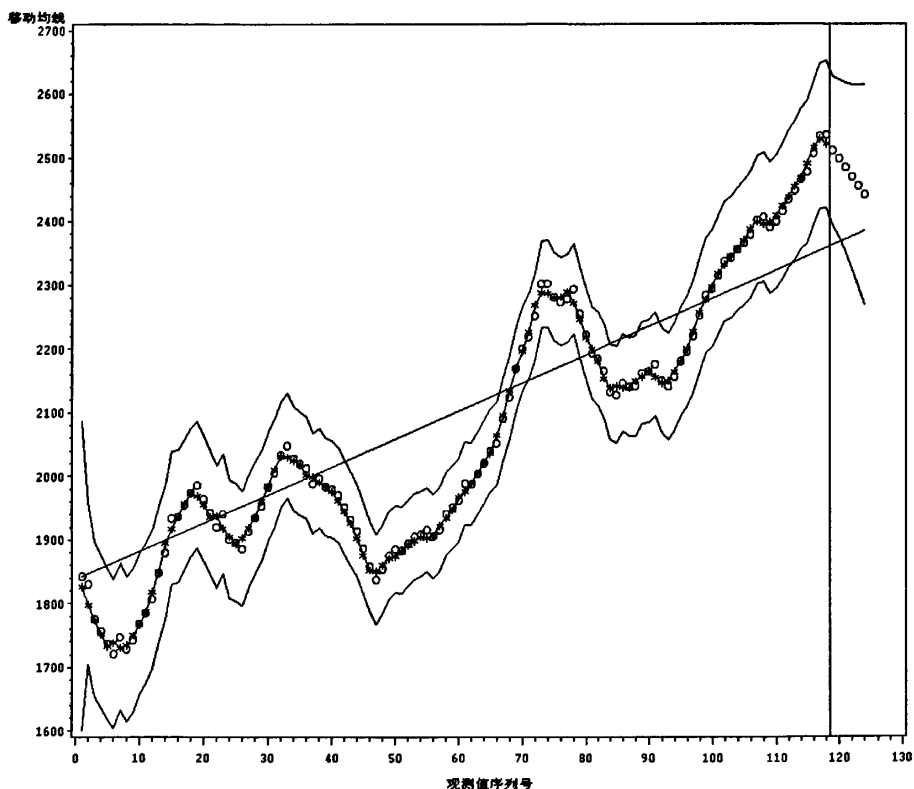


图 6.7: AR-GARCH 模型拟合与预测效果图

由拟合与预测效果图可以看出, AR-GARCH 模型的拟合效果非常好, 由作出的短期预测图可知, 未来 6 天内 7 日均线是一个下降的态势, 此时我们就应该停止买入, 以回避风险。

第三节 小结

本章所建立的模型都取得了很好的拟合效果, 也具备比较好的短期预测效果, 但是发现随着预测步长的增加, 预测的准确性会降低, 而且随着样本数据的变化, 那么模型的参数和结构也会发生变化。这说明股票价格趋势的波动模式具有易变性与短期稳定性, 即模型对样本具有敏感性。

利用单一模型进行预测的步长越长, 预测效果越差。金融市场是一个复杂的系统, 市场所表现出来的行为实际上是投资者群体行为的一种统计结果。价

格趋势的发展过程体现了投资者群体心理特征的变化轨迹，在一个阶段是一种心态，而在另外一个阶段则是另外一种心态，正是这种心态的变化导致了价格的变化，进而导致了预测模型的变化。我们进行趋势分析，就是要及时把握群众心理的这种变化，从而预测未来的行情趋势可能向什么方向发展。

股票价格预测是一个充满争议的问题，时间序列分析方法是对市场行为进行统计分析的有效工具，有着广阔的前景。实证研究表明，股票价格波动具有内在规律性，我们通过对历史价格时间序列数据可以建立价格波动模型，进行短期预测，可以帮助我们发现投资机会、回避投资风险，但在应用过程中，要坚持多方法验证。基本面分析、政策面分析仍然有重要的应用价值，从长期来看，价格是由内在价值决定的，但从短期来看，价格与内在价值关系不大，特别对中小投资者，应该充分发挥自身资金量不大，进出市场比较容易的优势，进行波段操作，最大程度地提高投资收益。本文所建立的时间序列模型，具有较好的短期预测效果，可以实现动态的连续预测，对投资者回避风险和把握买卖时机是有重要实用价值的。

