

分 类 号_____

学号 M201174105

学校代码 10487

密级_____

华中科技大学

硕士学位论文

基于股指期货的套利策略设计

学位申请人： 邹志超

学 科 专 业： 金融硕士

指 导 教 师： 刘方池 副教授

答 辩 日 期： 2013 年 5 月 20 日

**A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Finance**

**Design of Arbitrage Strategy
Based on Stock Index Futures**

Candidate : ZhiChao Chow
Major : Master of Finance
Supervisor : FangChi Liu

Huazhong University of Science & Technology
Wuhan 430074, P. R. China
May,20th 2013

独创性声明

本人声明所呈交的学位论文是我个人在导师的指导下进行的研究工作及取得的研究成果。尽我所知，除文中已标明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到，本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并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。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 _____ 年解密后适用本授权书。

本论文属于

不保密 ☐。

（请在以上方框内打“√”）

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

华中科技大学硕士学位论文

摘要

沪深 300 指数期货推出以来,我国基于股指期货进行套利越来越受到广大的投资者的关注。本文主要研究股指期货期现套利的策略,并且把全文重点放在期现套利的程序化策略设计方面,其中以构建现货组合跟踪沪深 300 指数为全文最主要的创新点。不同于大部分实证研究者在处理期现套利的实证所采用的遗传算法,本文采用更贴近跟踪模拟目的的聚类分析算法来选取成分股参与跟踪模拟,之后同时遗传算法和使用在数值逼近和函数模拟方面表现更为优秀的 BP 神经网络算法来模拟沪深 300 指数。采用聚类分析算法和遗传算法分别对采用成分股跟踪模拟沪深 300 指数实现期现套利的策略和采用 ETF 跟踪模拟沪深 300 指数实现期现套利策略进行程序化实证,并对实证结果进行分析。因此,本文分别讨论了单纯使用成分股跟踪沪深 300 指数进行期现套利和单纯使用 ETF 跟踪沪深 300 指数进行期现套利两个情况。

一方面,在使用成分股跟踪沪深 300 指数时,本文采用基于 2011 年 10 月到 2012 年 10 月共计一整年的沪深 300 指数和股指期货的日度交易数据,分析了股指期货的定价模型和股指期货期现套利的现货组合的构建模型,再者确定无套利区间的参数,最后通过程序化的方法把股指期货期现套利策略展现出来。另一方面,在使用 ETF 跟踪沪深 300 指数时,本文采用基于 2011 年 10 月到 2012 年 10 月的四只 ETF 和沪深 300 指数、股指期货当月连续的交易数据,分析了使用 ETF 跟踪沪深 300 指数再进行期现套利的套利机会及分析了最终的套利效果。

为了更实际的研究套利策略的优劣,本文最终通过 MATLAB 软件把套利策略程序化。研究发现使用 ETF 跟踪指数进行期现套利相比于使用成分股跟踪指数进行期现套利能发现更多的套利机会,能更好的利用期现价格的不一致获取利润。在分析期现套利的同时本文还发现了期货市场的到期日效应和年末效应。

关键词: 股指期货; 期现套利; 聚类分析; 遗传算法; MATLAB

Abstract

Since the CSI 300 index futures arbitrage, the vast number of investors of our country have paid more and more attention on stock index futures. This paper mainly studies the stock index futures arbitrage strategy, for this reason, and full-text focus on procedural strategy of arbitrage design, then the full text of the main innovation is to build a stock portfolio tracking CSI 300 Index. Different from Genetic algorithm which is the main tool of empirical researchers in the empirical treatment of arbitrage, the analysis algorithm to stocks selected in tracking simulation is close to the cluster tracking simulation purposes, and then, this paper use genetic algorithm and BP neural network algorithm which is more excellent in the numerical approximation and function simulation to simulate CSI 300 index. Using cluster analysis algorithm and genetic algorithm Respectively to study track and simulate the CSI 300 index by CSI 300 stocks or ETF. For this reason, this paper discussed two cases: the simple use of index tracking CSI 300 index on arbitrage and simple use ETF to track CSI 300 index on arbitrage.

On the one hand, in the use of index tracking in CSI 300 index, this paper use the data of daily transactions of CSI 300 index and stock index futures from Oct 2011 to Oct 2012, to analysis the model of the pricing model of stock index futures and stock index arbitrage of stock portfolio, in addition to determine parameters of arbitrage-free interval; Finally, through the procedures of the method to the stock index arbitrage strategies show. On the other hand, in the use of the ETF tracking in CSI 300 index, In this paper, based on the daily transaction data of four ETF and CSI 300 index stock futures from Oct 2011 to Oct 2012. This paper has analysed using ETF to track CSI 300 index to arbitrage opportunities and analyze the final arbitrage effect.

