



学 校 代 码 : 10459

学号或申请号 : 201312141648

密 级 : _____

郑 州 大 学

硕 士 学 位 论 文

基于稳定分布的高频交易研究

作 者 姓 名: 孙 晓 姿

导 师 姓 名: 李 华

学 科 门 类: 理 学

专 业 名 称: 应 用 数 学

培 养 院 系: 数 学 与 统 计 学 院

完 成 时 间: 2016 年 5 月

A thesis submitted to
Zhengzhou University
for the degree of Master

High Frequency Trading Analysis Based on Stable Distribution

By xiaozi sun

Supervisor: Prof. Hua Li

Applied Mathematics

School of Mathematics and Statistics

May, 2016

原创性声明

本人的学位论文是在导师指导下独立撰写并完成的，学位论文没有剽窃、抄袭等违反学术道德、学术规范的侵权行为，否则，本人愿意承担由此产生的一切法律责任和法律后果，特此郑重声明。

学位论文作者：

年 月 日

学位论文使用授权声明

本人在导师指导下完成的论文及相关的职务作品，知识产权归属郑州大学。根据郑州大学有关保留、使用学位论文的规定，同意学校保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权郑州大学可以将本学位论文的全部或部分编入有关数据库进行检索，可以采用影印、缩印或者其他复制手段保存论文和汇编本学位论文。本人离校后发表、使用学位论文或与该学位论文直接相关的学术论文或成果时，第一署名单位仍然为郑州大学的，适用以上学位论文使用授权声明。保密论文在解密后应遵守此规定。

学位论文作者：

年 月 日

摘 要

本文主要研究如何通过构造服从非正态稳定分布的资产组合，来建立收益稳定的高频交易策略组合。根据有效市场理论，有效或弱有效的市场中资产收益率呈正态分布。而行为金融研究发现，由于市场对信息过度反应，时常会出现效率缺陷，表现为资产收益率的尖峰厚尾现象。交易策略的设计一般都是针对某种市场效率缺陷或厚尾现象，且高频数据比低频数据更容易呈现尖峰厚尾现象，因此本文主要研究高频交易策略。

本文通过组合不同的资产，构造厚尾程度较高的稳定分布，然后按照组合的配置，分配各个资产的高频交易资金，从而建立效率相对较高的高频交易策略组合。实证研究发现，厚尾分布程度越强，高频交易策略组合的效率越高，因此我们可以通过构造厚尾分布，提高高频交易组合的效率。

关键词：高频交易 厚尾 稳定分布 有效市场假说 投资组合

Abstract

This paper studies how to improve the return of the portfolio of high-frequency trading strategies by constructing the trading assets portfolio of non-normal stable distributions. According to Efficient Market Hypothesis (EMH), asset returns are normally distributed in efficient or weak efficient market. Comparatively, behavioral finance researchers have found that the market is frequently inefficient because of the participants' over-reaction to information, and that assets return exhibits the severe phenomena of excess kurtosis and heavy tails. In general, trading strategies are usually designed to capture certain market efficiency or heavy tail phenomena. On the other hand, compared with the low-frequency data, the high-frequency data exhibit the feature of heavier tail, so in this paper we mainly focus on high frequency trading strategies.

Then, we first select a portfolio of different assets such that the portfolio return follows a stable distribution with a high degree of heavy tail, then according to the proportion of the portfolio we allocate investment capital in every portfolio of high-frequency trading strategies. Thus, we establish a relatively efficient portfolio of HFT strategies. The empirical study shows that, the greater the degree of fat-tail, the more efficient the high-frequency trading strategies. So we may improve the efficiency of high-frequency trading strategies by selecting a fat-tailed distribution portfolio.

Key words: High frequency trading; Heavy tail; Stable distribution; Efficient market hypothesis; Investment portfolio

目录

第一章 绪论	1
1.1 选题背景及意义	2
1.1.1 高频交易的魅力	2
1.1.2 高频交易的成功典范	4
1.1.3 高频交易带来的挑战	5
1.1.4 高频交易存在的风险	6
1.1.5 高频数据的统计特征	7
1.2 文献综述	8
1.2.1 关于高频交易及稳定分布的国外研究情况	8
1.2.2 关于高频交易及稳定分布的国内研究情况	10
1.3 本文基本框架、研究方法与创新之处	12
1.3.1 本文基本框架	12
1.3.2 本文研究的方法及创新之处	13
第二章 有效市场假说与高频数据	14
2.1 有效市场假说	14
2.2 高频数据潜在获利机会及其游程检验	17
2.2.1 高频数据潜在获利机会	17
2.2.2 游程检验	18
第三章 稳定分布的理论介绍	25

目录

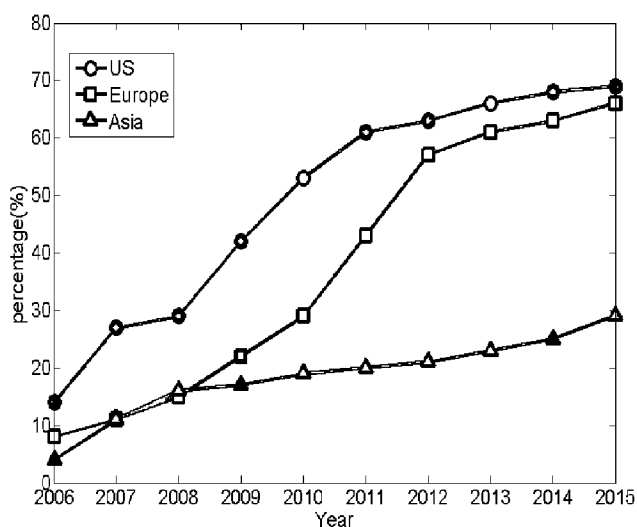
3.1	稳定分布的定义	25
3.2	稳定分布的几种特殊情况	26
3.3	稳定分布的性质	27
3.4	稳定分布的参数估计及数值计算	29
3.4.1	稳定分布的参数估计	29
3.4.2	稳定分布的数值计算	31
3.5	稳定分布与市场数据的拟合	33
3.5.1	稳定分布与指数的拟合	33
3.5.2	稳定分布与高频数据的拟合	34
第四章	构造稳定分布的高频数据及实证研究	40
4.1	数据构造模型介绍	40
4.2	实证分析	41
4.2.1	全球股指期货组合	42
4.2.2	国内跨市场指数组合	47
4.2.3	国内商品组合	50
第五章	结论与展望	55

第一章 绪论

近年来，高频交易(High frequency trading) 在全球金融市场上得到了迅速发展，其主要原因在于它能给投资者带来丰厚的回报。据《Alpha》报道，高频交易的支持者吉姆·西蒙斯(Jim Simons) 仅在2008 年就获得25亿美元的盈利。在2008 年，做高频交易的基金经理大部分都取得了正收益，而这一年70% 的低频交易者的收益率都是负的。高频交易具有巨大的盈利能力，能使投入的资金成指数式的增长，所以吸引了大批的交易者投入大量的精力和金钱对其研究及交易。据2009年Aite集团报告，交易所中60%的交易量都来自于高频交易，目前交易所中高频交易量占比更大，图 1.1 显示了近几年高频交易占总交易量的比例。

高频交易这个词频繁出现在众多研究报告中，何为高频交易？高频是与低频相对而言的，它是一种由复杂的运算和强大的计算机系统所主导，能在毫秒之内自动完成大量买、卖以及取消指令的交易。它能对变化的市场迅速做出反应，并且实现资金的快进快出。高频交易捕捉的是市场细微的获利机会，因此交易次数比较多；由于快进快出的特性，平均每次盈利不会太多。总体而言，高频交易的收益是很可观的，某些高频交易策略能使资金呈指数式增长。

得益于金融市场结构的不断完善、交易品种的日益多样化以及计算机、无线通信和互联网技术的发展和革新，高频交易在全球金融市场得到了快速发展。在中国，2010年4月沪深300股指期货上市，这使我国市场具备了完善的高频套利机制，高频交易也进入了一个快速发展的阶段。高频交易革新了证券交易的方式，使证券交易由传统的经纪人竞价、交割与清算发展成了全电子化的竞价、撮合方式。本章的第一节主要介绍选题的背景及意义，第二节主要介绍研究的目的和方法，第三节介绍了本文所做的工作和主要结论。



图片来源: TABB Group:BCG analysis

图 1.1: 美国、欧洲和亚洲的证券市场高频交易比例 (%)

1.1 选题背景及意义

据有效市场理论, 有效或弱有效的市场中资产收益率呈现出正态分布。而行为金融研究发现, 由于市场易对信息过度反应, 时常会出现效率缺陷, 表现为资产收益率的尖峰厚尾现象。交易策略的设计一般都是针对某种市场效率缺陷或厚尾现象, 且高频数据比低频数据更容易呈现尖峰厚尾现象, 因此本文主要研究高频交易策略。下面就从高频交易的魅力及成功典范、高频交易带来的挑战及存在的风险等几个方面来详细叙述本文选题的背景及意义。

1.1.1 高频交易的魅力

高频交易像风暴一样席卷全球, 它的魅力何在? 能够给投资者带来较高的年均回报只是一方面, 对个人投资者和机构而言, 高频交易能给他们带来以下几点优势:

- (1) 规避隔夜风险

随着资本市场全球化，大多数的交易头寸都可以延长至1天甚至更长的时间，但市场存在巨大的波动，某些交易品种具有跳空现象，隔夜持仓对投资者来说具有很大的风险；尤其对于发行基金产品的机构来说，一个巨大的跳空，基金产品将面临被清盘的风险。高频交易一般不会留隔夜头寸，所以能很好地规避隔夜风险。

（2）节约成本

持仓过夜的头寸除了保证金之外的部分需要按照隔夜利率（Overnight carry rate）支付利息，隔夜利率是一种通常比LIBOR 更高的利率。隔夜利息增加了交易成本，压缩资金的获利空间，通常最后使得资金管理人无利可图。所以高频交易能规避隔夜持仓的成本，尤其在资金紧张、利率偏高的情况下，能给投资者节省一大笔成本。

（3）充当另类资产

对于一个充分分散化的投资组合，除了配置股票、债券和现金等传统资产外，还需要配置另类资产。做高频交易的资产可以作为一种很好的另类资产放在整个组合中，这样既可以使投资组合充分分散化，又可以提高整个组合的收益。一般高频交易策略的资金承载量比较小，由于年化收益较高，即使做高频交易的资金占整个投资组合的比例比很小，也可以很大程度上提高整个投资组合的收益。

高频交易不仅对个人投资者和金融机构有很大好处，而且给整个金融市场带来了很多有益的方面：

（1）为市场提供流动性

高频交易的特征就是交易量大，所以高频交易的参与，能够使得那些不活跃的品种变得活跃起来，这样使得市场运行更加平稳，同时为投资者降低了摩擦成本。

（2）纠正市场错误定价

很多高频交易策略就是利用市场的短暂失效而研发出来的，所以高频交易的运行能促进市场价格迅速反映市场信息，市场价格在很短的时间内恢复正常，所以高频交易能够纠正市场的错误定价，从而起到了稳定整个市场体系的作用。

（3）促进计算机的创新

高频交易之所以能正常运行，离不开计算机技术的大力支撑。对每一个高频交易策略，首先需要编写成计算机语言，利用计算机检测交易信号，然后发单交易；同时计算机还帮我们收集日内的小时、分钟、秒等级别的数据，可以接收每一笔成交记录。大部分高频交易策略是靠速度取胜，也就是说为了胜利，需要争取千分之一秒的优势。所有的这些工作都离不开计算机，需要强大的计算机系统帮我们完

成。随着策略的需要，对计算机的要求会越来越高，所以高频交易的发展能反过来促进计算机的创新。

高频交易为投资者开辟了新的领域，使得人们可以进一步研究市场微观结构及其运行机制，据此研发出了各种各样的高频交易策略。高频交易现在已经涉及到众多的金融市场，比如股票、期货、外汇以及衍生品市场，在各个市场上都有比较成功的高频策略在运行。高频交易策略本身的属性是：适用于有足够流动性的金融工具。目前全球比较知名的高频交易公司主要包括DE Shaw、Worldquant和Millennium 等。这些公司为了取得信息传输上的优势，主要设于伦敦、芝加哥、新加坡等交易所附近。大多数的高频交易公司是对冲基金公司。目前市场上存在的高频交易策略主要可以分为4 类[1]：自动提供流动性(automated liquidity provision)、事件交易(event trading)、市场微观结构交易(market microstructure trading)、偏差套利(deviations arbitrage)。表 1.1简要介绍了4种策略的主要性质。

表 1.1: 高频交易策略分类

策略	描述	典型持仓时间
自动提供流动性	利用量化算法优化做市头寸的报价和执行	小于1分钟
市场微观结构交易	通过对观测到的报价进行逆向工程解析	小于10 分钟
事件交易	利用宏观事件进行的短期交易	小于1 小时
偏差套利	对价格均衡的偏差进行统计套利：例如三角交易、差交易等	小于1天

1.1.2 高频交易的成功典范

目前做高频交易的人越来越多，其中吉姆·西蒙斯(Jim Sions)教授是高频交易的一个成功典范。他既是一位声名显赫的数学家，又是一个杰出对冲基金经理人。他被称为整个量化投资的鼻祖级人物，被《金融时报》评为地球上最聪明的亿万富翁。长期以来，西蒙斯教授是高频交易的拥护者和实践者。他创立了文艺复兴科技公司(Renaissance Technologies Corp)，又在1988 年成立大奖章基金，仅在2008年

一年时间，西蒙斯教授就盈利25亿美元。有资料显示：1988年以来大奖章基金的年均复合回报率高达35%，这比索罗斯等投资大师同期的回报率高出10%，比同期标准普尔500指数的复合回报率高出20%，比股神巴菲特近20年的复合回报率高出18%。大奖章优异的表现，使得它成为收费最高的对冲基金，大奖章基金向投资者收取5%的资产管理费加上44%的投资收益分成。总而言之，西蒙斯的成功主要得益于他的一套量化投资模型，被相关学者称为“壁虎式交易”——一种非常短线的高频交易。

1.1.3 高频交易带来的挑战

高频交易给投资者带来了很大的机遇，同时也带来了一系列前所未有的挑战。首先，高频交易给投资者带来的挑战就是高频数据。由于金融市场连续和频繁的交易，导致数据更新频率极高，高频数据因此而得名。学术界所说的高频数据主要有两种形式：一种就是按时间等间隔记录的数据，是最早出现的高频数据形式，每条记录都是以分钟为基本时间单位，一般被称为分钟数据。另外一种是按每笔交易逐笔记录的数据，称为逐笔数据，一般以秒为单位，有的品种一秒能接到几个这样的数据。学者为了强调这种极高的频率的逐笔交易数据，并和分钟数据加以区别，称之为“超高频数据网”，这种数据形式是目前比较常用的数据。在本文研究中，分钟数据和逐笔数据都统称为高频数据。

大部分高频交易策略就是通过分析日内的数据而研发的，所以数据是基础。在交易时段每时每刻接收、处理、分析大量的数据是一件极富有挑战性的工作。相对于低频交易而言，高频交易的日内数据繁多，并且数据点的时间间隔也不同，一般情况下我们一秒接收1-2根tick数据，有的品种比如恒生股指期货一秒能接收到10几根tick数据。采用传统的方法每天都处理大量的数据是不行的，所以必须开发出新的能够快速接收、处理高频数据处理的工具。一般做过高频交易的投资者都了解，要开发出一个高频交易策略，首先要对历史数据进行回溯，观察回溯的结果是否符合策略的思路。一般情况下，需要长度在两年以上的数据进行回溯测试，两年的回溯数据几乎是一个庞大的数据库（相比低频而言）。如果是从数据服务商那里购买数据就可以减少一点压力，否则需要编写接收数据的程序，下载高频数据，这是一个既费时又繁琐的工作。拿到高频数据之后，根据策略对数据级别的要求，需要合成数据，经过合成处理后的数据才能用来进行回溯测试。所以在数据这方面需要不

