

学校编码: 10384
学号: 200215097

分类号 _____ 密级 _____
UDC _____

厦 门 大 学
硕 士 学 位 论 文

基于日内高频数据的股票收益率
分布特征与变化研究

A study on the Character and Movement of Stock
Returns Distribution Base on Intraday High Frequency
Data

罗智超

指导教师姓名: 吴世农 教授

专业名称: 工商管理(MBA)

论文提交日期: 2005年5月

论文答辩时间: 2005年6月

学位授予日期: 2005年 月

答辩委员会主席: _____

评 阅 人: _____

2005年5月

厦门大学学位论文原创性声明

兹呈交的学位论文，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的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文而产生的权利和责任。

声明人（签名）：

年 月 日

摘 要

在现代金融理论中，对资产收益率的分布研究是一个非常重要的分支。本文利用中国股市 1998 年 1 月 1 日至 2003 年 12 月 31 日的 5 分钟高频交易数据，对沪深两市市场 30 只股票的 5 分钟收益率及世界主要市场指数（包括新兴市场和成熟市场）日收益率的特征与变化进行了详细的基础性研究。研究内容主要包括：1、不同市场态势下收益率分布的统计特征；2、高频数据下的收益率日内、星期效应；3、拟合收益率分布函数（正态、指数、韦伯、分形）；4、检验股价动态行为是否满足随机游走模式；5、比较分析世界主要股票市场指数收益率特征及变化。

经研究，得出以下结论：1、不同市场态势下不同的行业股票收益率呈不同分布特征，正态性随时间间隔增加而增加；2、半小时内的 5 分钟收益率序列呈显著自相关；3、收益率均值和方差都存在显著日内效应；4、正态分布、指数分布、Weibull 分布都与实际的分布差异较大，而用稳定分布拟合效果较好，股票价格呈有偏随机游走态势；5、与钟蓉萨（1999）研究结果比较，新兴市场和成熟市场指数收益率分布近年来都发生了显著的变化。

本文共分四章：第一章前言对论文的选题及研究内容、研究意义做简要的描述并对国内外对收益率分布及其特征的研究文献做简要的综述；第二章对中国股票市场收益率进行系统详细的研究，并介绍了论文的研究数据、方法、步骤详细；第三章对世界上 16 个地区（10 个新兴市场、6 个成熟市场）的股票指数的收益率分布进行横向和纵向的比较分析，揭示世界股票市场收益率分布的特征和变化趋势。第四章对本文的研究内容进行总结并提出进一步研究方向。

关键词：股票市场 收益率分布 日内高频

Abstract

The returns distribution research is a very important branch of contemporary finance theory. This thesis uses the 5 minute interval high frequency data of Chinese stock market from Jan 1st 1998 to Dec 31st 2003. And selects 30 stocks from all kinds of industries and the daily data of mainly global stock market index including emerging market and developed market. The main research contents include: 1.the different statistic character of returns distribution during different time range. 2. the intraday and weekly effect under the high frequency data. 3. fit the distribution function with different functions. 4. test the behavior of stock price. 5.compare the different of returns distribution character between stock market index.

After the research, the thesis has concluded that: 1. the distribution character of different stocks during different time are discriminative.2. the coefficients of 5 minute time auto regression series are correlate during 30 minutes. 3. both the average and variance of returns have intraday effect. 4. the stable distribution fit the stocks returns distribution best. 5. compare with the research of Zhongrongsha(1999), the characters of market index distribution have changed evidently recent years, including both the emerging and developed market.

This thesis consists of four chapters. Chapter one is an introduction, discussing the research contents、 research meanings and reviewing the related literatures. Chapter two makes a detail research of the returns distribution of Chinese stock market. And also introduces the research data、 method and process. Chapter Three compares the stock index of 16 countries or districts including 10 emerging market and 6 developed countries. And discovers the trend of the change. Chapter Four is the conclusion and further research fields.