In order to study more realistic arbitrage strategies, this paper finally through the MATLAB software to the arbitrage strategy program. Study found that the use of the ETF tracking index arbitrage compared to the index tracking index arbitrage opportunities can be found more, and make better use of current period price not consistent profits. This article also found that futures market's maturity effect and the yearend effect.

Keywords: Stock index futures, Arbitrage, Cluster analysis, Genetic algorithm, MATLAB

华中科技大学硕士学位论文

目 录

摘要.....	I
Abstract.....	II
1 绪论	1
1.1 研究背景和意义	1
1.2 文献综述.....	2
1.3 全文结构和研究方法	7
2 股票指数期货的定价模型研究.....	9
2.1 股指期货定价的经典模型.....	9
2.2 股指期货定价模型的不足.....	11
3 构建期现套利指数现货组合.....	13
3.1 现货组合的构建模型	13
3.2 沪深 300 指数现货组合构建的实证分析.....	14
4 期现套利相关参数及套利区间.....	32
4.1 期现套利相关参数	32
4.2 期现套利无套利区间	34
5 期现套利策略、程序化和套利效果分析.....	38
5.1 期现套利策略流程简介.....	38
5.2 期现套利策略流程的程序化.....	39
5.3 期现套利效果分析	41
6 结论和建议.....	53
6.1 结论.....	53
6.2 不足之处及建议	54

华中科技大学硕士学位论文

致谢.....	56
参考文献	57
附录：沪深 300 成分股.....	61

1 绪论

1.1 研究背景和意义

股指期货是当代资本市场发展到一定阶段之后必然会衍生出来的金融衍生产品。20 世纪 70 年代,由于石油危机的影响,欧美各发达国家的资本市场上各种不稳定因素频生,经济发展波动剧烈,尤其是利率的波动,随着股票价格波动性逐渐增大,投资者急需一种可以对冲股票波动,进行资产套保的金融工具。在这种环境下,股指期货便在不久随即产生了;世界上最早的股指期货合约出现在美国堪萨斯期货交易所,它就是 1982 年 2 月 24 日出现的价值线综合指数期货。紧接其后出现的期货合约是在 1982 年 5 月 6 日由纽约期货交易所(NYBOT)推出的以纽约证券交易所(NYSE)的综合指数为标的的股指期货合约。

股指期货的推出无疑取得了投资者的热捧,对于新生的衍生金融产品股指期货投资者表现出高度的关注和热情,就仅仅价值线综合指数期货,在其刚推出的当年就取得了三十五万张成交量的好成绩。而更为受投资者追捧的标普 500 指数的股指期货合约一经推出之后,其成交量在第一年就突破了一百五十万张。股指期货在美国出现之后,极大的促进了美国金融市场的发展,整个美国期货市场的交易规模迅速增加,各种股指及其期货的流动性也得以提升。其他发达国家和发展中国家也纷纷加入到开发自己的股指期货的行列当中。自股指期货出现之后,股指期货交易已成为全球金融衍生产品市场最具有活力的交易品种之一。虽然股指期货相对于金融期货中其他品种来说出现较晚,但股指期货却不可否认是金融创新中出现的最重要、最成功的产品之一。

我国的两大证券交易所,上海证券交易所和深圳证券交易所,自成立以来,由于我国金融市场和金融体制的不健全,在两大交易所成立之初,各股票的价格波动剧烈,经常大起大落,市场监控机制尚待健全当中的情况下,金融市场上急需能够分散风险的投资工具。最先提出开设股指期货交易的交易中心为海南证券交易中心。海南证券交易中心在没有得到人民银行海南省分行完全授权的情况下,自行开展股

指期货交易，并允许买空卖空双向操作。虽然成交量在推出之后逐月增加，但是交易量仍然不大。

我国的股票市场发展尚未成熟，很多方面都需要进一步完善，由于股票市场的不完善，投机行为常常发生，导致本来就不稳定的股票市场波动不断，因此，我国急需能够稳定股价的投资工具。而随着我国股票市场的不断完善，股票的监管人员和投资者的股市知识和素质的不断提升，加之我国股权分置改革的推进，各项监管体系的完善等因素的影响，使得我国推出股指期货的条件逐步成熟。而此时，国际上其他的市场也已经开展了对我国金融产品设计的开发。其中最早对我国金融产品进行设计并推出相应股指期货的是新加坡交易所，2006年9月份，新华富时推出了以我国A股市场的50只成分股为标的的指数的股指期货。我国推出属于我国市场的股指期货交易迫在眉睫。在这样的环境下，我国为了尽快的开展我国自己的股指期货的交易，2006年9月8日，中国金融期货交易说正式成立并开始运作，中国金融期货交易所成立之后不久我国即开始开展期货合约的仿真交易。经过将近4年的前期准备阶段，于2010年3月26日证监会批准推出股指期货。而我国正式推出第一只我国的金融衍生产品——沪深300股指期货——是在2010年的4月26日，沪深300股指期货的推出填补了我国长期以来金融产品的匮乏，也极大的促进了我国金融市场的活跃度，为完善我国的金融市场和促进我国金融机制的健全带来了新的动力。