断地更新技术，提高数据的质量，为策略的回溯提供保障。

高频交易的另一个挑战就是交易信号的精确性。我们希望能当计算机检测出交易信号时，能够立刻建立起头寸，如果出现了信号延迟问题，已经看到的盈利可能瞬间变成亏损。我们发现在很多时候，应该能止盈出来的一笔交易，最终以亏损结束头寸。所以这就需要在接收数据质量上和计算机运算效率上加以改进，确保每个信号的精准性。此外，执行的速度也是高频交易遇到的一个巨大挑战，它和信号的精准性同样重要。有的高频交易策略是靠速度取胜，需要争取千分之一秒的优势才能获利出场。所以高频交易不能采用人工下单的方式，而是靠计算机自动产生下单指令并自动执行，在这方面非常考验策略程序的编写人员的能力，要求编程人员在软件开发方面具有高超的技能，从而确保计算机运算的速度和效率。

1.1.4 高频交易存在的风险

在高频交易的过程中，高效的风险管理同样非常重要。高频交易能够获得稳定收益的保障是：设计一份合理的并且能严格执行的风险管理方案。2010年5月美国证券市场发生的闪电崩盘激发了公众、学术界与业界对高频交易的广泛探讨，SEC联合美国商品期货交易委员会(CFTC)对此专门成立调查小组，目的是考察其与证券市场结构的可能关联；美国国会为此还举行了一系列听证会，以期揭露可能的监管失误等。2012年BATS、骑士资本接连出现技术失误并且扰乱了市场，以及2013年4月美联社Twitter帐号风波，则进一步呈现了电子化交易的复杂性及脆弱性，发生的这些问题都加剧了公众对高频交易担忧和质疑。在中国，2013年8月16日光大证券程序化交易系统出现失控，这使国内得投资者也亲身经历一次交易技术失误所带来的代价[2]。

由以上这些例子可知，高频交易存在巨大的风险，非常必要建立一套完整的风险管理体系。在高频交易中一个微小的差错就可能带来巨大的损失，所以风险管理的首要目的就是控制潜在损失。风险管理中最重要工作就是风险的度量，依据风险的类型采用不同的度量的方法。表 1.2 描述了会对金融资产造成影响的几种风险：下面具体说一下操作风险，它是高频交易中出现比较频繁的一种风险。高频交易涉及硬件、软件及网络通信等多个基础构件，来自于硬件损坏、软件错误、通信中断、断电和黑客攻击等的不利影响是最直接的风险，通常将此类风险统称为操作风险。Sornnette and Von der Becke(2011)[3]首次定量分析了操作风险的特征，

表 1.2: 各种风险类型

风险类型	描述
市场风险	由市场价格变动所导致的风险
信用及交易对手风险	对手没有履行其责任的能力和意愿
流动性风险	交易过程中迅速结束头寸的能力
操作风险	日常操作失误造成的风险
法律风险	诉讼费用支出的风险

发现其服从一个尾部指数为0.7 的幂律分布，这意味着操作风险在未来出现的次数将越来越多，其影响也将会变得越来越大。事实上，因为操作风险市场上发生了很多大的异动，像前述BATS与骑士资本的交易失误，以及我国的光大证券“8.16事件”。

目前度量风险的方法主要有四大类：统计模型、数量模型、情景分析、因果模型，表 1.3对其进行了详细描述。

1.1.5 高频数据的统计特征

高频数据是高频策略的基础，所以研究高频数据本身的统计特征是非常重要的。大量的实证研究表明金融资产收益率呈现尖峰厚尾等非正态特性，符合稳定分布的特征。稳定分布由levy 在二十年代提出来的，具有良好的性质，并且对金融市场的拟合程度比较高，因此一经提出，就在金融市场得到广泛的应用。早有学者的研究表明，与低频金融数据(如月度数据)服从高斯分布的特征不同，金融高频数据是不稳定的，在较短期间内有着明显的厚尾趋势(heavy tail)。

表 1.3: 度量风险的主要方法

类型	描述
统计模型	统计模型是根据历史数据对未来可能发生的最坏情况进行预测,最常见的风险度量工具是风险价值度(VaR)。
数量模型	数量模型是把可预见的最大损失表示成一些参数的百分比形式,比如收入、营业成本,一般数量模型通常用来估计操作风险。
情景分析	情景分析用于划分关键风险指标一般、最好、最坏的情况,这种指标可以量化所有类型的风险,情景分析通常也称“压力测试”,也常常用在投资组合中,观察不同的进场时间,对整个组合的影响。
因果模型	因果模型是找出潜在损失的前因后果,可以构建一个整合各种驱动因素的动态模拟模型用来度量和和管理信用风险,以及操作风险和法律风险。

1.2 文献综述

目前国内外学者对高频交易和稳定分布已经进行了深入全面的研究,并为我们提供了大量有用的参考文献。下面两节主要综述了国内外对高频交易和稳定分布的研究。

1.2.1 关于高频交易及稳定分布的国外研究情况

最近几年许多国外学者对高频交易的理论作了大量研究,大大地促进了高频交易的发展。Cvitanic和Kirilenko(2010) [4] 首次给出高频理论分析的模型。Chaboud(2011) 等[5] 发现高频交易与算法交易、程序化交易等有较大的重叠,所以认为高频交易是算法交易的一种特殊情形。Brogaard[6] 在2011 年研究发现由于信息技术和计算机的发展,使得高频交易成为美国证券交易的主要部分;截止到2010 年,高频交易量几乎占美国证券交易量的70%。Gomber和Haferkorn(2012)[7] 认为高频交易是一种通过设计复杂算法实现快速下单、撤单而获利的交易方式;高频

交易的交易标的是高流动性的资产，通过自有资金进行快速的买卖头寸。Scholtus (2012) 等[8] 在对高频交易的速度与绩效的关系进行了实证研究时发现高频交易的速度对交易策略的回报产生了显著的影响，所以高频交易的速度很重要。Menkveld(2012)[9] 的分析显示大量高频交易的出现、新市场的产生与买卖价差下降50% 同时发生；并得出以下结论：高频交易量占有所有交易的14.4%，80%的高频交易策略是被动的，高频交易速度非常快，每一笔交易仅获得很少的收益。Easley(2012) 等[10]在研究基于订单驱动市场的知情交易概率的测度模型(VPIN) 在高频交易中的作用中发现：源于毒性对流动性供给产生的作用，高水平VPIN 会带来价格大幅波动的风险；而开发依据毒性来改变指令交易方式的算法将会缓解这一风险。所以交易所可以将VPIN 方法用于流动性准备源中，在流动性遭受打击下降时另一方会加速流动性供给，这样就可以保持做市商的活性。Gomber and Haferkorn(2013)[7] 的分析可以将高频交易的主要策略分为三大类：流动性提供即做市商交易，统计套利(Statistical Arbitrage)，盘口交易(Trading the Tape)。

关于高频交易的研究大部分学者认为高频交易对市场和交易者是比较有利的，他们大致都得出这样的结论：高频交易能够增加市场的流动性、减少短期波动以及能给交易者带来丰厚回报等。2010年，Brogaard[11] 研究认为高频交易者避免了和具有内部消息的交易商交易，并且为市场提供流动性。Aldridge[12] 在2010年也认为高频交易在增加市场流动性、提升市场效率、稳定市场以及信息技术创新等方面都有着正面、积极的作用。Hasbrouck 和Saar(2012)[13] 研究表明交易商把交易主机托管具有明显的时间优势；同样场内交易商比场外交易商更具备时间优势。所以现在很多交易上把服务器部署在交易所附近是为了争取时间优势。Hanson (2011) [14]发现在一个模拟的市场环境，高频交易者的参与提高了市场流动性和效率。Esipov(2012)[15] 认为利用高频价格波动和自动交易获利的市场参与者和做市商同样决定了市场是否有效率，这是从市场效率的角度研究了高频交易对市场质量的影响。Angel 和McCabe(2013)[16] 认为在过去的十年高频交易明显提升了美国的市场深度。Brogaard(2010)[11] 研究还认为2008 年的卖空禁令很大程度上限制了高频交易的参与，使得日内波动性明显提升；他还发现高频交易有助于发现市场的价格。Chaboud(2011) 等[5] 研究了从2003 年到2007 年算法交易在外汇市场中的作用。他们用三组外汇高频交易数据，发现高频交易下外汇交易盈利的方差比手工交易下要小，说明人仍旧是一个会利用内部消息的交易者。Foucault(2011) 等[17]认为投资者在高频交易中能抓住投资机会并从交易中获利。

当然也有的学者认为高频交易对市场有负面影响。Zhang(2010)[18] 分析了在不同时期的市场中高频交易对股市波动的长期作用,认为高频交易提高了股市的长期波动性;他还认为如果市场中存在较多的高频交易者就会造成基本面消息对股价产生过度的响应,所以高频交易干扰市场的价格发现过程。Grillet-Aubert(2010)[19] 在研究中指出,由于高频交易者可以在10 微秒内撤销交易指令,这使得指令的撤销率变得极高。Robert(2011)等[20] 认为高频交易能够扰乱市场。Kirilenko等(2011)[21] 通过研究2010 年5 月美国股市的“闪跌”认为高频交易增加了股市的短期波动性。Dichev等(2011)[22] 研究发现高频交易大幅增加的交易量增强市场波动性。Boehmer等(2013)[23] 对全球39 个交易所数据进行了分析之后得出:算法交易与高频交易能够系统性地增强市场的短期波动。Martinez and Rosu(2013)[24] 的研究也表明高频交易明显提高市场短期波动性。

对高频数据的分析也成为很多领域的热点。首先高频数据本身具有时间间隔不等、取值离散等方面的问题,如何处理这种数据对于很多传统分析方法来说是一个挑战。再者高频数据中的波动性能够反映市场微观结构的重要信息,有利于投资者理解市场的微观结构。所以金融高频数据的出现使得各种基础理论和计量模型得到创新和发展,并且方便了人们对市场微观结构的研究。实证研究表明金融资产收益率呈现出尖峰、厚尾等非正态特性,这与稳定分布的特征一致。稳定分布是由列维在二十年代提出来的,一经提出就在金融市场得到广泛的应用;因为它具有良好的性质,并且对金融市场的拟合程度比较高。Mandelbrot(1968) [25]等提出了著名的分形理论,并指出金融资产呈现出尖峰、厚尾等非正态特征。Peters(1994)[26] 论述了美国股票收益率具有稳态的特性。Hurst(1999) 等[27] 研究了对数稳定分布下的期权定价公式。Khindanova(2001)[28] 等讨论了稳定分布的VaR 模型。2003 年,Jacquier 等[29]的研究发现SP500 指数的日收益数据具有非正态性。

1.2.2 关于高频交易及稳定分布的国内研究情况

国内对高频交易策略的研究起步比较晚,朱建平、来升强(2009)[30] 在套利定价理论的框架下,对基于市场中性的微观交易策略的含义和本质进行了深入的探讨。镇磊(2010)[31] 在《基于高频数据处理方法对A股算法交易优化决策的量化分析研究》一文中,介绍了高频交易的兴起和发展现状,以及当前高频交易的常用高频和主流设计思想,并通过对股票市场的发展分析,对在股票市场进行高频交

易所需要注意的问题进行分析总结。张彦(2011)[32] 对CSI300 股指期货与现货间关系进行了研究,使用广义自回归异方差模型(GARCH)模型分析了期货的推出给现货带来的影响,证明二者之间存在强相关性,但不存在相互决定关系。胡天福,何晓彬(2012)[33] 的高频交易在中国证券市场的应用研究,采用趋势跟踪策略对股指期货进行模拟。曾勇(2012)[34] 等基于上证50与深证成指指数成分股的高频交易数据,实证考察了阻力线(支撑线)与市场流动性指标之间的关系。任品和苟开桂(2013)[35]对目前国内期货市场的微观结构进行了相关研究,更好地认识了市场微观构件如何影响市场价格变动趋势,有助于我们评价期货市场质量和效率。包思(2013)[36] 通过平稳化MACD 指标,并将其应用于高频交易中,得到了较好的收益。徐冉、戴伟辉(2013)[37] 提出了基于TDF (Theory-Data-Feedback)框架的建模分析新方法对沪深300 股指交易指数收益率进行了实证分析,取得了很好的效果。陈炜、孙梦荣(2013)[38] 等以投资组合相关理论为基础,构建高频交易策略组合配置模型,利用人工蜂群方法对模型进行了实证研究。张涤新和李超(2013)[39] 研究了我国股指期货与股票组合的高频套利,证明了中国国内市场具有较好的套利机会。左顺根(2013)[40] 将理论与实证相结合,使用G-S、Geweke 线性反馈、DSEM 等模型从股指期货价格发现、套期保值、股票与股指期货市场价格波动等多个方面对期货现货市场进行了研究和总结。在价格预测方面,曹柯(2013)[41]等人提出了经验模态分解和相关向量机的方法对期货价格进行预测,并证明该方法能取得较好的效果。燕汝贞等(2014)[42] 通过优化交易算法,构建交易策略模型,能为投资者减少交易成本。韩芳婕(2014) [43] 等用历史数据进行回溯检验,结果表明:基于过滤理念的高频交易策略能够显著提高策略的夏普比率,能有效的控制交易风险。同时实证结果表明对于基本的日内策略,加入复合止损条件可以更加有效的降低风险,降低策略的最大亏损。陈贻娟(2014)[44] 等利用列联表检验及独立性检验,发现量化交易策略与价格波动之间极有可能存在着一定的非独立的关系。梁满发、潘晔(2014)[45] 提出通过Kelly策略优化单一高频策略模型的收益,并且建立基于高频交易的投资组合策略。田波平等(2015)[46] 在基于高频交易策略的BIAS技术指标统计分析一文中,就国内高频交易的现状通过技术分析的手段对高频交易策略进行研究,并利用我国股票市场的高频交易数据对策略进行实证检验。陈艳、王宣承(2015)[47] 就我国金融衍生品市场的现状提出了基于LASSO 变量选择方法和遗传网络规划的期货高频交易策略。

国内有关稳定分布及金融数据统计特征的文献也有很多,徐剑刚和唐国

兴(1997) [48] 可能是我国最早的一篇关于我国股票市场波动性GARCH-M 建模的文章; 此后, 严太华(2000) [49]、杜海涛(2000) [50]、唐齐鸣和陈健(2001) [51]、温素彬(2002) [52]、陈守东和俞世典(2002) [53]、金洪飞(2004) [54]、王佳妮和李文浩(2005) [55] 等都对我国证券市场建立了时间序列模型并在实证分析中取得了不错的效果; 徐绪松(2006)[56] 等在非正态稳定分布条件下的投资组合模型: 均值- 尺度参数模型一文中提出在非正态稳定分布条件下研究投资组合模型, 通过拟合优度检验发现我国的股票收益率与非正态稳定分布的拟合效果非常好; 研究了非正态稳定分布条件下投资组合收益和风险的度量, 建立了均值- 尺度参数投资组合模型; 通过实证分析发现均值- 尺度参数模型能够解释资产配置之谜。艾克凤(2006)[57] 采用Mantegna 和Stanley (1995)提出的方法, 得到上证综指收益率的特征指数估计为1.4837, 我国股票市场收益率不服从正态分布, 收益率分布呈现尖峰厚尾特征。钟蓉萨(1999) [58]研究了沪、深市场部分股票收益率的各阶矩的特征。龚锐、陈仲常和杨栋锐(2005) [59]、马丹(2004) [60]假定误差序列服从t 分布、GDE分布等非正态分布建立了我国股市的GACRH 类模型, 取得了满意的实证分析结果。武东等(2006)[61]、[62]以及周孝华和王小庆(2009) [63] 基于稳定分布进行了股票收益率的特征分析并应用于风险值VaR 的计算, 验证了稳定分布的可靠性和可行性; 稳定分布的分布指数很好地刻画的分布的厚尾性。李志慧和卢新生(2011) [64] 研究了玉米等中国农产品期货的多重分形性及其成因。史文静和高岩(2014) [65] 研究了中国股指期货价格的多重分形性及其成因。陈伟娜 (2015) [66]引入稳定分布, 用向前违约概率的方法, 将多重积分转化为递推的单重积分, 从而为CDO定价建立一个兼具精确性和实用性的动态稳定因子Copula模型, 并且适用于基于同一信用组合的不同期限的CDO 定价。

1.3 本文基本框架、研究方法与创新之处

1.3.1 本文基本框架

本文在稳定分布理论的基础上, 介绍了稳定分布的性质、参数估计以及数值计算方法。用稳定分布对指数以及期货高频数据分别进行拟合分析, 并利用稳定分布特征值 α 的性质对股指和国内指数高频数据进行了实证分析, 从金融数据的本质特