参考文献

- [1]高惠璇等. SAS 系统 SAS/ETS 软件使用手册[M]. 北京: 中国统计出版社, 1998.
- [2]朱世武, 严玉星. 金融数据库[M]. 北京: 清华大学出版社, 2007.
- [3]王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2008.
- [4]朱世武. 基于 SAS 系统的金融计算[M]. 北京: 清华大学出版社, 2004.
- [5]朱世武. SAS 编程技术教程[M]. 北京: 清华大学出版社, 2007.
- [6]马超群, 兰秋军, 陈为民. 金融数据挖掘[M]. 北京: 科学出版社, 2007.
- [7]王振龙著. 时间序列分析[M]. 北京: 中国统计出版社, 2002.
- [8]Engel. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K Inflation [J]. Econometrica, 1982, (50): 987—1008.
- [9]Bollerslev Tim. Generalized Autoregressive Conditional Heteroscedasticity [J], Journal of Econometrics, 1986: 31—32.
- [10]Christian Gouriéroux. ARCH Models and Financial Applications [J]. New York: Springer—verlag, 1997, 27—28.
- [11]刘业政, 王毅, 杨攀. 集成时序分析与形态理论的证券价格预测方法[J]. 应用数学与计算数学学报, 2006.6, 10~18.
- [12]章劲松. 金融时间序列分析的非线性方法研究[D]. 安徽: 中国科技大学, 2002.
- [13]任彪. 资本市场非线性特征及预测理论的若干问题研究[D]. 天津: 天津大学, 2005.
- [14]表二苏. 证券价格预测方法研究[D]. 天津: 天津大学, 2007.
- [15]韩文蕾. 中国股票市场的非线性分析与预测[D]. 西安: 西北工业大学, 2005.
- [16]陈胜荣. 技术分析有效性的实证研究[D]. 厦门: 厦门大学, 2001.
- [17]顾岚. 时间序列分析预测与控制[M]. 北京: 中国统计出版社, 1997.
- [18]Andrew W.Lo, A.Crnig Mackinlay, John Y.Campbell, The Econometrics of Finacial market, Princeton University Press, 1997.
- [19]李红梅. 简介一种预测股票价格的方法[J]. 中国科技信息, 2006(17):161.
- [20]胡桔州. 金融时间序列的数据挖掘技术与经典统计模型比较[J]. 005.6; 95-98.
- [21]杨钟瑾. 浅谈时间序列的分析预测[J]. 中国科技信息, 2006(14) 207~208.
- [22]赵志峰. 对建立中国股票价格指数时间序列模型的探讨[J]. 统计与信息论坛, 2003.1:

66~69.

- [23]向书坚. 20 世纪 90 年代时间序列预测领域主要研究动态[J]. 中南财经大学学报, 2001(12),31~36.
- [24]张大海, 江世芳等. 灰色预测公式的理论缺陷及改进[J]. 系统工程理论与实践, 2002(8).
- [25]刘文抒. 基于灰色理论和 ARIMA 模型的上证指数研究[D]. 南京: 海河大学, 2005.
- [26]李红梅. 股票分析和预测系统[D]. 吉林: 吉林大学, 2004.
- [27]彭章艳. 沪深股票市场非线性特征及其预测模型研究[D]. 武汉: 武汉理工大学, 2004.
- [28]Eugene F.Fama, Efficient Capital Market: A Review of Theory and Empirical Work, Journal of Finance, 1970, Vol.25(2):383~417.
- [29]Fama E F, Fisher L, etal, The Adjustment of Stock Prices to New Information, International Economics Review, February 1969:1~21.
- [30]Peters E.E., Chaos and Order in the Capital Market(Second Edition). John Wiley & Sons, Inc., New York, 1996.
- [32]Goetzmann, W., Patterns in Three Centuries of Stock Market Prices, Journal of Business 66(2), 1993: 249~337.
- [33]Willey T, Testing for nonlinear dependence in daily stock indices, Journal of Economics and Business, 1992(44): 63~74.
- [34]Mandelbrot B.B.,Taqqu M.S.Robust R/S Analysis of Long Run SerialCorrelation[J].Bulletin of the International Statistical Institute,1979,49:69~99.
- [35]Cocharane John H.By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. Journal of Political Economy[J], April 1999:205-251.
- [36]Michael D McKenzie.Non-Periodic Australian Stock Market Cycles: Evidence from Rescaled Range Analysis[J].The Economic Record,2001,77:393-406.
- [37]Sentana E., Wadhwani S. Feedback Traders and Stock Return Autocorrelatoins: evidence from a Century of Daily Data. The Economic Journal.1992,102(411): 415-425.
- [38]Werbos P. Beyond regression: new tools for prediction and analysis in the behavioral sciences[D].Ph D Dissertation, Harvard University,1974.