Key Words: Stock Market; Returns Distribution; Intraday High Frequency Data

目 录

第一章 研究绪论	1
第一节 本文研究的主要问题及意义.....	1
第二节 研究框架.....	5
第三节 研究文献评述.....	7
第二章 中国股票市场收益率研究	12
第一节 研究样本及数据.....	12
第二节 研究步骤与方法.....	13
第三节 实证研究结果及分析.....	15
第三章 世界股票市场收益率研究	33
第一节 收益率基本统计特征及差异研究.....	33
第二节 股票市场指数统计特征比较研究.....	34
第四章 结论与启示	42
第一节 研究结论.....	42
第二节 启示.....	44
第三节 研究局限.....	44
第四节 进一步研究方向.....	46
附 录	47
参考文献	51
后 记	53

Contents

Charter 1 Introduction	1
1. Major research contents and meanings.....	1
2. The framework of research.....	5
3. Literature review.....	7
Charter 2 Research on Chinese Stock Returns Distribution	12
1. Sample and data of research.....	12
2. Process and method of research.....	13
3. Empirical results and analysis.....	15
Charter 3 Research on Global Stock Market Returns Distribution	33
1. Basic statistical characters.....	33
2. The difference between emerging and developed market.....	34
Charter 4 Conclusions and Suggestions	42
1. Conclusions.....	42
2. Reveal.....	44
3. Deficient.....	44
4. Further research.....	46
Appendix	47
Reference	51
Postscript	53

第一章 研究绪论

在现代金融理论中，对资产收益率的分布研究是一个非常重要的分支。在许多经典的理论常常假定股票收益率服从正态分布，但许多专家学者对这一经典假设作了大量的理论与实证研究后，发现许多研究结论并不支持收益率正态分布的假说，从而对经典的理论的假设基础提出了质疑。到底收益率分布满足什么分布？收益率分布具备什么特征？不同类型的股票在不同时段的收益率有什么区别？不同国家的股票收益率有什么差别？不同时间频率下的收益率分布有什么不同？本文的选题正是基于对以上问题的思考。

目前对中国股票市场收益率分布的研究，大多数局限于股票指数的分布研究，并且在研究时间上采用的是日或月收益率等低频数据。而对基于高频数据下个股的收益率分布及其特征变化的研究比较少。因此本文利用中国股票市场 1998 年至 2003 年的 5 分钟高频交易数据，对不同行业的 30 只样本股票在不同的市场态势（熊市和牛市）下的收益率分布情况进行研究，并拟合部分个股的收益率曲线，计算个股收益率分布曲线在高频数据下的特征值，检验个股收益率的日内效应，检验股票价格的动态行为，最后比较研究了新兴市场 and 成熟市场指数收益率分布的特征与变化。

通过对以上问题的研究，本文发现了许多在低频数据下无法发现的现象，另一方面，通过对以上对金融市场微观结构方面的基础性研究也希望有利于其他研究人员进一步研究与检验金融市场的其他理论。

第一节 本文研究的主要问题及意义

John Y. Campbel (1997)¹认为在金融计量研究中收益率分析之所以占据主导地位有两个方面的原因：其一，对大多数投资者而言，金融市场也

1. Campbell, J. The Econometrics of Financial Markets, Princeton University Press, 1997.

许被认为是接近完全竞争的，所以投资规模并不会影响股价的变化，因此投资技术不受投资规模的影响，收益率是对市场投资机会最完美的反映；其二，无论是出于理论分析的需要还是由于经验研究的方便，收益率相对股票价格有更诱人的统计特征，如平稳性与遍历性。本文将从以下几个方面对研究中国股票市场收益率的意义展开讨论。

一、揭示中国股票收益率在高频数据下的分布特征

高频数据主要指以小时、分钟甚至每一笔交易作为记录对象，它与传统的低频数据（如周、日数据）相比，具备一些独特的特征。随着信息技术的不断提升与普及，高频数据的获得也不再是件困难的事情，因此高频数据的研究已经引起学术越来越浓厚的兴趣。