征上分析了高频交易策略盈利的原因。

第一章主要从高频交易的魅力、成功典范、存在的风险以及高频数据的统计特征等方面介绍了写作的背景及意义，阐述了国内外对高频交易和稳定分布的研究情况，并阐述了本文的研究方法及创新之处。第二章介绍了有效市场理论和高频数据的游程检验，本章具体介绍了几种常见的检验方法，并对商品期货高频数据进行了检验分析。第三章主要介绍了稳定分布的理论，从稳定分布的定义、几种特殊情况、性质以及参数估计和数据算法进行了具体介绍，最后分别对指数数据和股指高频数据进行了拟合分析。第四章主要介绍了根据稳定分布特征值 α 的性质构造新的数据序列，通过分析 α 的大小，反向推出不同 α 的投资比例，运用高频交易策略进行组合分析。第五章主要介绍了本文研究的结论：数据的稳定分布的 α 越小，运用到策略上盈利的可能性越大。最后阐述了本文的缺陷和不足之处，并提出了未来的研究方向。

1.3.2 本文研究的方法及创新之处

我们知道 α ($0 < \alpha \leq 2$) 为稳定分布的特征指数 (Characteristic exponent)，描述尖峰厚尾程度， α 越小，峰越尖，尾越厚且越长；反之，随着 α 增大，分布的尾将变薄。本文主要通过线性构造的方法构造出不同 α 的数据序列，然后把该序列运用到高频交易策略上，观察分析不同 α 的数据序列的投资效果。希望得出这样的结论： α 越小，投资效果越好。

高频交易具有高盈利性，那么它盈利的依据是什么？市场的短暂失效。所以要想知道它盈利的内在逻辑，还要在高频数据上加以研究和探索。本文的创新之处在于把稳定分布用在了高频数据的研究上，从高频数据的分布上阐述高频交易盈利的本质。

第二章 有效市场假说与高频数据

有效市场假说可以追溯到20世纪初，法国学者路易·巴舍利耶（Louis Bachelier）（1900）在其博士论文《投机理论》中最早使用统计方法分析金融资产的收益率序列，认识到资产价格的变化过程是布朗运动。他同时提出投机市场价格行为的基本原理：投机应该是公平游戏，其期望利润应该等于零。我们从现代随机过程理论可以知道，由Bachelier提出的投机基本原理其实就是鞅过程：给定目前信息的条件下，如果所有的信息瞬间体现在证券价格中，那么证券价格未来的期望值永远等于当前价格本身。1964年巴舍利耶的论文被库特（Paul H.Cootner）收录到《股票市场价格的随机性》（The Random Character of Stock Market Prices）一书中。该书一出版，人们对股票市场价格行为的探究就达到一个高潮，该书也为尤金·法马（Eugene Fama）在1965年所提出的有效市场假说提供了理论基础。在2013年，为了表彰学者们对“资产价格的实证分析”所作出的卓越贡献，尤金·法马(Eugene Fama)与皮特·汉森(Peter Hansen) 以及罗伯特·J·席勒(Robert J.Shiller)共同获得了诺贝尔经济学奖。

本章主要介绍有关市场有效假说的发展，以及基于有效市场假说中弱有效性对高频数据的检验，期望得到高频数据在什么频率上是“可测的”（即不是随机的），这样就可以在该频率上建立相应的统计套利模型。

2.1 有效市场假说

尤金·法马(Eugene Fama)（1965）[67]在前人研究的基础首次正式提出了有效市场的概念：在一个资产市场上，如果资产价格能够完全充分地反映了所有可获得的信息，那么这个市场就称为有效市场。所以我们可以理解为在一个有效的市场

上，价格的变动依赖于与资产有关的所有新信息，但是新信息的发生是不可以预测的，即不可以真正预测资产的价格。有效市场理论说明了市场价格变化的过程是一个鞅过程，这是有效市场假说在经济学上的意义所在。尤金·法马(Eugene Fama)(1970) [68]提出有效市场的三个层次：弱有效市场、半强有效市场以及强有效市场，它们分别对应于三个不同的信息集合，具体如下：

1. 弱式有效市场

弱有效市场对应的信息集是指包括资产历史价格、收益率、交易量、短期利率等与资产相关的所有历史信息。对弱有效市场的检验主要是看能否通过资产的历史价格或收益率序列对未来收益率进行预测。

2. 半强式有效市场

半强有效市场的信息集除了与证券有关的所有信息以外，还包括所有已经公开的信息：公司的财务报表、管理层的管理水平、产品的特点、针对该行业的国家经济政策以及未来盈利预测等各种用于基本面分析的信息。

3. 强式有效市场

强有效市场是指资产价格包含了所有与之有关的信息，不仅包括历史信息和所有公开的信息，还包括一些只有公司内幕人员掌握的非公开的信息。

Malkiel(2003)[69]对有效市场重新定义：有效的市场就是与资产价格相关的信息能在价格中完全反映；严格地说来，金融市场对某个给定的信息集合是有效的才能说明金融市场是有效的，如果对市场上所有的参与者公开这些信息，资产价格则不受影响；金融市场对某个给定的信息集合是有效的，那么在市场投资者使用该信息集进行交易不可能获得任何利润，也就是说没有投资者可以获取超过风险调整后的平均收益。

在强有效市场中，证券价格一旦偏离其价值，市场参与者就会通过迅速的买卖行为消除这种差异。从这一角度来看，高频交易无疑会提高市场的效率。当市场出现微小的差价，就会给投资者无风险套利的空间，市场参与者利用高频交易的方法获取这种差价，这样就会迅速消除微小的差价。高频交易在提高市场效率的同时，也增加了金融市场的流动性。高度竞争的流动性会促使价格与价值趋于一致，从而增加了市场的有效性。

在有效市场中，证券有关的所有的可用信息会立即反应在其价格上。如果出现信息慢慢整合到证券价格中上这种情况，就会出现套利机会，此时就可以认为该市场是无效的。图 2.1展示了有效市场和无效市场的区别，寻找无效市场的过程也就是在寻找具有套利机会的市场。换句话说，部分套利机会就是在市场无效的时候发生的。图 2.1(a)展示了对“利好消息”有效市场和无效市场推动证券价格上涨所

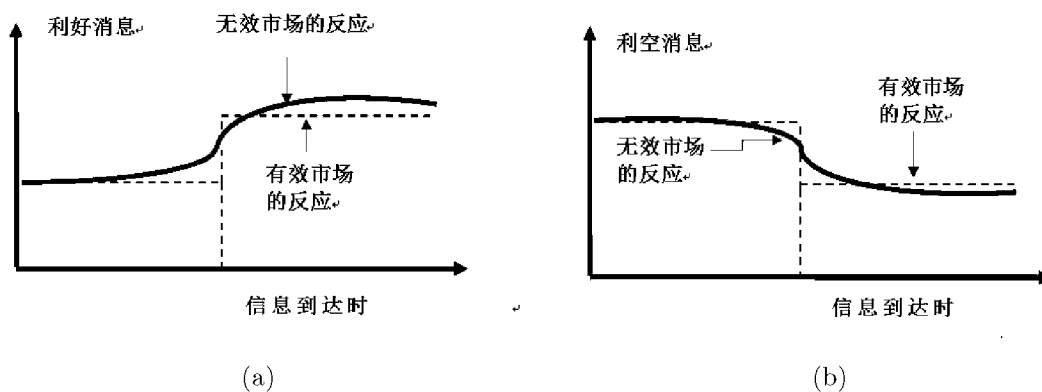


图 2.1: 有效市场和无效市场下市场对信息的反映

发生的变化过程，图 2.1(b) 展示了“利空消息”有效市场和无效市场推动证券价格下跌的变化过程。其中，实线表示的是无效市场的反应，虚线表示的是有效市场的反应。在有效市场中，当信息出现时，证券价格会立即调整到新的水平；在无效市场中，由于信息泄露等原因，该证券会在信息还未公开之前就开始调整其价格，所以信息公开后，价格通常还会有短暂的过度反应过程[1]。

金融市场上价格容易出现过度反应，有利好消息时，价格通常先是过度上涨然后再回落；有利空消息时，通常是过度下跌随后再反弹。这种现象在市场的微观结构的研究中已经被证明是经常出现的，但持续时间不会太久。有效市场假说总体来讲是成立的，在市场上不会长时间存在大的套利机会。但是小的短暂的套利机会还是经常存在的，也是转瞬即逝的，而这种瞬间的套利机会需要高频的交易方法去捕捉才能获利。

2.2 高频数据潜在获利机会及其游程检验

高频交易存在巨大的潜在盈利，是因为通过高频数据可以捕捉到很小的获利机会。高频交易机会具有很广的覆盖范围，比如做市商利用秒级的价格运动获利；交易者根据市场微观结构理论预测分钟级别的价格趋势进行交易从而获得利润；交易者通过统计几个小时级别的市场运动规律开发出统计套利策略，从而赚取价格偏离统计值的利润。高频交易策略的盈利能力主要受到交易频率的限制。比如在以日为级别的交易，最大收益和亏损不会超出当日的价格波动幅度；在以小时级别的频率下，价格运动的幅度会缩小，最大收益和亏损不会超出当个小时的价格波动幅度。如果是小时级别数据，一天的潜在收益等于一天之中所有小时价格变动幅度的加和。如果是在更高的频率下，价格运动的幅度随着数据频率的提高而进一步缩小，从而有更多的盈利机会。具体一个高频交易策略适合哪个频率的数据，一个标的在什么频率下适合做高频交易？这都需要分析高频数据存在的潜在获利机会以及在什么频率下是“可测的”。与当代其他有关市场有效性的文献一样，本文只对弱式有效市场进行检验。

2.2.1 高频数据潜在获利机会

由于标的所处的市场不同，其交易者的习惯也不同，以及其他原因，使得交易标的流动性不同，从而导致每天的波动情况不同，甚至每时每刻的波动也不同。有些交易标的比较活跃，在很短的时间就可波动很大的幅度，理论上这种交易标的存在巨大的获利潜力。在每种频率下，用价格变动“幅度和”来计算最大收益。通过开盘价进行标准化，从而得到百分比的相对收益。每种频率的最大潜在收益则是该频率下每个时间区间的价格变动幅度总和。表 2.1展示了2016年2月26日恒生股指期货在不同频率下的最大潜在收益以及其他一些有关的统计量。

从表 2.1中可以看到，高频之下的潜在收益是巨大的。10秒级别的每日最大潜在收益是每日的127倍。像这样小的、短的套利机会是经常有的，不过转瞬即逝，这些需要高频的交易方法去捕捉。当然潜在亏损也很大，所以要想高频交易发挥出最大的收益潜能，需要交易者精心设计高频交易策略、进行大量的历史回溯以及严谨的风险管理方法。

表 2.1: 2016年2月26日恒生股指期货在不同频率下最大潜在收益

	“10秒”	“1分钟”	“10分钟”	“1天”
每日最大潜在收益	34.23%	27.70%	9.12%	0.27%
日内周期数	4590	765	75	1

开发高频交易策略的关键是在高频数据中发现重复出现且能够获利的交易机会。交易策略的盈利能力受到所选择的品种以及交易频率的限制。至于什么频率下才会有套利机会，这些是需要对大量的数据进行挖掘。首先好的高频策略能够发出稳定收益的信号。由于高频机会每次的获利不大，所以需要足够多的交易机会才能提高整体收益，而交易机会的多寡，可以用市场有效程度来衡量。以下两节主要采用游程检验的方法来检验不同频率数据是“可测的”还是“随机的”。

2.2.2 游程检验

市场越是无效，存在的可预测性的交易机会也就越多。本节主要研究在什么频率下，高频交易才是最有利可图的。如果价格变动是随机的，那么想找到可稳定获利的交易机会的可能性微乎其微。反之，如果价格变化不是“随机的”（即是“可测的”），那么该标的就具有持续的可预测性，值得投资者进行更加深入的研究分析。对市场效率进行检验可以帮助我们发现市场在多大程度上存在可预测的交易机会。经过学者的多年研究，现在已经有很多种检验市场有效性的方法，比如基于自回归的检验、基于鞅假设的市场有效性检验、基于协整性的市场有效性检验、非参数游程检验以及随机游动检验。本节选取两种检验方法，非参数游程检验和随机游动检验，以帮助投资者挑选出最具有盈利性的投资标的。

非参数游程检验方法

非参数游程检验方法由Louis Bachelier在1900年提出，并且是最早的一种对市场有效性进行检验的方法。该方法主要对一系列连续正的或负的价格变动所出现的概率进行度量。给定一个价格变动频率，多个连续的具有相同方向的价格变动就提供了一个交易机会，前提是交易所获得的盈利能够覆盖交易成本。对给定频率下的价格运动样本数据进行分析，首先需要记录严格同方向价格运动序列的个数。比如

给定的频率是“1分钟”，就要记录1分钟间隔下价格方向变动严格为正或为负的序列；如果给定的频率是1笔交易，那么所记录的游程就是每个交易价格方向变动严格为正或为负的一个序列。下面以恒生股指期货2016年2月26日1分钟频率级别的价格变动为例，进行非参数游程检验过程如下：

1. 对样本数据标记方法：

给定的数据频率是1分钟，价格变动计算方法是：如果这1分钟价格高于前一分钟的价格，则这1分钟的价格变动方向记为“+”，否则记为“-”。若连续为正，则记一次正游程（P）；价格连续为负，记一次负游程（N）。样本中所有记录的游程个数记为 μ ，价格变动为正的个数 n_1 ，价格变动为负的个数 n_2 。

2. 样本数据标记过程及统计结果：

表 2.2 展示的是2016年2月26日9:40到10:40恒生股指期货价格变动情况。在这1小时的统计结果中，共有27个正向价格变动，30个负向价格变动。16个正的游程价格上涨，16个负的游程价格下跌，即： $\mu=32$ ， $n_1=27$ ， $n_2=30$ 。

3. 假设检验：

如果价格变动是完全随机的，游程总数的期望值： $x = \frac{2n_1n_2}{n_1+n_2} + 1$ ，标准差为： $\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}$ ，构造的双边假设检验统计量： $Z = \frac{|\mu-x|-0.5}{\sigma}$ ， α 为显著性水平，如果在95%的置信度下，游程数目与均值的偏差大于1.645倍标准差，则认为该频率下的游动是非随机的，即可测的。所以：

如果 $|Z| \geq N^{-1}(\frac{\alpha}{2})$ ，拒绝原假设，即为可测的；

否则 $|Z| < N^{-1}(\frac{\alpha}{2})$ ，接收原假设，即为随机的。

4. 非参数游程检验结果：

表 2.3 展示的是在2015年2月26日到2016年2月26日恒生股指期货在置信度 $\alpha=95\%$ 的不同数据频率的游程结果，可以看出5秒、10秒、1分钟数据频率是可测的；10分钟、30分钟数据是随机的，根据检验原理可知高于30分钟数据频率的数据是随机的，所以在建立关于恒生股指期货高频统计套利模型时，数据频率尽量不要低于10分钟级别。

表 2.2: 2016年2月26日9:40到10:40恒生股指期货1分钟价格收盘价

日期	收盘价	1分钟价格变动单位	变动符号	1 游程
2016/02/26 9:40	19235	1	+	P_1
2016/02/26 9:41	19252	17	+	P_1
2016/02/26 9:42	19237	-15	-	N_1
2016/02/26 9:43	19221	-16	-	N_1
2016/02/26 9:44	19213	-8	-	N_1
2016/02/26 9:45	19191	-22	-	N_1
2016/02/26 9:46	19186	-5	-	N_1
2016/02/26 9:47	19175	-11	-	N_1
2016/02/26 9:48	19179	4	+	P_2
2016/02/26 9:49	19183	4	+	P_2
2016/02/26 9:50	19193	10	+	P_2
...
2016/02/26 10:30	19224	13	+	P_{13}
2016/02/26 10:31	19223	-1	-	N_{14}
2016/02/26 10:32	19218	-5	-	N_{14}
2016/02/26 10:33	19229	11	+	P_{14}
2016/02/26 10:34	19220	-9	-	N_{15}
2016/02/26 10:35	19220	0		
2016/02/26 10:36	19226	6	+	P_{15}
2016/02/26 10:37	19216	-10	-	N_{16}
2016/02/26 10:38	19218	2	+	P_{16}
2016/02/26 10:39	19219	1	+	P_{16}
2016/02/26 10:40	19222	3	+	P_{16}