我国推出沪深300指数的股指期货之后，有关期货套利的研究也就成为了投资者和理论研究者争相关注的焦点。本文针对股指期货的三大投资工具之一的套利进行研究，希望通过研究股指期货的套利策略来了解期现套利对的实务操作。

1.2 文献综述

对于股指期货套利的研究在国外很早就已经开始，而在国内相对较晚。对于股指期货的研究主要围绕股指期货的定价模型、套利现货组合的构建模型、套利时机和交易成本等方面展开。鉴于此，本文主要从股指期货定价模型、套利现货组合的构建模型以及套利时机和交易成本等方面分别进行文献综述。

1.2.1 股指期货定价模型研究文献

Cornell 和 French(1983)¹两人最早提出期货定价模型,也就是最早出现、并且目前研究者广泛使用的持有成本模型。在他们的模型中,假定市场是完美的,没有交易摩擦、交易成本、投资者可以自由买卖资产。但是,他们两人还进一步进入了股息率和税收等因素,通过实证分析,他们发现股指期货的实际价格明显低于完美市场模型的理论价格。

Modest(1984)²对于股指期货的无套利区间的确定,Modest 认为应该考虑不连续股利支付对股指期货定价的影响。

Mackinlay 和 Ramaswamy(1988)³在研究标普 500 指数期现价格的日内数据时,发现美国股指期货和现货指数价格之间存在较大的差异,并且这种差异随着到期时间的延长而不断增加。

Stoll 和 Whaley(1990)⁴在研究股指期货的一般定价公式时,没有使用价格而是采用收益率进行替代,发现股指期货的收益率刚好等于指数现货的收益率和净持有成本之差的规律。

Hsu 和 Wang(1998)⁵认为套利不是无摩擦的,存在交易成本。他们给予偏微分方程,将预期引入到股指期货的定价模型的推导当中,使得期货的定价模型得以进一步完善。

对于股指期货定价模型的研究国内研究起步比较晚,并且对于股指期货的定价模型国内的研究者大多以持有成本模型作为基础来进行扩展和完善改进。而对于股指期货的套利方面的研究跟国外的研究类似,采用无套利原理和一价定律来确定无套利区间和进行期货套利。

张宗成、苏振华(2004)⁶在研究期货和现货之间的关系时从期货市场和现货市

¹Cornell B and K.R French.The pricing of stock index futures [J].Journal of Futures Markets,1983

²Modest,D.M.On the Pricing of Stock Index Futures[J].Journal of Portfolio Management,1984

³Mackinlay,C,and K. Ramaswamy.Index-futures arbitrage and the behavior of stock index futures prices[J].Review of Financial Studies,1988

⁴Hans R.Stoll, Robert E.Whailey.The Dynamics of Stock Index and Stock Index Futures Returns[J].The Journal of Financial and Quantitative Analysis:1990

⁵Hsu Hsinan,Janchung Wang. The Pricing Model of Stock Index Futures in Imperfect Markets and Analysis of Price Expectation[J].Journal of National Cheng Kung University,1998

⁶张宗成,苏振华.运用股指期货对证券的复合套期保值战略,2004

场的关联关系入手，考虑到两个相关的市场之间可能存在一定的相互影响的关系，称之为均衡关系，而这种相互关系可能影响到股指期货的定价，从而通过研究期货市场和现货市场之间的相互关系来探讨了期现市场的均衡关系对定价误差的各方面的影响。

1.2.2 现货构建研究文献

Rudd(1980)研究现货组合构建模型时，考虑了两种不同的模型：一、剔除股价指数中占权重较小的股票，而仅仅按照指数中权重较大的股票来构建现货组合的“分层模型”；二、选择指数中 β 值为 1 的股票，按照最小化残差风险的原则来选择股票构建现货组合的“最优化模型”。Rudd 选取 1976 年到 1978 年的月度数据作为样本数据进行研究，以 S&P500 指数为模拟对象，实证结果表明，按照上述两种模型构建的投资组合，投资组合包含的公司家数越多，模型效果会越好，但是两种模型对比时，“最优化模型”对指数的跟踪模拟效果更好。