近年来，高频数据成为学术界研究的焦点主要有以下两个原因。其一，高频数据具有低频数据所不包含的独特属性。如，除了交易价格、交易量以外，还包括询价和报价、每笔交易量、交易间隔等；其二，金融高频交易数据是研究市场微观结构的最好的原始素材。传统理论认为股价的波动是由于噪音引起的，没有必要研究“黑盒”里面发生的数据。而实际上高频数据恰恰包含了引起股票波动等现象的最重要的信息。高频数据的记录使学者能够对“黑盒”内部发生的事件的原因与规律展开研究。

目前国内用高频数据系统的研究股票收益率分布的文献还不多见，本文希望能够通过高频数据的角度分析股票收益率分布特征。

二、拟合中国股票收益率的分布函数

对股票收益率分布函数的拟合与预测是十分重要的。因为，概率密度最好的描述了不确定性（Rothschild & Stiglitz, 1970）。当收益率序列服从高斯分布时，不确定性可以由方差来刻画。但在非高斯分布情况下，不确定的刻画就必须由其他方法来完成。

在宏观经济研究领域，研究通货膨胀、产出及失业率的分布函数有利

于政策制定者拟定合理的政策，如控制通货膨胀率和失业率。

在金融研究领域，分布函数的研究也广泛应用于投资组合、公司债券、流动性度量和风险控制理论。

因此，通过高频数据来描述中国股票收益率的分布函数，必然可以为其他研究领域的进一步研究提供基本数据。了解不同板块的股票组合的收益率分布及其参数估计。

三、为风险管理及投资组合管理提供基础研究

研究收益率分布的对称性是进一步研究收益率分布的基础，而收益率分布函数的确定可以说是整个金融计量研究的基础。不同的收益率分布导致不同的风险状况。传统的均值一方差方法和 β 系数度量风险的方法只有在收益的概率分布是正态分布的条件下才有意义。对于正态分布而言，只要确定期望收益率水平和方差就可以确定投资者面临的风险状况。但如果收益率是非正态分布，就不能仅仅由期望收益率和方差来确定投资者面临的风险。

研究风险计算的方法有很多种，如 VaR 的计算原理主要是基于证券组合价值变化的统计分布图，如果给出某一证券组合的价值变化的统计分布图，根据 VaR 的定义就可直观找到与置信度相应的分位数即 VaR 值。因而在计算股票收益率的 VaR 值时对收益率分布的假定是一个关键，不同分布假定下的 VaR，计算的结果将会不同。

实证分析表明，收益率分布的尾部越厚，表明该分布假设下小概率事件出现的频率相对正态分布越高，估计的风险也就越大，最终的投资策略也应该越保守。因此，分析收益率的分布对正确的投资决策至关重要。

1952 年 Markowitz 在“Portfolio Selection”中提出了用股票收益的方差来衡量风险，组合中的收益为组合中所有股票期望收益率的加权平均，组合的方差为组合中单个资产收益的方差和各个资产收益之间的协方差。实证研究发现 Markowitz 的理论基础股票收益率服从正态分布的假设与实

际情况不一致。因此，股票收益分布的多样性与复杂性也决定了投资组合收益率及风险的准确计量的困难。

四、为股票市场非线性系统理论研究提供证据

Mandelbrot 首先将分形的概念引入经济学。1963 年，他研究发现市场商品价格变化与时间之间呈现出统计自相关的性质。求得 19 世纪棉花价格变化的分维 $D_e=1.7$ ，并研究了利润变化、股票市场的行情、居民的收入及资产和财产的分布等经济问题，并认为所有这些问题都与分形有关。1964 年，他提出资本市场收益率服从稳定帕累托的分布，即分形分布，在均值处有高峰、胖尾，其特点是趋向有趋势和循环，也有突然和不连续的变化，而且它们可按偏斜度调整。这些胖尾分布常常显示出由非线性随机过程所产生的一种具有长期记忆的迹象。分形分布具有以下几个特征：（1）自相似性；（2）可加性；（3）粗尾形态。