表 2.3: 2015年2月26日到2016年2月26 日恒生股指期货非参数游程检验结果

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30 分钟”
边界值	1.960	1.960	1.960	1.960	1.960
统计量	14.640	11.632	4.883	-1.279	-0.401
游程检验结果	“可测的”	“可测的”	“可测的”	“随机的”	“随机的”

随机游动检验

Lo和MacKinlay在1988年[70]提出了一种检验价格走势是否服从随机游动的方法，即随机游动检验。该方法是一种更先进的检验市场有效性的方法。这种方法可以帮助投资者评估市场状态，将资金配置到有效性更差的市场上去，也就是获取利润机会更多的市场中去。该方法的原理是：如果在某一给定的频率下，比如1小时频率的价格变动是随机的，在更低的频率下（2小时，3 小时…）价格的变动应该也是随机的，并且1 小时价格变动的方差和2小时的有确定的关系。反之，则不成立，即如果2 小时频率数据是随机的，不能说明1小时数据频率是随机的，且1小时价格变动的方差和2小时的不存在某种特定的关系。

当价格随机变化时，一般价格是随机游动的。随机游动过程可以表示为：

$\Delta \ln P_t = \ln P_t - \ln P_{t-1} = \mu + \epsilon$ 其中 P_t 为 t 时刻的价格， μ 为漂移项， ϵ 为残差项。假设式子中残差项 ϵ_t 是一系列独立同分布正态分布，且均值为0，方差为 σ_0^2 的随机变量，即 $\epsilon \sim i.i.d.N(0, \sigma_0^2)$ 。下面是随机游动检验的理论及检验过程：

1. 检验基于的估计量：

$$\begin{aligned}
 \hat{\mu} &= \frac{1}{2n} \sum_{k=1}^{2n} (\ln P_k - \ln P_{k-1}) = \frac{1}{2n} (\ln P_{2n} - \ln P_0), \\
 \hat{\sigma}_a^2 &= \frac{1}{2n} \sum_{k=1}^{2n} (\ln P_k - \ln P_{k-1} - \hat{\mu})^2, \\
 \hat{\sigma}_b^2 &= \frac{1}{2n} \sum_{k=1}^n (\ln P_{2k} - \ln P_{2k-2} - 2\hat{\mu})^2.
 \end{aligned} \tag{2.2.1}$$

Lo和MacKinlay [70]在文章中证明了参数 σ_0^2 , σ_a^2 , σ_b^2 之间渐进服从如下分布:

$$\begin{aligned}\sqrt{2n}(\hat{\sigma}_a^2 - \sigma_0^2) &\stackrel{a}{\sim} N(0, 2\sigma_0^4), \\ \sqrt{2n}(\hat{\sigma}_b^2 - \sigma_0^2) &\stackrel{a}{\sim} N(0, 2\sigma_0^4).\end{aligned}\quad (2.2.2)$$

2. 构造检验统计量:

$$J_r \equiv \frac{\sigma_b^2}{\sigma_a^2}, \sqrt{2n}J_r \stackrel{a}{\sim} N(0, 2) \quad (2.2.3)$$

进行双边检验, $J = \sqrt{2n}J_r, \alpha$ 为显著性水平:

如果 $|J| \geq N^{-1}(\frac{\alpha}{2})$, 拒绝原假设, 即为可测的;

如果 $|J| < N^{-1}(\frac{\alpha}{2})$, 接受原假设, 即为随机的。

3. 检验结果: 同样选取在2015年2月26日到2016年2月26日恒生股指期货, 置信度 $\alpha = 95\%$, 表 2.4展示了不同数据频率的游程结果。与非参数游程检验不同的是, 随机游动检验1分钟数据是随机的。

表 2.4: 2015年2月26日到2016年2月26 日恒生股指期货随机游动检验结果

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30 分钟”
边界值	3.920	3.920	3.920	3.920	3.920
统计量	-49.710	-23.010	-6.772	0.623	0.556
游程检验结果	“可测的”	“可测的”	“可测的”	“随机的”	“随机的”

其他市场检验

用两种方法对恒生股指期货进行有效性检验, 从得出的结果看出5秒、1分钟数据都是可测的。非参数游程检验和随机游动检验的结果有所不同, 因为随机游动的检验方法的理论根据更强, 结果会更加可靠, 所以本文更偏重随机游动的检验方法。为了寻找更多具有可测性的品种, 本小节用随机游动检验方法对我国的沪深300股指期货和商品期货进行检验。时间选取2015年2月26日到2016年2月26日, 置信度 $\alpha = 95\%$ 。表 2.5、表 2.6、表 2.7、表 2.8、表 2.9、表 2.10分别是沪深300股指期货、甲醇、PTA、橡胶、螺纹钢及白糖的结果。可以看出5秒、10秒数据是可

测的，1小时数据是随机的，根据随机游程检验的原理可知低于1小时频率的数据是“随机的”。所以在构建高频策略时要考虑到标的适用频率，这样才能找到更多的套利机会，获得更多的利润。

表 2.5: 沪深300股指期货

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30分钟”	“1 小时”
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	5.479	36.708	-2.064	-1.908	-1.969	-1.921
游程检验结果	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”	“随机的”

表 2.6: 甲醇（MA）

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30分钟”	“1 小时”
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	-200.971	-21.056	0.829	2.487	-0.560	2.533
游程检验结果	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”	“随机的”

表 2.7: PTA（TA）

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30分钟”	“1 小时”
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	-209.598	-27.534	-2.706	-2.927	-1.429	-1.051
游程检验结果	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”	“随机的”

表 2.8: 橡胶 (ru)

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30分钟”	“1 小时”
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	-116.036	-14.579	-3.602	1.529	-3.090	-1.065
游程检验结果	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”	“随机的”

表 2.9: 螺纹钢 (rb)

数据频率	“5秒”	“10秒”	“1分钟”	“10分钟”	“30分钟”	“1 小时”
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	-351.40	-45.19	-8.57	1.07	1.72	-1.82
游程检验结果	“可测的”	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”

表 2.10: 白糖 (SR)

数据频率	5秒	10秒	1分钟	10分钟	30 分钟	1小时
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	-97.19	-9.56	-9.36	-1.78	0.13	-0.70
游程检验结果	“可测的”	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”

第三章 稳定分布的理论介绍

稳定分布又称分形分布，是由经济学家帕雷托(Pareto)(1897) 在研究收入分布时首次发现的；在20世纪20年代，法国数学家列维（Levy）首次提出并研究稳定分布的数学性质，因此稳定分布也被称为帕雷托分布、稳定帕雷托分布和帕雷托——列维分布。从30 年代以来学者们对稳定分布的性质进行了深入的研究。Mandelbort （1968）提出分形理论，该理论在金融和其他科技领域应用广泛。Peters （1991）[71] 研究发现无论是发达国家或发展中国家，资本市场均呈现出分形特性。由于稳定分布具有卷积不变性，通过参数调整能够很好地适应尖峰、厚尾的分布情况，所以在金融领域的研究中可以用稳定分布进行建模。

3.1 稳定分布的定义

稳定分布的定义有很种，下面我们只简单介绍最常见的三种定义。

定义1：称随机变量 X 服从稳定分布，如果存在一个与随机变量 X 独立同分布的随机变量序列 Y_1, Y_2, \dots, Y_n ，对于任意的 $n \geq 2$ ，存在常数 $c > 0$ 和 $d \in R$ ，使得 $Y_1 + Y_2 + \dots + Y_n \xrightarrow{f} cX + d$. 其中符号 \xrightarrow{f} 表示依分布收敛。若 $d = 0$ ，则称 X 服从严格稳定分布。

定义2：称随机变量 X 服从稳定分布，若对于任意的常数 $a > 0$ 和 $b > 0$ ，存在着一个常数 $c > 0$ 和 $d \in R$ ，使得： $aX_1 + bX_2 \stackrel{d}{=} cX + d$ 。这里 X_1 和 X_2 是与 X 独立同分布的随机变量，其中 $\stackrel{d}{=}$ 表示同分布。若 $d = 0$ ，则 X 服从严格稳定分布。

除了高斯分布、柯西分布、levy 分布，稳定分布没有解析的密度函数和分布函数，但是其特征函数有解析形式，因此对其性质的研究往往通过其特征函数。Nolan(1997)[72] 给出了稳定分布的以下定义方法。

定义3: (基于特征函数的定义) 假设随机变量 X 服从稳定分布, 记为 $x \sim S(\alpha, \beta, \gamma, \delta, k)$, 参数 k 可以等于0、1、2, 这里只对 $k=1$ 的情况进行研究, 下面简称 $x \sim S(\alpha, \beta, \gamma, \delta)$, 其特征函数为:

$$\chi_{S(\alpha, \beta, \gamma, \delta)}(x) = \begin{cases} \exp(-\gamma^\alpha |x|^\alpha [1 - i\beta(\tan \frac{\pi\alpha}{2}) \text{sign}(x)] + i\delta x) & \alpha \neq 1, \\ \exp(-\gamma |x| [1 + i\beta \frac{2}{\pi} (\ln|x|) \text{sign}(x)] + i\delta x) & \alpha = 1. \end{cases} \quad (3.1.1)$$

从第三个定义的表达式中, 我们可以看出, 服从稳定分布的随机变量是由4个参数 α 、 β 、 γ 、 δ 描述的, 其中:

1. α 为特征指数 (Characteristic exponent), $0 < \alpha \leq 2$, 描述尖峰、厚尾程度。 α 越小, 峰越尖, 尾越厚越长; 反之, 随着 α 增大, 分布的尾将变浅。
2. β 为对称参数 (Symmetry Parameter), $-1 \leq \beta \leq 1$, 决定了分布的偏斜程度。当 β 为0时, 分布是对称的; 当 $\beta > 0$ 时分布右偏, 当 $\beta < 0$ 时分布左偏。
3. γ 为尺度参数 (Scale Parameter), $\gamma \geq 0$, 描述了分布偏离其均值的程度。 γ 类似于高斯分布的方差, 但是在分布为高斯分布时 γ 为方差的两倍。
4. δ 为位置参数 (Location Parameter), $\delta \in R$, 反映概率密度函数在 x 轴的偏移。当 $\gamma = 1$, $\delta = 0$ 时, 稳定分布是标准的。

3.2 稳定分布的几种特殊情况

当 $\alpha = 2$ 时稳定分布为正态分布(方差为 $\sigma^2 = 2\gamma^2$, 均值为 δ); 当 $\alpha=1$, $\beta=0$ 时稳定分布为柯西分布; 当 $\alpha=1/2$, $\beta=1$ 时稳定分布为levy 分布。

稳定分布随机变量的概率密度函数存在并且连续, 但是除了正态分布, 柯西分布, 列维分布外, 没有解析表达式。

1. 高斯分布 (Gaussian Distribution) $S(2, 0, \gamma, \delta) = N(\delta, 2\gamma^2)$, 其概率密度函数为:

$$f(x) = \frac{1}{2\gamma\sqrt{\pi}} \exp(-\frac{(x-\delta)^2}{4\gamma^2}).$$

2. 柯西分布 (Cauchy Distribution) $S(1, 0, \gamma, \delta)$, 其概率密度函数为:

$$f(x) = \frac{\gamma}{\pi((x-\delta)^2 + \gamma^2)}.$$

如果 $x \sim S(1, 0, \gamma, 0)$, 则对于 $x > 0$, 有:

$$P(X \leq x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{\gamma}\right).$$

3. 列维分布 (Levy Distribution) $S(\frac{1}{2}, 1, \gamma, \delta)$, 其概率密度函数为:

$$f(x) = \left(\frac{\gamma}{\pi}\right)^{1/2} \frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\delta)}\right).$$

其中在区间 (δ, ∞) 上收敛。如果 $x \sim S(\frac{1}{2}, 1, \gamma, 0)$, 则对于任意的 $x > 0$, 有:

$$P(X \leq x) = 2(1 - \phi(\sqrt{\frac{\gamma}{x}})).$$

其中: ϕ 为标准正态分布 $N(0, 1)$ 的分布函数。

图 3.1 显示了稳定分布的以上几种特殊情况的密度函数:

3.3 稳定分布的性质

性质1: 对于两个独立的稳定分布, X_i 满足稳定分布 $S(\alpha, \beta_i, \gamma_i, \delta_i)$, c_i 不同时等于0, $i=1, 2$, 则 $c_1 X_1 + c_2 X_2$ 也满足稳定分布 $S(\alpha, \beta, \gamma, \delta)$, 其中

$$\begin{aligned} \beta &= \frac{(|c_1| \gamma_1)^\alpha \text{sign}(c_1) \beta_1 + (|c_2| \gamma_2)^\alpha \text{sign}(c_2) \beta_2}{(|c_1| \gamma_1)^\alpha + (|c_2| \gamma_2)^\alpha}, \\ \gamma &= [(|c_1| \gamma_1)^\alpha + (|c_2| \gamma_2)^\alpha]^{\frac{1}{\alpha}}, \\ \delta &= c_1 \delta_1 + c_2 \delta_2. \end{aligned} \tag{3.3.2}$$

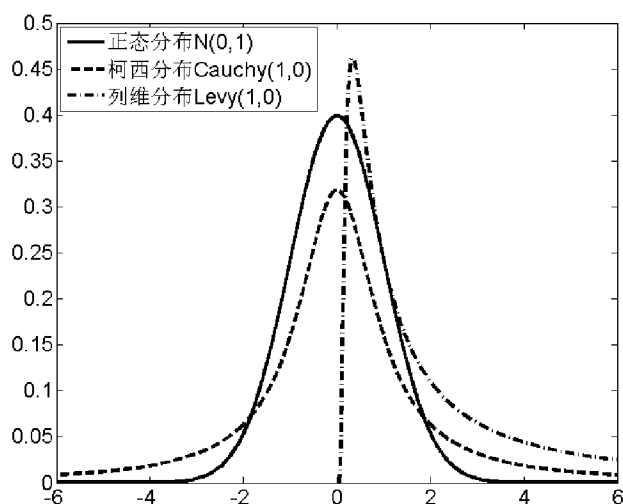


图 3.1: 稳定分布几种特殊情况的密度函数。

对于多个独立的稳定分布, X_i 满足稳定分布 $S(\alpha, \beta_i, \gamma_i, \delta_i)$, c_i 不同时等于 0, $i=1, 2, \dots, n$, 则 $c_1X_1 + c_2X_2 + \dots + c_nX_n$ 也满足稳定分布 $S(\alpha, \beta, \gamma, \delta)$, 其中

$$\begin{aligned}\beta &= \frac{\sum_{i=1}^n (|c_i|\gamma_i)^\alpha \text{sign}(c_i)\beta_i}{\sum_{i=1}^n (|c_i|\gamma_i)^\alpha}, \\ \gamma &= \left[\sum_{i=1}^n (|c_i|\gamma_i)^\alpha \right]^{\frac{1}{\alpha}}, \\ \delta &= \sum_{i=1}^n c_i\delta_i.\end{aligned}\tag{3.3.3}$$

性质2: 若 $x \sim S(\alpha, \beta, \gamma, \delta, 1)$, 那么对于任意的 $a \neq 0, b \in \mathbb{R}$,

$$ax + b = \begin{cases} (S(\alpha, (\text{sign} a)\beta, |a|\gamma, a\delta + b, 1)) & \alpha \neq 1, \\ (S(1, (\text{sign} a)\beta, |a|\gamma, a\delta + b - \frac{2}{\pi}\beta\gamma a \ln|a|, 1)) & \alpha = 1. \end{cases}\tag{3.3.4}$$

性质3: 稳定分布 $S(\alpha, \beta, \gamma, \delta, 1)$ 除在 $\alpha = 1$ 外, 其特征函数、密度函数和分布函数在其它任何点都连续。

3.4 稳定分布的参数估计及数值计算

3.4.1 稳定分布的参数估计

由于稳定分布不存在显性的概率密度函数和分布函数，仅有特征函数表示，所以估计稳定分布的参数比较困难。目前，存在三种常用的估计稳定分布参数的方法。

1. 基于分位数的参数估计方法

Fama和Roll (1968) [73]最早提出了对称稳定分布 ($\beta = 0$) 的基于分位数的参数估计方法。McCulloch (1986) [74]推广和改进了Fama和Roll 提出的估计方法，提出了目前常用的基于分位数的参数估计方法。基于分位数的参数估计方法能够给出稳定分布的四个参数：特征指数 α 、偏斜指数 β 、尺度参数 γ 和位置参数 δ 的一致估计量。记 x_a 为随机变量 x 的第 a 分位数， \hat{x}_a 为 x_a 的样本估计值，McCulloch 定义了如下的两个统计量：