附录

实证研究所用 SAS 代码

一、ARIMA 建模所用代码

```

proc gplot data=resdat.ma7;
plot ma7*date=1;
symbol1 c=black v=star i=join;
label ma7="移动均线" date="日期";
run;
data a;
set resdat.ma7;
dma7=dif(ma7);
run;
proc gplot data=a;
plot dma7*date=1;
symbol1 c=black v=none i=needle;
label dma7="移动均线一阶差分" date="日期";
run;
proc arima data=resdat.ma7;
identify var=ma7(1) nlag=10 outcov=out;
run;
proc gplot data=out;
plot corr*lag=1 partcorr*lag=1;
symbol1 c=black v=none i=needle;
label corr="自相关系数" lag="滞后阶数";
label partcorr="偏自相关系数" lag="滞后阶数";
run;
proc arima data=resdat.ma7;
identify var=ma7(1) nlag=10 outcov=out;
estimate p=1 q=1 noconstant;
run;
forecast lead=6 id=date out=results;
run;
proc gplot data=results;
plot ma7*date=1 forecast*date=2/overlay href='21apr2009'd;
symbol1 c=black i=join v=none;
symbol2 c=black i=join v=star;
label ma7="移动均线" date="日期";
run;

```

二、AR-GARCH 建模所用代码

```

proc gplot data=resdat.nma7;
symbol1 v=star i=join c=black;
symbol2 v=none i=r c=black;
plot ma7*t=1 ma7*t=2 /overlay;
label ma7="移动均线" t="观测值序列号";
run;

proc autoreg data=resdat.nma7;
model ma7=t / dw=5 dwprob;
run;

proc autoreg data=resdat.nma7;
model ma7=t / method=ml nlag=5 backstep;
run;

proc autoreg data=resdat.nma7;
model ma7=t / nlag=(1 2 5) archtest dwprob;
run;

proc autoreg data=resdat.nma7;
model ma7=t / nlag=(1 2 5) garch=(q=1,p=1);
run;

data b;
ma7=.;
do t=119 to 124;
output;
end;
run;

data b;
merge resdat.nma7 b;
by t;
run;

proc autoreg data=b;
model ma7=t / nlag=(1 2 5) garch=(q=1,p=1);
output out=p p=yhat pm=ytrend lcl=lcl ucl=ucl;
run;

proc gplot data=p;
plot ma7*t=1 yhat*t=2 ytrend*t=3
lcl*t=3 ucl*t=3 / overlay href=118.5;
symbol1 v=star i=join c=black;
symbol2 v=circle i=none c=black;
symbol3 v=none i=join c=black;
label ma7="移动均线" t="观测值序列号";
run;

```

致谢

光阴荏苒，在美丽的厦大不知不觉已经生活学习了三年，而这三年也是我人生重要的转折点。在这写下致谢词的同时，也意味着这一段学习生活的结束。在这段时间里，特别要感谢导师陈珍珍教授的悉心指导，让我能够在这短短的三年里，学习到了许多的知识。在本论文的完成过程中，从选题、结构设计、修改直到最终定稿，倾注了导师的大量心血和汗水，陈老师精益求精的研究精神、孜孜不倦的工作态度深深地影响了我，使我在学习和生活中受益匪浅。我还要感谢陪我走过陪我经历这一切的所有朋友、同学、老师、家人。感谢你们对我的帮助，感谢你们给我的教导，感谢你们给我的温情。

三年时间转瞬即逝，回首往事感慨万千。研究生阶段只是我追求的开始，在以后的日子里我会更加勤奋的工作，努力去征服一个个的困难，实现我人生的梦想。

最后，再次谢谢所有帮助过我的人们，谢谢……