非线性系统理论的应用以及 20 世纪 90 年代初艾德嘉.E.彼得斯针对有效市场假说的非现实性提出了一种崭新的资本市场假说——分形市场假说的提出，为资本市场的研究构建了一个新的理论框架，为资本市场特性与结构的研究提出了新方法。诸如市场的分形维数和非线性结构，市场对于信息反应的非线性因果关系，混沌、突变、行为金融学、协同市场假说等非线性理论等问题的研究，对于无论是投资者还是市场的监管者和政策的制定者，都具有重要的意义。

资本市场是一个复杂的系统，它包括贪婪非理性的、谨慎理性的投资者，包括长期的和短期的投资者，仅仅用线性模型来简化复杂的资本市场是具有一定局限性的，尝试着用非线性的分析方法来讨论资本市场问题也是一个很好的创新。

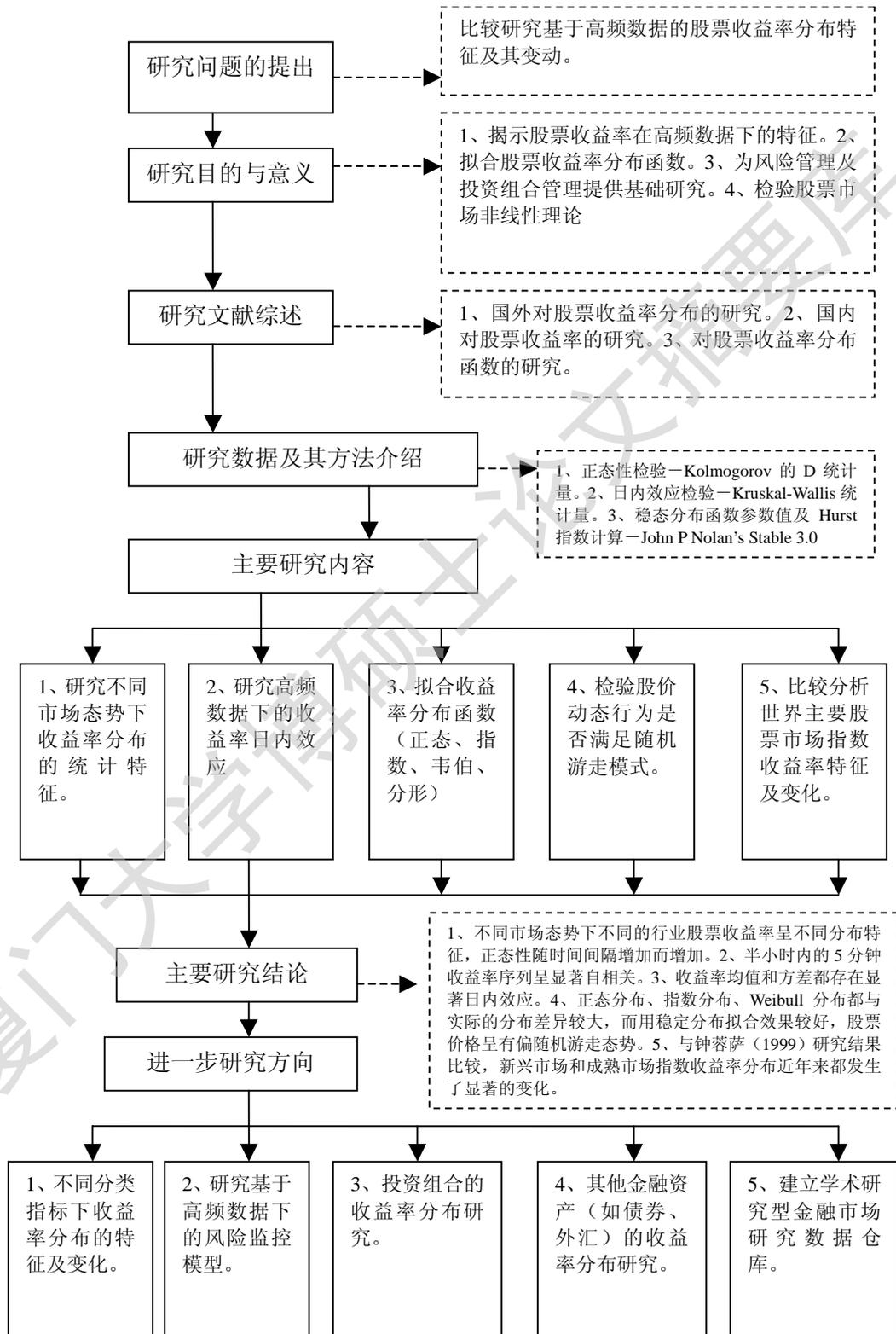
本文尝试使用中国股票市场 5 分钟的交易数据，拟合股票收益率的分形分布模型，并估算相应的参数，验证与传统 EMH 相违背的现象。