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}, \quad v_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.5}}{x_{0.95} - x_{0.05}}. \quad (3.4.5)$$

v_α 和 v_β 是独立于尺度参数 γ 和位置参数 δ 。McCulloch 证明了特征指数 α 和偏斜指数 β 与 v_α 和 v_β 之间具有如下的函数关系：

$$\alpha = \psi_1(v_\alpha, v_\beta), \quad \beta = \psi_2(v_\alpha, v_\beta) \quad (3.4.6)$$

根据 v_α 和 v_β 的样本值 \hat{v}_α 和 \hat{v}_β 和McCulloch提供的线性插值表可以得到特征指数 α 和偏斜指数 β 的估计值 $\hat{\alpha}$ 和 $\hat{\beta}$ 。

采用类似的方法可以得到尺度参数 γ 和位置参数 δ 的估计值 $\hat{\gamma}$ 和 $\hat{\delta}$ ：

$$v_\gamma = \frac{x_{0.75} - x_{0.25}}{\gamma}, \quad v_\delta = \frac{\delta - x_{0.5}}{\gamma}. \quad (3.4.7)$$

则：

$$\hat{\gamma} = \frac{\hat{x}_{0.75} - \hat{x}_{0.25}}{\phi_3(\hat{\alpha}, \hat{\beta})}, \quad \hat{\delta} = \hat{\gamma}\phi_4(\hat{\alpha}, \hat{\beta}) + \hat{x}_{0.5}. \quad (3.4.8)$$

2. 基于特征函数的参数估计方法

Kourtouvelsi(1980)[75] 采用回归分析提出了一种基于特征函数的参数估计方

法。由Samorodnitsky-Taquu 形式的特征函数可以得到:

$$\ln(-\ln|\phi(t)|^2) = \ln(2\gamma^\alpha) + \alpha \ln|t|. \quad (3.4.9)$$

当 $\alpha \neq 1$, 取特征函数的实部和虚部:

$$\begin{aligned} R(\phi(t)) &= \exp(-|\gamma t|^\alpha) \cos[\delta t + |\gamma t|^\alpha \beta \text{sign}(t) \tan \frac{\pi t}{2}], \\ I(\phi(t)) &= \exp(-|\gamma t|^\alpha) \sin[\delta t + |\gamma t|^\alpha \beta \text{sign}(t) \tan \frac{\pi t}{2}]. \end{aligned} \quad (3.4.10)$$

则:

$$\arctan\left(\frac{I(\phi(t))}{R(\phi(t))}\right) = \delta t + \beta \gamma^\alpha \tan \frac{\pi t}{2} \text{sign}(t) t^\alpha, \quad (3.4.11)$$

则随机变量的样本特征函数为:

$$\hat{\phi}(t) = \frac{1}{n} \sum_{i=1}^n \exp(itx_i), \quad (3.4.12)$$

其中 x_1, x_2, \dots, x_n 为 x 的一组样本。样本的特征函数 $\hat{\phi}(t)$ 为 $\phi(t)$ 的一致估计量。式子 3.4.11仅与特征指数 α 和尺度参数 γ 有关, 可以通过回归分析得到 α 和 γ 的估计量 $\hat{\alpha}$ 和 $\hat{\gamma}$, 然后利用式子 3.4.12 进行回归分析得到偏斜指数 β 和位置参数 δ 的估计量 $\hat{\beta}$ 和 $\hat{\delta}$ 。以 $\hat{\alpha}, \hat{\gamma}, \hat{\beta}, \hat{\delta}$ 为初始估计量多次进行回归, 直到达到设置的收敛目标。

Kogon&Williams(1968)[76] 对Koutrouvelis(1980) [75]的估计方法进行了改进, 不仅使计算时间有所减少, 而且在 $\alpha = 1$ 和 $\beta \neq 1$ 的附近参数估计得结果更好。

3. 极大似然估计

在统计学中,极大似然估计是最常用的参数估计方法,且极大似然估计有非常好的渐进分布性质。因此,可以应用极大似然估计方法估计稳定分布的四个参数:特征指数 α 、偏斜指数 β 、尺度参数 γ 和位置参数 δ 。记 x_1, x_2, \dots, x_n 是随机变量 x 的一组样本,则 α 、 β 、 γ 和 δ 的极大似然估计值是使对数似然函数:

$$L(\alpha, \beta, \gamma, \delta) = \sum_{i=1}^n \ln \tilde{f}(x_i; \alpha, \beta, \gamma, \delta),$$

取最大时的参数值，其中 $\tilde{f}(\cdot; \alpha, \beta, \gamma, \delta)$ 是稳定分布随机变量 x 的密度函数。极大似然估计的方法思路清晰，计算简单。但对于稳定分布来说，应用极大似然估计方法的最大问题是：稳定分布的密度函数没有显性的表达式，即不知道 $\tilde{f}(\cdot; \alpha, \beta, \gamma, \delta)$ 的具体函数形式。所以需要采用数值方法与之近似逼近。目前常用的数值方法是Nolan(2001)[77]提出的直接积分法。

在以上三种参数估计方法中，基于分位数的参数估计方法最简单、计算速度最快，但是准确度最差；基于特征函数的参数估计方法所需时间较长，与基于分位数的参数估计方法相比提高了准确度；而极大似然估计所需时间最多，准确度也最高。所以本文采用极大似然估计方法来估计稳定分布的参数。

3.4.2 稳定分布的数值计算

Nolan(1997)[8]给出了稳定分布密度函数和分布函数的计算公式，并给出了详细的证明过程。当 $\gamma = 1$ ， $\delta = 0$ 时稳定分布为标准稳定分布，标准稳定分布的特征函数为：

$$\chi_{S(\alpha, \beta)}(x) = \begin{cases} \exp(-|x|^\alpha [1 + i\beta(\tan \frac{\pi\alpha}{2})\text{sign}(x)(|x|^{1-\alpha} - 1)]) & \alpha \neq 1, \\ \exp(-|x| [1 + i\beta \frac{2}{\pi}(\ln|x|)\text{sign}(x)]) & \alpha = 1. \end{cases} \quad (3.4.13)$$

文献中给出标准稳定分布的密度函数和分布函数表示如下：

$$f(x; \alpha, \beta) = c_2(x, \alpha, \beta) \int_{-\theta_0}^{\frac{\pi}{2}} g(\theta; x, \alpha, \beta) \exp(-g(\theta; x, \alpha, \beta)) d\theta. \quad (3.4.14)$$

$$F(x; \alpha, \beta) = c_1(\alpha, \beta) + c_3(\alpha) \int_{-\theta_0}^{\frac{\pi}{2}} \exp(-g(\theta; x, \alpha, \beta)) d\theta, \quad (3.4.15)$$

其中：

$$\begin{aligned}
 \zeta &= \zeta(\alpha, \beta) = \begin{cases} -\beta \tan \frac{\pi\alpha}{2} & \alpha \neq 1, \\ 0 & \alpha = 1. \end{cases} \\
 \theta_0 &= \theta_0(\alpha, \beta) = \begin{cases} \frac{1}{\pi} \arctan(\beta \tan \frac{\pi\alpha}{2}) & \alpha \neq 1, \\ \frac{\pi}{2} & \alpha = 1. \end{cases} \\
 c_1(\alpha, \beta) &= \begin{cases} \frac{1}{\pi}(\frac{\pi}{2} - \theta_0) & \alpha < 1, \\ 0 & \alpha = 1, \\ 1 & \alpha > 1. \end{cases} \\
 V(\theta; \alpha, \beta) &= \begin{cases} (\cos \alpha \theta_0)^{\frac{1}{\alpha-1}} \left(\frac{\cos \theta}{\sin \alpha(\theta_0 + \theta)} \right)^{\frac{\alpha}{\alpha-1}} \frac{\cos(\alpha \theta_0 + (\alpha-1)\theta)}{\cos \theta} & \alpha \neq 1, \\ \frac{2}{\pi} \left(\frac{\frac{\pi}{2} + \beta \theta}{\cos \theta} \right) \exp\left(\frac{2}{\pi} \left(\frac{\pi}{2} + \beta \theta \right) \tan \theta\right) & \alpha = 1, \beta \neq 0. \end{cases} \\
 c_2(x; \alpha, \beta) &= \begin{cases} \alpha / (\pi |\alpha - 1| (x - \zeta)) & \alpha \neq 1, \\ 1 / (2|\beta|) & \alpha = 1. \end{cases} \\
 c_3(\alpha) &= \begin{cases} \text{sign}(1 - \alpha) / \alpha & \alpha \neq 1, \\ 1 / \pi & \alpha = 1. \end{cases} \\
 g(\theta; x, \alpha, \beta) &= \begin{cases} (x - \zeta)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta) & \alpha \neq 1, \\ e^{-\frac{\pi x}{2\beta}} V(\theta; 1, \beta) & \alpha = 1. \end{cases}
 \end{aligned} \tag{3.4.16}$$

$$\begin{aligned}
 c_3(\alpha) &= \begin{cases} \text{sign}(1 - \alpha) / \alpha & \alpha \neq 1, \\ 1 / \pi & \alpha = 1. \end{cases} \\
 g(\theta; x, \alpha, \beta) &= \begin{cases} (x - \zeta)^{\frac{\alpha}{\alpha-1}} V(\theta; \alpha, \beta) & \alpha \neq 1, \\ e^{-\frac{\pi x}{2\beta}} V(\theta; 1, \beta) & \alpha = 1. \end{cases}
 \end{aligned} \tag{3.4.17}$$

对于非标准稳定分布，我们可以根据稳定分布的性质，把随机变量进行标准化。例如：若随机变量 $X \sim S(\alpha, \beta, \gamma, \delta)$ ，则 $\frac{X-\delta}{\gamma} \sim S(\alpha, \beta, 1, 0)$ ，其中 $\frac{X-\delta}{\gamma}$ 为标准稳定分布。我们参考Nolan(1997)[72]，运用高斯数值积分方法，得到稳定分布的密度函数和累积分布函数。从图 3.2a 和图 3.2b 可以看出随着 α 的变小，分布的峰越来越尖，尾部越来越厚。

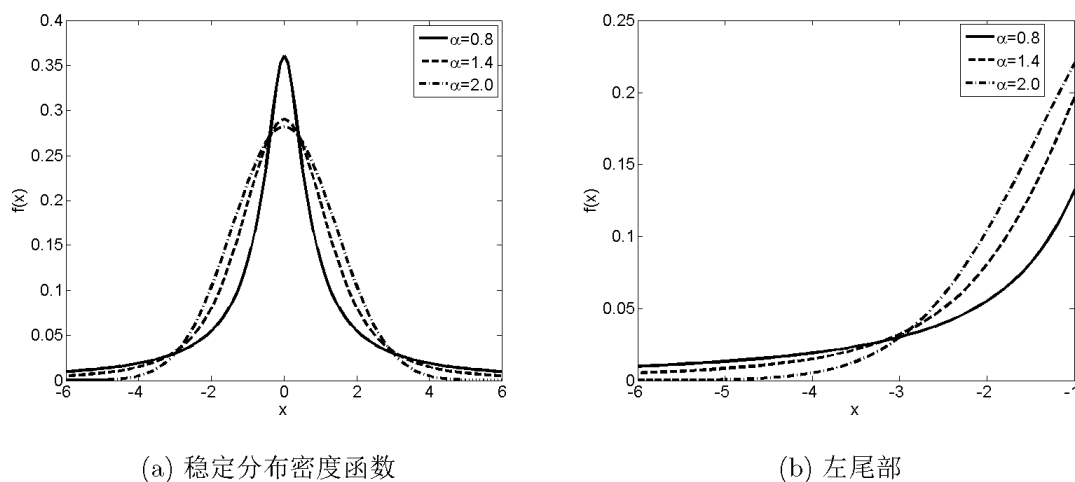


图 3.2: 稳定分布 $\alpha = 0.8, 1.4, 2.0, \beta = 0, \gamma = 1, \delta = 0$ 的密度函数。

3.5 稳定分布与市场数据的拟合

在金融领域的研究中，经常假设资产收益率的分布服从正态分布，但是大量的实证研究表明金融数据的收益率具有尖峰、厚尾等非正态特征。与正态分布相比，稳定分布能更好的拟合金融市场数据。因此，有关资产收益模型选择稳定分布更加合理。在本节中分别采用正态分布和稳定分布对上证指数、恒生指数、标普500指数的日数据以及沪深300股指期货高频数据进行拟合，通过观察分析，可以看出稳定分布能够更好地拟合市场数据。

3.5.1 稳定分布与指数的拟合

用极大似然法对上证指数、恒生指数、标普500指数的日数据进行参数拟合。选取2000年1月到2016年3月这段长数据来观察长期数据的统计特征及分布情况；选择2013年1月到2016年3月这段数据作为短数据分析最近三年指数数据所呈现出的统计特征。从表 3.1 可以看出在长期看来三个指数的 α 都在1.45到1.5这个范围内，其中标普500指数的 α 最小；从短期来看，上证指数的 α 最小，说明最近三年上证指数的收益率序列尖峰后尾程度比较大，存在的套利机会比较多，这可能与中国市场

不太完善有很大关系。图 3.3， 3.4， 3.5 分别是上证指数、恒生指数、标普500 指数的稳定分布和正态分布拟合图，直观地表现了市场数据的尖峰厚尾统计特征，从图中也可以看出稳定分布比正态分布能够更好的拟合市场数据。

表 3.1: 上证指数、恒生指数及标普500 指数正态分布和稳定分布的拟合参数(日数据)

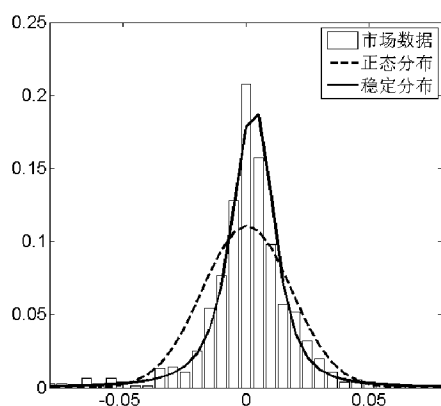
参数 \ 指数		上证指数		恒生指数		标普500指数	
		长数据	短数据	长数据	短数据	长数据	短数据
高斯分布	μ	2E-4	3E-4	3E-5	-2E-4	8E-5	4E-4
	σ	0.017	0.018	0.015	0.011	0.013	0.009
稳定分布	α	1.453	1.331	1.490	1.519	1.450	1.495
	β	-0.095	-0.116	-0.078	-0.171	-0.132	-0.044
	γ	8E-3	8E-3	8E-3	5E-3	6E-3	6E-3
	δ	9E-4	1E-3	4E-4	8E-4	7E-4	-4E-5

3.5.2 稳定分布与高频数据的拟合

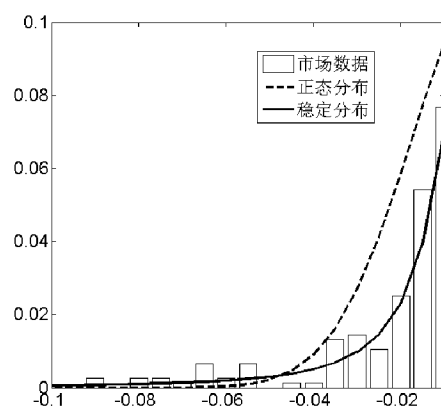
由第二章《有效市场假说与高频数据》的介绍可知，通过游程检验可以判断出高频数据在什么频率下是“可测的”，所以在用稳定分布对高频数据进行拟合之前，选出标的高频数据的可测频率，然后再对高频数据的统计特征进行具体分析。希望得出这样的结果：可测的高频数据，也具有尖峰厚尾的特征。

高频数据的游程检验

本文采用“非参数游程检验”和“随机游动检验”两种方法对高频数据进行“有效性”检验，由第二章叙述可知，“随机游动检验”这种检验方法理论性更强，得出的结果会更可靠。所以下面对沪深300股指期货（IF）进行“随机游动检验”，