第二节 研究框架

本文的组织是这样安排的：第一章前言对论文的选题及研究内容、研究意义做简要的描述并对国内外对收益率分布及其特征的研究文献做简要的综述；第二章对中国股票市场收益率进行系统详细的研究，其中先对本论文的研究数据、方法、步骤进行详细的介绍。其次在以下几个方面展开研究：不同市场态势下收益率分布的统计特征研究；高频数据下的收益率日内效应检验；利用正态、指数、韦伯、分形等分布对收益率分布进行拟合；检验股价动态行为是否满足随机游走模式。第三章对世界上 16 个地区（10 个新兴市场、6 个成熟市场）的股票指数的日收益率分布进行横向和纵向的比较分析，揭示世界股票市场收益率分布的特征和变化趋势。第四章对本文的研究内容进行总结。首先提出本文的研究得到的几个结论，其次对提出本文的研究的启示，接着对研究局限性进行分析，最后提出笔者在该研究领域将进一步研究的工作。图 1.1 详细的描述了本文的总体分析思路。

本文的总体分析思路为：首先提出研究的问题——高频数据下的股票收益率分布的特征及其变动；其次，介绍研究工具及其方法；再次，对收益率特征及其变动的 5 个方面展开研究；最后得出结论及启示。

图 1.1 本文研究分析思路框架图 (—▶表示分析框架主线---▶表示解释说明)



第三节 研究文献评述

一、国外对股票收益率分布的研究

Fama(1965)²研究了道琼斯工业平均指数的 30 只股票的收益率序列的分布特征与序列相关性,验证了随机游走模型的适用性。确定了随机游走模型在描述股价行为中的核心地位。

Summers(1986)、Fama 和 French(1988)、Lo 和 Mackinlay(1988)及 Poberba 和 Summers(1988) 发现美国股票市场股票收益不服从随机游走的证据。

Alexander (1961)认为尖峰、肥尾是金融资产收益率的基本特征,用正态分布来描述金融资产的短期收益率是不太合适的。Mandelbrot(1963) 研究了投机资产的价格行为,提出了用稳定的 Paretian 分布来描述收益率。Gray 和 French(1990) 等提出了用 t 分布或几个正态分布的组合来描述这些资产的收益率,从而考虑收益率的尖峰、肥尾特征。

二、国内对股票收益率分布的研究。

陶亚民、蔡明超和杨朝军(1999)³分别运用柯莫哥洛夫拟合优度检验法和异方差的t 检验法对上海股票市场股票收益率的分布特征进行了实证分析。研究发现在排除异常事件干扰的情况下,收益率服从正态分布。

徐龙炳(2001)⁴应用稳态分布理论实证研究中国股票市场股票收益的特征。研究表明,中国股票市场股票收益构成的时间序列呈现类峰、厚尾,具有稳态特征,根据稳态分布理论,此时方差不确定或无穷,用它作为风险尺度的样本方差近乎无意义

徐迪、吴世农(2002)⁵以上证综指为样本,应用分形市场理论对上海股票市场进行R/S分析,得出上海股票市场Hurst指数为 0.61,分形维数为 1.64,平均记忆长度为 350 天。这进一步说明上海股市不同时间资产收益率

2. Fama The Behavior of Stock-Market Price Journal of Business,1965.

3. 陶亚民、蔡明超和杨朝军. 上海股票市场收益率分布特征的研究. 预测. 1999,(2):57.