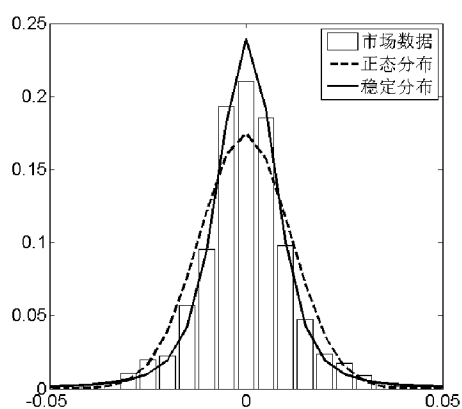


(a) 不同分布对真实市场数据拟合

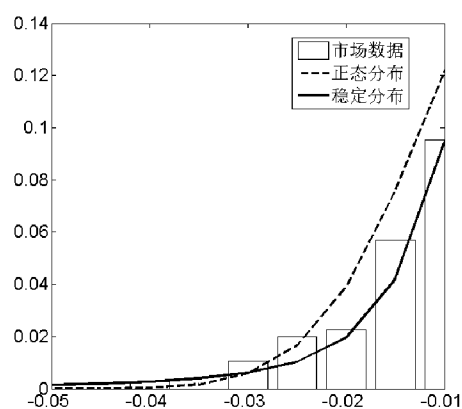


(b) 不同分布对真实市场数据拟合（尾部）

图 3.3: 上证指数日数据
(时间:2013/1-2016/3 ,数据来源:Bloomberg)



(a) 不同分布对真实市场数据拟合



(b) 不同分布对真实市场数据拟合（尾部）

图 3.4: 恒生指数日数据
(时间:2013/1-2016/3 ,数据来源:Bloomberg)

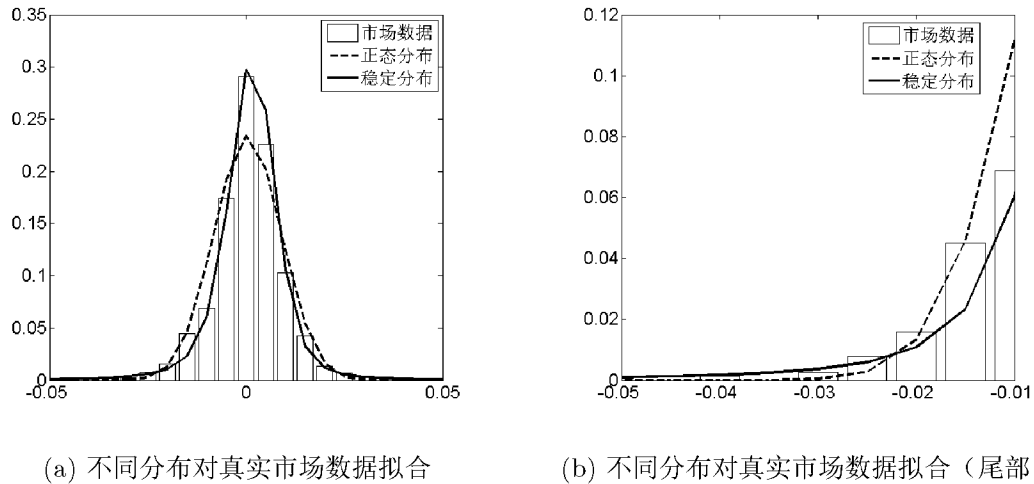


图 3.5: 标普500指数日数据

（时间:2013/1-2016/3 ,数据来源:Bloomberg）

从表 3.2可以看出“IF”在“5 秒”、“30 秒”、“1分钟”数据是“可测的”，这也说明在这三个频率的高频数据中市场效率短暂丧失，存在转瞬即逝的交易机会，低于“1 分钟”频率的数据都是“不可测的”，在构建关于沪深300股指期货的高频策略时主要考虑高于“1分钟”频率的数据。

表 3.2: IF（2015年7月到2015 年8月）

数据频率	5秒	30秒	1分钟	10分钟	30 分钟	1小时
边界值	3.920	3.920	3.920	3.920	3.920	3.920
统计量	5.479	36.708	-5.064	-1.908	-1.969	-1.921
游程检验结果	“可测的”	“可测的”	“可测的”	“随机的”	“随机的”	“随机的”

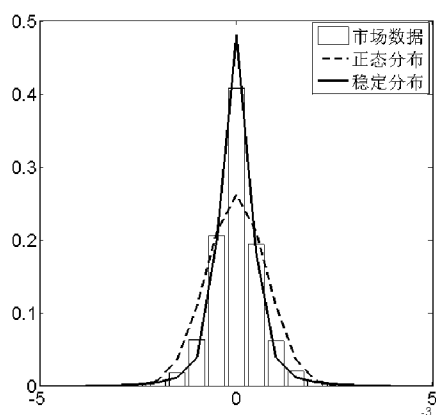
稳定分布与高频数据的拟合

下面对沪深300股指期货（IF）5 秒、30 秒、1 分钟高频数据同样在2015年7月到2015 年8月这段时间进行研究分析，从表 3.3 可以看出：5 秒、30 秒、1 分钟

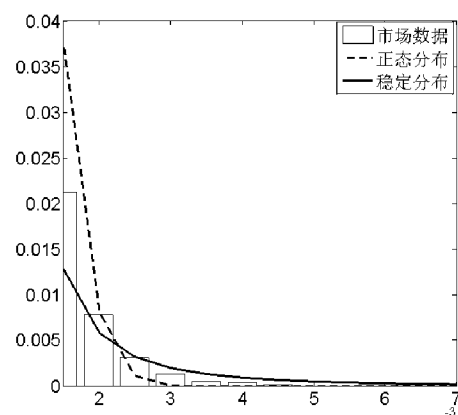
数据的 α 都处在1.33到1.46范围内，并且随着数据频率递减而递减。图 3.6， 3.7， 3.8分别是不同分布对5 秒、30 秒、1 分钟数据的拟合，可以看出稳定分布也能够很好地拟合高频数据，从图中可以看出高频数据具有尖峰厚尾的特征，这说明高频数据存在很大的套利空间。基于高频数据的统计特征，投资者可以开发能够带来正收益的高频数据套利策略。

表 3.3: 沪深300股指期货（IF）正态分布和稳定分布的拟合参数(高频数据)

“IF” 参数		5秒	30 秒	1分钟
高斯分布	μ	-1.7912E-5	-3.7107E-6	-6.7826E-6
	σ	0.0012	0.0019	0.0026
稳定分布	α	1.4579	1.3631	1.332
	β	0.0384	0.055	0.0264
	γ	0.0003	0.0007	0.0011
	δ	-0.4E-5	-0.6E-5	-0.9E-5

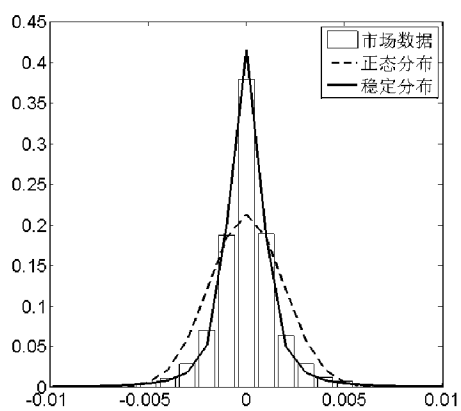


(a) 不同分布对真实市场数据拟合

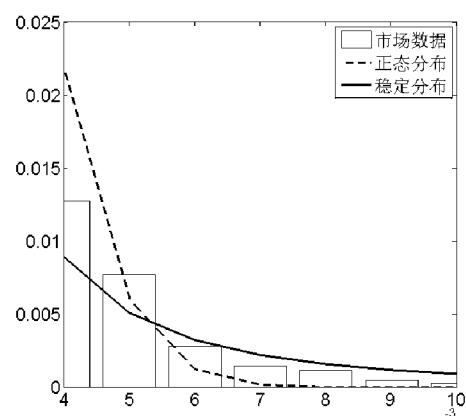


(b) 不同分布对真实市场数据拟合（尾部）

图 3.6: 沪深300股指期货5秒数据
(时间:2015/7-2015/8,数据来源: Bloomberg)



(a) 不同分布对真实市场数据拟合



(b) 不同分布对真实市场数据拟合（尾部）

图 3.7: 沪深300股指期货30秒数据
(时间:2015/7-2015/8,数据来源: Bloomberg)

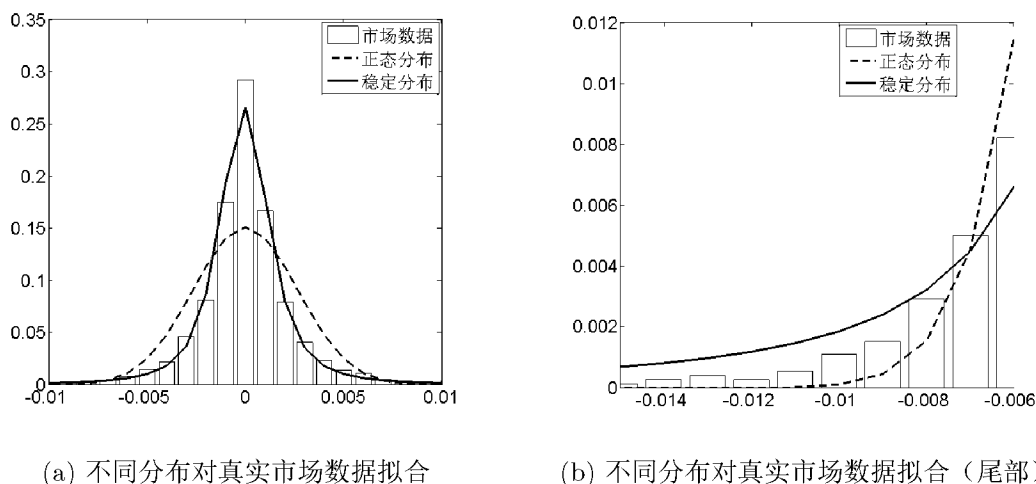


图 3.8: 沪深300股指期货1分钟数据

(时间:2015/7-2015/8,数据来源: Bloomberg)

小结

从以上分析可以看出,“可测的”的高频数据,也具有尖峰厚尾的统计特征。事实上,高频数据的“可测性”是从市场有效性方面来说明,市场在短暂的时间上丧失效率;高频数据具有尖峰厚尾统计特征,并且更符合稳定分布,是从数据的本质上说明高频数据存在潜在的交易机会。以Robert J. Shiller 为代表的“行为金融理论”从一种新的视角正确反映市场主体实际决策行为以及准确描述市场运行的状况。行为金融学就是将心理学尤其是行为科学的理论与金融学相融合,从微观个体行为以及产生这种行为的心理等动因来解释、研究和预测金融市场的发展,这一新的研究方向通过分析金融市场主体在市场行为中的偏差和反常,来寻求不同市场主体在不同环境下的经营理念及决策行为特征。

第四章 构造稳定分布的高频数据及实证研究

通过对第二章有关市场有效假说理论的了解，可以知道在有效的市场中不可能获得任何利润，也就是说没有投资者可以获取超过风险调整后的平均收益；只有在市场无效时才可以找到套利的机会，并且市场越无效，具有可预测性的套利机会越多。从第三章稳定分布理论我们可以知道稳定分布比其它分布能更好的拟合市场真实数据，并且能更好的展示出真实金融市场数据尖峰后尾的统计特征，尖峰后尾程度是从另一个侧面展现市场的无效性。所以构造出峰越尖，尾巴拖的越长越厚的数据序列具有更多的套利机会。用具有这种特征的数据建立策略模型会有巨大的盈利能力。所以本章主要从构造更具稳定分布特征的高频数据序列、用MACD策略进行实证检验分析这两个方面进行了具体论述。

4.1 数据构造模型介绍

研究表明市场越无效，有可预测性的交易机会越多；市场无效表现在数据尖峰厚尾的统计特征；根据第三章关于稳定分布的介绍，我们知道 α ($0 < \alpha \leq 2$) 为特征指数 (Characteristic exponent)，描述尖峰厚尾程度， α 越小，峰越尖，尾越厚且越长；反之，随着 α 增大，分布的尾将变浅。本文希望通过品种叠加，构造出 α 比较小的数据序列。然后将其运用到高频统计套利策略上，观察对比分析策略的效果。

假设有 n 个标的， R_1, R_2, \dots, R_n 是各个标的的收益率序列， $W=w_1, w_2, \dots, w_n$ 是

对应系数, $R_p = w_1 R_1 + w_2 R_2 + \dots + w_n R_n$; 那么给定一组系数就计算出一个 R 序列, 每一个 R 序列带入 *stable* 软件, 得出一个 α , 本文研究主要模型:

$$\begin{aligned} & \min_W \alpha_p \\ & s.t. \begin{cases} W^T I = 1, \\ -1 \leq \rho_{i,j} \leq m, m < 1, \end{cases} \end{aligned} \quad (4.1.1)$$

其中 I 为单位矩阵, α_p ($0 < \alpha_p \leq 2$) 为 R_p 对应的 α , $\rho_{i,j}$ 为 R_i 与 R_j 的相关系数 ($i, j = 1, 2, 3, \dots, n, i \neq j$)。

式 (4.1.1) 是一种简单的线性构造方式, 我们希望得到一个收益率序列 R , 它的特征是: α 比较小。当给定一个 α , 我们就可以找到一组与之对应的比例系数 w_1, w_2, \dots, w_n , 按照这个比例构造出了一个新的高频收益率序列, 这也等价于按照这个比例进行组合。

4.2 实证分析

根据Markowitz投资组合理论, 假设我们用期望值和标准差来描述不同的投资机会, r_1 与 r_2 分别表示两个标的的投资回报率, 如果我们把 w_1 的比例投入到第一种标的, 把 $w_2 = 1 - w_1$ 的比例投入到第二种标的。这样组合的回报率为: $w_1 r_1 + w_2 r_2$, 投资组合的期望回报值是: $\mu_p = w_1 \mu_1 + w_2 \mu_2$, 式中, μ_1 为第一个标的的回报期望值, μ_2 为第二个标的的回报期望值。所以组合的回报标准差为:

$$\sigma_p = \sqrt{w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2\rho w_1 w_2 \sigma_1 \sigma_2} \quad (4.2.2)$$

式中, σ_1 与 σ_2 分别表示 r_1 和 r_2 的标准差, ρ 表示 r_1 和 r_2 的相关系数。所以 ρ 越小, 组合的标准差越小, 即组合的风险越小; 在选择标的的时候, 首先对标的的进性相关性分析, 本文选择 $\rho \leq 50\%$ 的标的组合。

所以选择组合的标的时, 首先要通过相关性分析, 然后根据数据构造模型构造出不同 α 的数据序列, 同时找到对应的投资比例, 最后把数据序列应用高频交易策略上, 最后分析不同 α 的数据序列的投资效果, 得出结论。本节分为两小节, 第一小节选择了全球几个主要股指期货进行了比较详细的实证研究, 第二小节选择国内不同市场的几个指数进行实证分析, 第三小节对国内几个商品期货进行简要分析。

4.2.1 全球股指期货组合

本小节主要选择了几个股指期货品种，同样从股指期货的标的相关性、用极大似然法估计股指期货的分布参数、构造出不同 α 的收益率序列以及投资组合效果分析等几个方面进行了研究分析。

标的选择（相关性分析）

股指期货与其标的的相关性非常高，并且其标的是连续的，所以用标的代替股指进行各品种间的相关性分析，沪深300股指期货（代码：IF）、标普500 股指期货（代码：SP1）、恒生股指期货（代码：HI1）以及日经225股指期货的标的分别是沪深300 指数（代码：000300）、标普500 指数（代码：SPX）、恒生指数（代码：HIS）、日经225 指数（代码：NK）。从表 4.1 可以看出日经225 指数与标普500 指数和沪深300 指数的相关系数在16%左右，其他各标的间的相关系数在5% 以下。综合比较，选取标普500 股指期货，沪深300 股指期货以及恒生股指期货进行下面的分析。

表 4.1: 相关系数矩阵（时间：2012/05/24-2015/8/14）

	“SPX”	“000300”	“HIS”	“NK”
“SPX”	1	0.0769	0.0724	0.1687
“000300”	0.0769	1	0.0482	10.1624
“HIS”	0.0724	0.0482	1	-0.0295
“NK”	0.1687	0.1624	-0.0295	1