4. 徐龙炳. 中国股票市场收益稳态特性的实证研究. 金融研究. 2001,(6).

5. 徐迪, 吴世农. 上海股票市场的分形结构分析. 中国经济问题. 2002,(1)..

之间不是相互独立的，不是服从正态分布。

王建华、柯开明和王玉玲（2004）⁶认为非对称拉普拉斯分布有明确的解析表达分布函数，只含有 2 个参数，能够较好拟合股票收益率分布的尖峰厚尾偏倚特性。

余卫军和张新生（2004）⁷提出了一种类似Weibull分布的函数来拟合上证指数收益率分布，模拟结果较好。并按通常的方法对尾部作了截尾处理，以更接近实际数据在尾部表现出来的“厚尾”现象。

三、主要股票收益率分布函数概述

（一）正态分布

1900年3月29日法国著名数学家Bachelier在其提交给法国巴黎科学院的博士论文中提出投机资产价格服从维纳过程，即投资资产的无条件分布为正态分布。Osborne（1959）⁸将描述股票收益率的模型改进为Bachelier-Osborne模型，认为用几何布朗运动来描述股价的变动是合理的，对数股价差分的无条件概率分布为正态。但是，近年来越来越多的研究发现股票收益率并不服从正态分布，而是具有“尖峰、厚尾”的特征。

（二）稳定分布

稳定分布也称分形分布，又称为 Pareto 分布、Pareto-Levy 分布或 Stable-Paretian 分布。这些分布的性质最早由 Levy 推导出来，而他的工作又是以 Pareto 有关收入分布的工作为基础的。近年年来，大量实证研究发现股票收益不是正态分布，相对正态分布来说，有明显的“尖峰”和“肥尾”，Mandelbrot 将之称为“稳定帕累托” (Stable Paretian) 分布。

稳定分布得到广泛应用的原因有：其一，许多研究表明收益率分布服从非Gauss稳态分布；其二，广义中心极限定理表明对大量的独立同分布

6. 非对称拉普拉斯分布在股票收益率中的应用. 武汉理工大学学报. 2004,(3).

7. 余卫军、张新生. 上证指数收益率分布的拟合. 经济数学. 2004,(3).

8. Osborne. Brownian Motion in Stock Market. Operations Research. 1959,(10).

随机变量的和进行标准化处理后，如果极限分布存在，则此分布一定属于稳态分布族；其三，由于许多大量数据集具有尖峰、厚尾的特征，用Gauss分布描述效果不好，但用稳态分布拟合效果很好Nolan（1999a）⁹。

稳定分布是含有4个参数的分布，一般没有解析表达式，由特征函数来描述。特征函数定义如下：

$$\lg f(x) = \begin{cases} i\delta t - \gamma|t|^\alpha (1 + i\beta(t/|t|) \tan(\alpha\pi/2)) & \alpha \neq 1 \\ i\delta t - \gamma|t|(1 + i\beta(t/|t|) \frac{\pi}{2} \lg|t|) & \alpha = 1 \end{cases} \quad (1)$$

稳定分布的4个参数分别为 $\alpha, \beta, \gamma, \delta$ 。 α 是稳定性指数，标志着分布的峰度及尾部的厚性，它也决定了分布的矩、随机变量和的分布以及其标准化等特征。 $\alpha \in (0, 2]$ ，当 $\alpha = 2$ 时稳定分布退化成正态分布，此时均值为 δ ，方差为 2γ ；当 $0 < \alpha < 2$ 时稳定分布相对正态分布有更厚的尾部， α 值越小，分布在尾部取值的概率越大。