构造收益率序列

下面主要是对三个股指5分钟、1分钟以及5 秒数据，在时间段2015/05/13-2015/10/30 上进行参数估计，从表 4.2 可以看出标普500 股指期货的 α 最小，与其他两个股指期货相比5分钟的标普500 股指期货数据尖峰厚尾程度更大。从表 4.3、4.4 可以看出所选的三个股指的 α 都在1.3左右，可见在1分钟、30秒数据频率上，三个股指的尖峰厚尾程度基本一致。对单个股指数据特性进行分析之后，随机模拟1000 组3 维系数，构造出1000 组收益率序列，相应的我们就得到1000 个组合 α ；

筛选3个标的的比例系数同时都在0.1 到0.8 范围内的组合 α ，可以找到 α 最小的一组 and α 最大的一组系数，如表 4.5 所示，以5 分钟数据为例： $\alpha = 1.428$ 是满足条件限制中最小的，77%、12%、11%分别是SP1、HI1、IF 所占的比例； $\alpha = 1.491$ 是满足条件限制中最大的，11%、53%、36%分别是SP1、HI1、IF 所占的比例。

表 4.2: “SP1”、“HI1” 及 “IF” 正态分布和稳定分布的拟合参数（5分钟）

参数 \ 股指期货		“SP1”	“HI1”	“IF”
高斯分布	μ	-3.99E-7	-1.2E-5	-5.74E-5
	σ	2.46E-3	1.5E-3	5.77E-3
稳定分布	α	1.1711	1.285	1.3687
	β	0.0015	-0.029	-0.04
	γ	0.0003	0.0004	0.0023
	δ	-1E-7	-4E-6	-2E-5

表 4.3: “SP1”、“HI1” 及 “IF” 正态分布和稳定分布的拟合参数（1分钟）

参数 \ 股指期货		“SP1”	“HI1”	“IF”
高斯分布	μ	-0.0000944	0.0000574	0.000113
	σ	0.00741	0.00423	0.00420
稳定分布	α	1.3379	1.2871	1.2994
	β	0.001	0.0055	-0.0077
	γ	0.003	0.0015	0.0014
	δ	-0.000001	-0.000003	0.000003

表 4.4: “SP1”、“HI1”及“IF”正态分布和稳定分布的拟合参数（30 秒）

参数 \ 股指期货		“SP1”	“HI1”	“IF”
高斯分布	μ	-0.0000944	0.0000574	0.000113
	σ	0.00741	0.00423	0.00420
稳定分布	α	1.3379	1.2871	1.2994
	β	0.001	0.0055	-0.0077
	γ	0.003	0.0015	0.0014
	δ	-0.000001	-0.000003	0.000003

表 4.5: 序列参数

数据频率	组合 α	“SP1” 占比(%)	“HI1” 占比(%)	“IF” 占比(%)
“5分钟”	1.4284	77	12	11
	1.491	11	53	36
“1分钟”	1.3042	16	13	71
	1.4384	23	67	10
“30秒”	1.2181	13	10	77
	1.3656	25	65	10

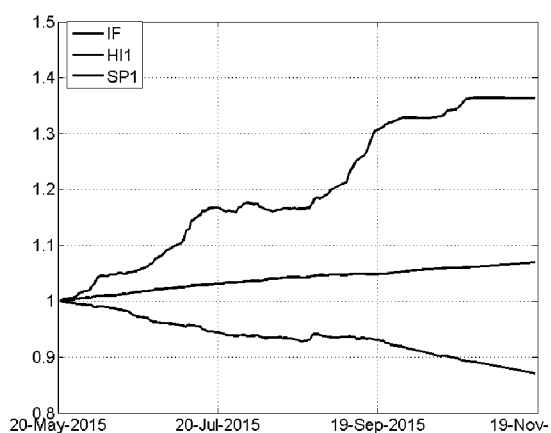
投资组合分析

通过“游程检验”发现在“1分钟”频率下，“SP1”、“HI1”及“IF”都是“可测的”。有关股指的高频交易模型有很多，本文选用MACD策略：MACD 指标出现金叉时进场做多，出现死叉时进场做空，用表 4.5 “1 分钟”频率参数进行投资组合。投资结果如表 4.6 所示，通过分析可以得出以下结果：

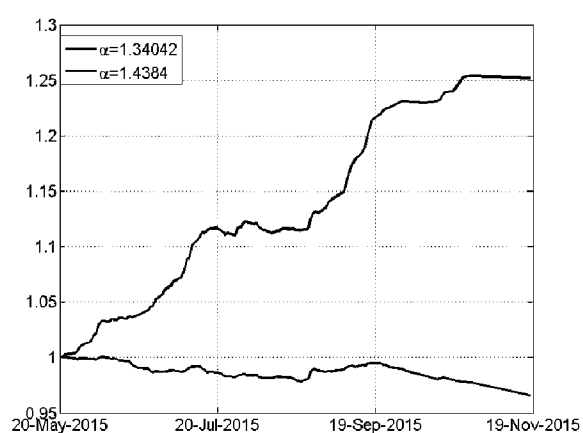
1. 单独投资于单个标的，从年化收益率指标来看，投资于IF的收益率最大： $r = 73\%$ ，投资于“HI1”的收益率最小： $r = -26\%$ ，而从表 4.3 可以看出“HI1”和“IF”1分钟数据的 α 相差不大，收益率却差很多，这可能和单个标的不能很好地分散风险有关系。从年化标准差指标来看，单独投资于“IF”和“HI1”的风险相对都比较大。从夏普比例来看除了投资HI1 的夏普比率比较小且是负数，其他两个都比较大，SP1夏普比率达到了10 以上。
2. “ $\alpha = 1.3042$ 的组合”是小 α 组合，“ $\alpha = 1.4383$ 的组合”是大 α 组合，从两个组合来看，有明显的差别，不仅小 α 组合的收益率比大 α 组合的收益率高出很多，并且夏普比例也比较高。
3. 小 α 组合不仅比大 α 组合有明显的优势，比单个标的投资效果也要好很多。它虽然没有单独投资于IF的收益率高，但夏普比率却高出2 点多；它没有单独投资于SP1的夏普比率高，但收益率却高出很多。所以综合来说小 α 组合具有明显的优势。从图 4.1可以更直观的看出小 α 组合的效果，可以看出小 α 组合相比其他投资方式具有明显优势。

表 4.6: 组合投资结果(无风险收益率=3%)

	年化收益率 (%)	年化标准差 (%)	夏普比率
“ $\alpha = 1.3042$ 组合”	51	4.9	9.771
“ $\alpha = 1.4383$ 组合”	-7	3	-3.3
“SP1”	14	1.5	10.84
“HI1”	-26	4.9	-5.97
“IF”	73	6.5	7.30



(a) 单个标的资金曲线图



(b) 组合资金曲线图

图 4.1: 全球股指期货资金曲线图

(时间:2015/05/13-2015/10/30,数据来源: Bloomberg)

4.2.2 国内跨市场指数组合

本小节选取了国内不同市场的五个指数，分别为：上证综合指数（代码：000001）、国债指数（代码：000012）、大宗商品指数（代码：399979）、创业板指数（代码：399006）以及中小板指数（代码：399005），从指数相关性、用极大似然法估计指数的分布参数、构造出不同 α 的收益率序列以及投资组合效果分析等几个方面进行详细研究分析，以下分析均用代码代替标的名称。

标的选择（相关性分析）

由表 4.7得出的相关性结果，综合比较，选“000001”、“000012”及“399979”这三个标的。

表 4.7: 相关系数矩阵（时间：2010/06/01-2015/06/08）

	“000001”	“000012”	“399979”	“399006”	“399005”
“000001”	1	0.0406	0.4941	0.5570	0.7606
“000012”	0.0406	1	0.0070	0.0133	0.0321
“399979”	0.4941	0.0070	1	0.6561	0.8375
“399006”	0.5570	0.0133	0.6561	1	0.8877
“399005”	0.7606	0.0321	0.8375	0.8877	1

构造收益率序列

下面主要是对“000001”、“000012”及“399979”三个标的的1分钟数据，在时间段2011年8月22到2016年3月16进行参数估计，表 4.8 可以看出“399979”的 α 最小；“000012”的 α 最大，并且达到了2，这可能与国债的交易不活跃有关。对单个股指数据统计特性进行分析之后，随机模拟1000组3维系数，构造出1000组收益率序列，相应的我们就得到1000个组合 α ；筛选3个标的的比例系数同时都在0.1到0.8范围内的组合 α ，可以找到 α 最小的一组 and α 最大的一组系数，如表 4.9 所示：“ $\alpha=1.3411$ 组合”是满足条件限制中最小的，11%、14%、75%分别是“000001”、“000012”及“399979”所占的比例；“ $\alpha=1.4540$ 组合”是满足条件限制中最大的，同样的，71%、18%、11%分别是“000001”、“000012”及“399979”

所占的比例。

表 4.8: “000001”、“000012”及“399979”正态分布和稳定分布的拟合参数（1分钟）

参数 \ 指数		“000001”	“000012”	“399979”
高斯分布	μ	6E-7	7E-7	-1E-6
	σ	0.001	0.00005	0.0011
稳定分布	α	1.4640	2	1.3278
	β	0.0134	0	0.0297
	γ	3E-4	3E-6	3E-4
	δ	1E-6	0	-4E-6

表 4.9: 序列参数

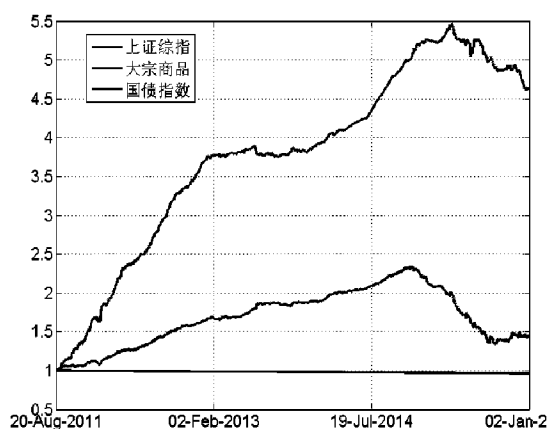
数据频率	组合 α	“000001” 占比(%)	“000012” 占比(%)	“399979” 占比(%)
1分钟	1.3411	11	14	75
	1.4540	71	18	11

投资组合分析

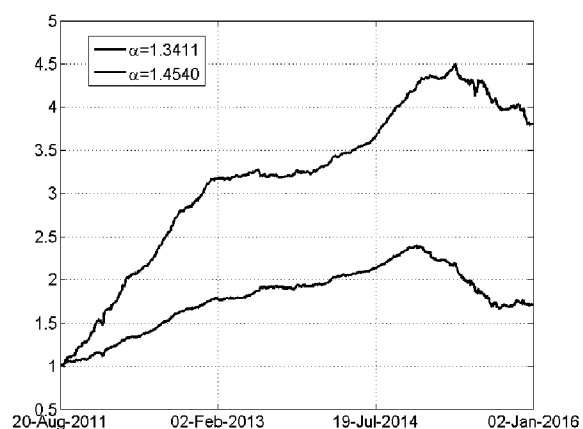
通过“游程检验”发现国内指数在“1分钟”频率数据也是“可测的”，同样地，选用MACD策略（不限于MACD 策略）：MACD 指标出现金叉时进场做多，出现死叉时进场做空，用表 4.9 参数进行投资组合。投资结果如表 4.10 所示，我们可以看出“ $\alpha = 1.3411$ 组合”的投资效果最好。

表 4.10: 组合投资结果(无风险收益率=3%)

	年化收益率 (%)	年化标准差 (%)	夏普比率
“ $\alpha = 1.3411$ 组合”	64	7	8.27
“ $\alpha = 1.4540$ 组合”	16	7	2
“000001”	10	9	0.81
“000012”	-1	0.1	-69.64
“399979”	83	8	9.53



(a) 单个标的资金曲线图



(b) 组合资金曲线图

图 4.2: 国内指数资金曲线图

(时间:2011/8/22-2016/3/16,数据来源: Bloomberg)

4.2.3 国内商品组合

由于篇幅原因，本节只对国内的几个商品期货的高频数据进行了相关性分析、数据构造两方面的阐述。

标的选择（相关性分析）

选择国内几个活跃的商品期货品种：橡胶（ ru ）、白糖（ SR ）、镍（ ni ）、甲醇（ MA ）、螺纹钢（ rb ）、豆粕（ m ）以及焦炭（ j ），表 4.11 与表 4.12 是对这几个标的相关性分析的结果。考虑到系数 $\rho \leq 50\%$ ，综合表 4.11 与表 4.12 的相关系数矩阵，最终选出橡胶（ ru ）、白糖（ SR ）以及豆粕（ m ）三个标的进行下面的分析（注：下面均用 ru 、 SR 、 m 对应表示三个标的）。

表 4.11: 相关系数矩阵（时间：2015/04/07-2015/12/24）

	rb	ru	j	m	SR	MA	ni
rb	1	0.51	0.54	0.29	0.28	0.51	0.49
ru	0.51	1	0.34	0.32	0.30	0.52	0.49
j	0.54	0.34	1	0.17	0.20	0.24	0.18
m	0.29	0.32	0.17	1	0.24	0.28	0.42
SR	0.28	0.30	0.20	0.24	1	0.29	0.28
MA	0.51	0.52	0.24	0.28	0.29	1	0.48
ni	0.49	0.49	0.18	0.42	0.28	0.48	1

数据构造

首先对 ru 、 SR 以及 m 三个标的 30 分钟、15 分钟、5 分钟、1 分钟数据用极大似然估计方法估计出稳定分布参数，如表 4.13、4.14、4.15、4.17 所示，可以看出这三个标的的 30 分钟和 15 分钟数据的 α 都在 1.2 与 1.4 之间，5 分钟数据的 α 都在 1.2 与 1.5 之间， m 1 分钟数据的 α 达到了 2。

对每个频率数据分别随机模拟出 1000 组 3 维系数，然后根据本章第一节数据构造模型，找出两组合适的参数，一组 α 比较小，另一组比较大，见表 4.16

表 4.12: 相关系数矩阵（时间：2011/04/18-2015/12/24）

	rb	ru	j	m	SR	MA
rb	1	0.51	0.58	0.19	0.21	0.42
ru	0.51	1	0.43	0.29	0.35	0.55
j	0.58	0.43	1	0.14	0.19	0.33
m	0.19	0.29	0.14	1	0.24	0.25
SR	0.21	0.35	0.19	0.24	1	0.32
MA	0.42	0.55	0.33	0.25	0.32	1

表 4.13: ru 、 SR 及 m 正态分布和稳定分布的拟合参数（30分钟）

商品期货		ru	SR	m
参数				
高斯分布	μ	-0.0000944	0.0000574	0.000113
	σ	0.00741	0.00423	0.00420
稳定分布	α	1.3379	1.2871	1.2994
	β	0.001	0.0055	-0.0077
	γ	0.003	0.0015	0.0014
	δ	-0.000001	-0.000003	0.000003

表 4.14: ru、SR及m 正态分布和稳定分布的拟合参数（15分钟）

参数 \ 商品期货		ru	SR	m
高斯分布	μ	-4.3793E-05	2.81316E-05	-5.26213E-05
	σ	0.005118448	0.002971731	0.002923731
稳定分布	α	1.3861	1.3252	1.2958
	β	0.0234	0.0055	0.0201
	γ	0.002	0.001	0.0008
	δ	-0.00004	-0.000007	-0.000006

表 4.15: ru、SR及m 正态分布和稳定分布的拟合参数（5分钟）

参数 \ 商品期货		ru	SR	m
高斯分布	μ	-1.42908E-05	9.6237E-06	-1.72546E-05
	σ	0.00293886	0.001718354	0.00169859
稳定分布	α	1.4657	1.4229	1.2771
	β	0.0347	0.0291	0.0026
	γ	0.0011	0.0006	0.0004
	δ	-0.00001	-0.000005	-0.0000004

表 4.16: ru、SR及m 正态分布和稳定分布的拟合参数（1分钟）

参数 \ 商品期货		ru	SR	m
高斯分布	μ	-2.85882E-06	1.8764E-06	-3.49052E-06
	σ	0.001369813	0.000772819	0.000794567
稳定分布	α	1.4683	1.3395	2
	β	0.0045	0.007	0
	γ	0.0005	0.0002	0.0004
	δ	-0.0000006	-0.0000005	0