当 $0 < \alpha \leq 1$ 时，不存在稳定均值，但此范围的 α 比较罕见；当 $1 \leq \alpha < 2$ 时，方差不确定或无穷，具有稳定均值，此时 α 对应于具有长期关联和统计上自我相似的分形布朗运动， α 就是时间序列的分数维。这种自我相似特征就是Mandelbrot(1983)意义下的分形。分形市场理论的研究表明，股票价格表现出长期依存的关系，具有状态持续性，表现为价格的连续涨跌，波动呈现集群性的特点。

β 是偏度（Skewness）指数， $\beta \in [-1, 1]$ ，当 $\beta = 0$ 时，分布是对称的； $\beta > 0$ 时，分布右偏， $\beta < 0$ 分布左偏， β 值越大偏离程度也越大。当 $\alpha = 1$ ， $\beta = 0$ 时，该分布为Cauchy分布；

γ 是稳定分布的尺度参数，表示曲线的宽度。 δ 是稳定分布的某一位

9. Nolan, J.P., 1997, "Numerical Computation of Stable Densities and Distribution Functions"

置参数, 当 $\beta = 0$ 时, δ 为分布的中位数。对于稳态分布可以通过 γ 、 δ 进行标准化, 通过标准化得到 $S(x; \alpha, \beta, \gamma, \delta) = S((x - \delta) / \gamma; \alpha, \beta, 1, 0)$ 。

(三) t 分布

t 分布包含三个参数, 即位置参数 m , 尺度参数 $h > 0$, 以及自由度参数 $v > 0$ 。其密度函数表达式为:

$$f(x | m, h, v) = \frac{v^{(1/2)v}}{B(\frac{1}{2}, \frac{1}{2}v)} [v + h(x - m)]^{-1/2(v+1)} \sqrt{h} \quad (2)$$

其中 $B(a, b)$ 为贝塔函数, $B(a, b) = \Gamma(a)\Gamma(b) / \Gamma(a + b)$, $\Gamma(\cdot)$ 为伽玛函数。

t 分布具有以下基本特征:

(1) 大于自由度 v 的高阶矩不存在, 只有当 $v > 1$ 时, 分布的均值才存在, $Var(x) = h^{-1}v / (v - 2)$;

(2) 当 $v = 1$ 时, t 分布即为 Cauchy 分布, 这与稳定 Paretian 分布相一致, 当 $v \rightarrow \infty$, t 分布即收敛于正态分布。

(四) 混合分布模型

由于正态分布和稳定分布模型都假定股票价格变化是独立同分布的随机变量, 且交易在时间上是均匀分布的。而实际上由于信息的到达时间以及不同时刻不同信息对股票的影响可能会不同, 交易时间的发生时间不均匀都可能导致同分布的假设不一定能满足。因此提出混合分布模型。

混合模型一般由两个分布复合而成, 其中一个为正态分布, 另一个则控制该正态分布的方差的变动。可以由一个从属的随机过程来描述。令 $\{x(s); s \geq 0\}, \{h(s); s \geq 0\}$ 表示两个随机过程, 定义从属于 $x(s), h(s)$ 的随机过程 $z(s)$, 令 $\{z(s) = x(h(s)); s \geq 0\}$ 。一般 $x(s)$ 表示在长度为 s 的区间上某股票的收益率, 假定 $x(s)$ 为高斯稳定过程, 并对其做适当的中心化处理, $E(x(s)) = 0$ 。 $h(s)$ 表示经济环境以及相关信息的变化, 由 $h(s)$ 来刻画 $x(s)$ 的方差变化。¹⁰

10. 封建强. 股价行为理论与实证分析. 经济科学出版社. 2004.

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

廈門大學博碩士論文摘要庫