表 4.17: 序列参数

数据频率	组合 α	ru 占比(%)	SR 占比(%)	m 占比(%)
30分钟	1.4005	11	17	72
	1.4427	40	49	11
15分钟	1.3715	71	11	18
	1.4901	21	40	39
5分钟	1.4589	10	74	16
	1.5518	20	36	44
1分钟	1.5685	11	79	10
	1.7298	13	11	76

小结

从全球股指期货组合和国内跨市场指数的组合投资效果来看， α 越小，投资的效果越好。从国内股票组合以及其他组合也得出了同样的结果，由于篇幅限制，在本文没能一一列出。

第五章 结论与展望

本文从金融市场的数据具有尖峰厚尾的统计特征，并且稳定分布能更好地对其收益率序列进行拟合出发，着重介绍了高频数据的统计特征，发现指数以及股指的高频数据也具有尖峰厚尾的特性。从有效市场理论出发，介绍了高频数据在市场短暂无效时存在交易机会，并通过对不同频率的高频数据进行有效性检验，发现某些标的在一定频率下是可测的，这也说明了高频数据存在巨大的套利机会，可以根据高频数据开发相应的统计套利模型。基于以上理论基础和研究发现，希望得出 α 的数据序列具有更好的交易机会，并且能够获得丰厚的回报。本文重点对全球股指组合以及国内不同指数组合进行了实证分析。通过构造不同 α 的收益率序列，对应找到组合的投资比例，本文用MACD策略对不同 α 的组合进行了投资分析，得出的结果和预期相符， α 小的组合投资效果明显比 α 大的组合要好很多。从年化收益率、年化标准差以及夏普比率等指标来看，综合分析比较得出： α 越小，投资效果越好。

从本文的分析结果来看确实得出了预想的结论，但也存在很多不足的地方，比如：(1) 本文只选择了全球股指期货以及国内不同指数进行组合分析，所以组合标的的选择不够全面。(2) 本文实证分析部分只对1分钟频率的数据进行比较完整的分析，缺少对其他频率的实证研究，所以在数据频率不够多样化。(3) 构造 α 组合时只用到线性组合的方法，没有用到非线性以及其他方法，所以存在构造数据方法太过简单的问题。当然还有存在一些其他问题。

由于期权的波动和杠杆比较大，存在的套利机会会比较多，所以在期权方面存在很大的研究空间；可以用熵（Entropy）的方法度量资产收益率的厚尾；把投资者的心理因素当做一个指标，运用在投资模型中。

参考文献

- [1] 艾琳.奥尔德里奇. 高频交易[M]. 北京: 机械工业出版社, page 3, 2011.
- [2] 蓝海平. 高频交易的技术特征—发展趋势及挑战. 证券市场导报, 2014.
- [3] S.Von der Becke D.Sornnette. Crashes and High Frequency Trading[R]. *Swiss Finance Institute Research*, pages 11–63, 2011.
- [4] Cvitanic, Jaks, and Andrei A Kirilenko. High Frequency Traders and Asset Prices[R]. *Working Paper(March 1).Available at SSRN*, 2010.
- [5] Chaboud, Alain, Hjalmarsson, and Erik. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market[R]. *FRB International Finance Discussion Paper No.980*, 2011.
- [6] J.A.Brogaard. The activity of high frequency traders[R]. *University of Washington Working Paper*, 2011.
- [7] P.Gomber,Martin Haferkorn. High-Frequency Trading[J]. *Business and Information Systems Engineering*, 2013.
- [8] Scholtus, L.Martin, Van Dijk, and J.C.Dick. High-Frequency Technical Trading:The Importance of Speed[R]. *Tinbergen Institute Discussion Paper*, 2012.
- [9] Menkveld,J.Albert. High Frequency Trading and the New- Market Makers[R]. *Associate Financial Advisor Conference2012 Paper*, 2012.

- [10] Easley, David, Lopez de Prado, and Marcos. Flow Toxicity and Liquidity in a High Frequency World[J]. *Review of Financial Studies*, 2012.
- [11] J.A.Brogaard. High Frequency Trading and Its Impact on Market Quality[R]. *University Kellogg School of Management Working Paper*, 2010.
- [12] I.Aldridge. High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems[M]. *John Wiley Sons, Inc., New Jersey*, 2010.
- [13] Hasbrouck, Joel, and Saar. Low-Latency Trading[R]. *Johnson School Research Paper Series No.35-2010,2012;AFA 2012 Chicago Meetings Paper*, 2012.
- [14] T.A.Hanson. The effects of high frequency traders in a simulated market[R]. *University of Kent State Working Paper*, 2011.
- [15] Esipov,Sergei. Weakly Inefficient Markets, Stability of High-Frequency Trading Strategies[J]. *Tribute to Y B Levinson issue. Lithuanian Journal of Physics*, 2012.
- [16] J.Angel,D.McCabe. Fairness in Financial Markets: the Case of High Frequency Trading[J]. *Journal of Business Ethics*, 112:585–595, 2013.
- [17] Foucault, Thierry, and Moinas. Equilibrium High Frequency Trading[R]. *International Conference of the French Finance Association (AFFI)*, 2011.
- [18] F.Zhang. High-Frequency Trading, Stock Volatility, and Price Discovery[R]. *Working Paper, Yale University*, 2010.
- [19] L.Grillet-Aubert. Equity Trading: A Review of the Economic Literature for the Use of Market Regulators[R]. *AMF,mimeo*, 2010.
- [20] Robert A Jarrow and Philip Prottery. A Dysfunctional Role of High Frequency Trading in Electronic Markets[R]. *Johnson School Research Paper Series*, 2011.

- [21] Kirilenko, A.Andrei, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The Flash Crash: The Impact of High Frequency Trading on an Electronic Market[R]. *Working paper. Available at SSRN: <http://ssrn.com/abstract=1686004>*, 2011.
- [22] I.D.Dichev,K.Huang,D.Zhou. The Dark Side of Trading[r]. *Emory Law and Economics Research Paper*, 2011.
- [23] E.Boehmer,K.Y.L.Fong,J.Wu. International Evidence on Algorithmic Trading[C]. *AFA 2013 San Diego Meetings Paper*, 2013.
- [24] V.H.Martinez ,I.Rosu. High Frequency Traders,News and Volatility[R]. *Research Paper, 2013*, 2013.
- [25] B.Mandelbrot. Fractional Brownian Motions,Fractional Noises and Applications. *SIAM Review*, 10:422–437, October 1968.
- [26] B.Mandelbrot. Fractal Market Analysis:Applying Chaos Theory to Investment and Economics. *New York: John Wiley and Sons,Inc*, 1994.
- [27] S.R.Hurst , E.Platten ,S.T.Rachev. Option Pricing for a Log Stable Asset Price Model[J]. *Mathematical and Computer Modeling*, 29, 1999.
- [28] I.Khindanova and S.Rachev. Stable Modeling of Value at Risk.Mathematical and Computer Modeling. *Mathematical and Computer Modeling*, 34:1223–1259, 2001.
- [29] E.Jacquier,N.G.Polson and P.E.Rossi. Bayesian analysis of stoehastic volatility models with fat-tails and correlated errors[J]. *Journal of econometrics*, 122:185—212, 2003.
- [30] 朱建平, 来升强. 高频数据交易策略与波动性分析[D]. 厦门大学, 2009.
- [31] 镇磊. 基于高频数据处理方法对a 股算法交易优化决策的量化分析研究[D]. 中国科学技术大学, 2010.

- [32] 张彦. 中国股指期货与现货关系实证研究——基于沪深300 股指期货[J]. 价值工程, 30:126–128, 2011.
- [33] 胡天福, 何晓彬. 高频交易在中国证券市场的应用研究[D]. 上海交通大学, 2012.
- [34] 曾勇, 李文峰. 基于技术分析的流动性提供与高频交易策略研究[D]. 电子科技大学, 2012.
- [35] 任品, 苟开桂. 超高频期货市场微观结构噪音实证研究[D]. 西南财经大学, 2013.
- [36] 包思, 郑伟安, 周瑜. 基于macd 的平稳技术指标在高频交易中的应用[J]. 华东师范大学学报(自然科学版), page 152—160, 2013.
- [37] 徐冉, 戴伟辉. 金融高频交易数据的tdf建模及软件实现[D]. 复旦大学, 2013.
- [38] 陈炜, 石诚, 孙梦荣. 基于人工蜂群算法的高频交易策略组合配置研究[D]. 首都经贸大学, 2013.
- [39] 张涤新, 李超. 我国股指期货与股票组合的高频套利[D]. 南京大学, 2013.
- [40] 左顺根. 中国股指期货市场操纵及其对市场功能的影响研究[D]. 华南理工大学, 2013.
- [41] 曹柯, 危启才, 田魁. 基于emd 和rvm 的股指期货价格预测[J]. 武汉工业学报, 2:115–118, 2013.
- [42] 燕汝贞, 李平, 曾勇. 一种面向高频交易的算法交易策略[J]. 管理科学学报, page 88—96, 2014.
- [43] 韩芳婕. 股指期货高频交易模型结构研究[D]. 华南理工大学, 2014.
- [44] 陈贻娟. 量化交易策略模型应用研究[D]. 云南财经大学, 2014.
- [45] 梁满发, 潘晔. 基于高频交易模式下期货投资组合策略研究[D]. 华南理工大学, 2014.

- [46] 田波平, 宋天琪. 基于高频交易策略的bias 技术指标统计分析[D]. 哈尔滨工业大学, 2015.
- [47] 陈艳, 王宣承. 基于变量选择和遗传网络规划的期货高频交易策略研究[D]. 上海财经大学, 2015.
- [48] 徐剑刚, 唐国兴. 我国股票市场报酬与波动的garch—m模型[J]. 数量经济技术经济研究, 12:28–32, 1997.
- [49] 严太华. 上海股市1992——1999周末效应的实证研究[J]. 经济科学, 2:77–85, 2000.
- [50] 杜海涛. Var模型及其在证券风险管理中的应用[J]. 证券市场导报, 8, 2000.
- [51] 唐齐鸣, 陈健. 中国股市的arch效应分析[J]. 世界经济, 3:29–36, 2001.
- [52] 温素彬. 我国股市波动的arch类模型分析[J]. 淮海工学院学报, 11:64–67, 2002.
- [53] 陈守东, 俞世典. 基于garch模型的var方法对中国股市的分析[J]. 吉林大学社会科学学报, 7, 2002.
- [54] 金洪飞. 上海股市(1997.7—2003.5)的波动性、风险溢价和周末效应分析[J]. 清华大学工作论文, 2004.
- [55] 王佳妮, 李文浩. Garch模型能否提供好的波动率预测[J]. 数量经济技术经济研究, 22:74–87, 2005.
- [56] 徐绪松, 侯成琪. 非正态稳定分布条件下的投资组合模型:均值-尺度参数模型[J]. 武汉大学经济与管理学院技术经济及管理研究所, 2006.
- [57] 艾克凤. 股票收益率的非正态性检验与分布拟合[D]. 上海理工大学理学院, 2006.
- [58] 钟蓉萨. 中国股票市场实证统计分析[M]. 北京:中国财政经济出版社, 1999.
- [59] 龚锐, 陈仲常, 杨栋锐. Garch族模型计算中国股市在险价值(var)风险的比较研究与评述[J]. 数量经济技术经济研究, 22:67–81, 2005.

- [60] 马丹. 沪市动态var计量模型分析案例[J]. 北京:中国统计出版社, 2004.
- [61] 武东, 张青, 汤银才. 基于稳定分布的股票收益率的特征分析[J]. 统计与决策, 19:92–93, 2006.
- [62] 武东, 汤银才. 稳定分布及其在金融中的应用[J]. 应用概率统计, 23:434–445, 2007.
- [63] 周孝华, 王小庆. 基于稳定分布的风险价值(var)研究与实证[J]. 统计与决策, 22:65–67, 2009.
- [64] 李志慧, 卢新生. 中国农产品期货市场的多重分形特征及其成因分析——基于mf-dfa分析法[A]. 统计与信息论坛, 26:84–89, 2011.
- [65] 史文静, 高岩. 基于mf-dfa的中国股指期货价格的多重分形实证研究[A]. 财务与金融, 5:11–15, 2014.
- [66] 陈伟娜. 动态稳定因子copula模型在信用衍生品定价中的应用[D]. 郑州大学, 2015.
- [67] Eugene F.Fama. The Behavior of Stock Market Prices[J]. *Journal of Business*.
- [68] Eugene F.Fama. Efficient Capital Markets:A Review of Theory and Empirical Work[J]. *Journal of Finance*, 25:383–417, 1970.
- [69] G.Malkiel Burton. The Efficient Market Hypothesis and Its Critics. *The Journal of Economic Perspectives*, 17:59–82, 2003.
- [70] Andrew W.Lo,A.Craig MacKinlay. Stock market prices do not follow random walks:Evidence from a simple pecification test. *Review of financial studies*.January, 1988.
- [71] E.Peters. Chaos and Order in the Capital Market. *J. WILEY*, 40(4):838–856, December 1991.
- [72] J.P.Nolan. Numerical Calculation of Stable Densities and Distribution Functions. *Common Statist-Stochastic Models*, 13(4):759–774, 1997.

- [73] E.F.Fama and R.Roll. Some Properties of Symmetric Stable Distribution. *Journal of the American Statistical Association*, 66:331–338, 1968.
- [74] J.H.McCulloch. Simple Consistent Estimators of Stable Distribution Parameter. *Communications in Statistics- Simulation and Computiom*, 27:1109–1136, June 1986.
- [75] I.A.Koutrouvelis. Chaos and Order in the Capital Market. *J. WILEY*, 75:918–928, 1980.
- [76] S.M.Kogon ,D.B.Williams. Characteristic Function Based Estimators of Stable Parameters. *Journal of the American Statistical Association*, 66:331–338, 1968.
- [77] J.P.Nolan. *Maximum Likelihood Estimation and Diagnostics for Stable Distributions*. Birkhauser, 2001.

致 谢

又一年海棠花开，我突然意识到自己在郑州大学的学习和生活已经接近尾声，感觉非常不舍。在这里我度过了四年美好的大学时光，有关我青春的记忆都发生在这里。三年的硕士研究生学习和生活，是我成长中最关键的一段时间，我除了跟着老师专心研究学术外，还把学到的知识与实践相结合，这让我感受到了学习的快乐以及知识的价值所在。如今离别在即，人生将踏上另一个转折点，此时的我心潮澎湃，思绪万千，昔日的点点滴滴，跃然而起。非常感谢母校为我们提供了优越的学习环境和条件，感谢数学与统计学院的老师们对我们的谆谆教诲，感谢身边的同学们陪我度过了温暖的研究生生活。我能够顺利读完这三年研究生，离不开大家的关心和帮助，在此一一表示感谢。

首先要感谢我的导师李华老师。李老师既是我学业上的恩师，又是我的精神与心灵上的导师。在学业上，李老师除了传授我们理论知识，还教我们如何把学到的知识与实践相结合。老师给我们提供实习的机会，使得我们学到的知识在实践中真正地发挥价值，这不仅激发了我们学习的热情，还让我们在学习中找到人生的价值及意义。在研究生期间老师不断地给我们锻炼和学习的机会，使得我在这三年增长了很多见识，并有了很大的进步。李老师把我们当做她的孩子一样引导和培养，在生活上给了我们很多的帮助，在精神上给了我们很大的鼓舞。李老师对知识不懈的追求、对工作的拼搏精神以及对生活的态度是我们学习的榜样。在此我要向李老师给以最诚挚的感谢。

此外我要感谢我尊敬的学长郑明浩、我亲爱的两个学姐陈伟娜和郭丽，谢谢你们毫无保留、不耐其烦地传授我学习的方法和经验。感谢我的两个伙伴儿马贵贵和庞丽媛，谢谢你们陪我度过三年充实的研究生生活。感谢郭亚丹、郭昭曼、郭峻诚、吕杨苗四个可爱的学妹学弟，是你们的帮助我的论文才得以顺利完成。我还要感谢我亲爱的朋友们，虽然你们有的已经工作，有的还在读书，每次与你们见面和交流都给了我很大的鼓舞和动力，不管遇到什么困难我都能坚持走下去。最后，我要感谢我的家人是你们多年的坚持和付出，才有了今天的我，你们是最坚强的后盾。

最后，我要感谢所有参加论文评阅及答辩的专家学者们，感谢你们对我的论文提出的宝贵意见。