Andrews、Ford 和 Mallison(1986)提出了目前较为常用的三种构建现货组合的模型：完全复制模型、优化抽样模型和分层抽样模型。三种模型中，后两种都为部分复制模型，考虑到完全复制模型的交易成本较高，样本选取模型的随机性太过于严重，分层化模型综合前两种模型的优点，因此在实际操作中使用较为广泛。

Meade 和 Salkin(1989)选取日本股市自 1985 年 1 月到 1987 年 9 月的周度数据作为样本期间，通过构建现货组合来模拟东京证券交易所股价指数(TOPIX)来研究分析现货组合构建模型的效果；再以 1987 年的周度数据作为评估绩效的模拟期，采用优化的样本选取模型，对于指数复制模型采用优化的“分层抽样模型”；提出了四种构建指数现货组合的优化部分复制模型：“未分层估计系数模型”、“分层估计系数模型”、“未分层市值加权模型”和“分层市值加权模型”。实证结果表明，如果按照最小化跟踪误差最为评价构建模型的跟踪模拟效果的判断原则，“未分层估计系数模型”的跟踪模拟效果最好，“分层市值加权模型”的跟踪模拟效果最差。同时实证结果还发现，构建指数现货组合时，如果给出的限制条件越多，跟踪模拟效果可能会越差，也就是说限制条件的增加会导致跟踪模拟效果变差。

Roll (1992)⁷研究最小化指数复制的跟踪误差问题时,以马克维茨的均值-方差模型为基础,定义跟踪误差为组合收益率与标的指数收益率之差的平方和。通过最小化跟踪误差来改善组合对标的指数的跟踪效果。

J.Shapcott(1992)、Dirk Eddebuttel 和 Marseilles(1996)等人在研究跟踪现货指数的最优化指数组合问题研究时引入了遗传算法来寻找最优解。Dirk Eddebuttel 和 Marseilles 则进一步引入混合遗传算法来解决最优化指数组合问题。

Markus R (1999)⁸对跟踪误差进行了重新定义,认为跟踪误差应当是收益率之差,按照 Markus 的定义,跟踪误差为现货组合的平均收益率跟表兄弟指数的收益率之差。基于这样的定义,Markus 提出了四个不同的线性跟踪误差最小化模型,并使用他们分别来进行实证检验,最后对比二次规划模型,发现,线性的跟踪误差定义比传统的二次平方和的定义更为接近现实。

王秀国、邱菀华(2005)对现货投资组合的风险收益一般关系进行了研究,提出了可以统一传统跟踪误差模型和均值方差模型两种不同模型的统一化模型形式。该模型对超额收益率进行了研究,特别是伴随着超额收益率的风险,包括相对风险和总体风险,并且这个统一化的模型对传统跟踪误差优化模型的一些缺点和不足进行了完善,并在统一化模型中引入多因素模型,减少了模型中的待估参数,同时也给出了模型最优解的显示表达式。

1.2.3 套利时机和交易成本研究文献

Neal(1995)使用标普 500 指数期货的一分钟连续报价的高频数据作为研究样本,通过对一分钟的高频数据来进行实证研究分析,并且把套利交易分为三种不同的形式:买方项目(Buy programs)、直接卖出项目(Directsell programs)、卖空项目(Short sell programs),通过实证研究结果,Neal 他指出美国大多数的股票掌握在机构投资者手中,所以卖空限制对于套利的的影响作用并不大。Neal 还发现,期货价格与理论价格的价差越大套利活动并不会同步的越大,他把这解释为套利机会稍纵即逝,投资者很难及时的投入大量的资金进行套利获利。

⁷Roll R. A mean/variance analysis of tracking error[J]. The Journal of Portfolio Management,1992

⁸Markus R, Hans-Jurgen W, Heninz Z. A Linear model for tracking error minimization[J]. Journal of Banking & Finance,1999

华中科技大学硕士学位论文

Draper 和 Fung(2002)在研究指数期货和期权的定价问题时,以伦敦金融时报指数和伦敦金融时报指数期权最为研究对象,并使用 1991 年 10 月到 1998 年 2 月的每日报价和成交价的日内数据作为样本数据,对期货和期权合约的套利策略的实际效果进行实证研究。实证结果表明,如果套利策略的平仓策略使用提前平仓策略,那么获得的套利利润明显高于采用持有到期策略所获得的套利利润,然而遗憾的是提前平仓的机会并不多见。

Jaehalee(2005)使用事后研究法对韩国市场(KOSPI)的 KOSPI200 股指期货的期现套利(主要指套利利润和套利时机)进行了实证研究。实证研究结果表明,期现套利的套利利润和套利机会经常出现后会持续一段时间,由此说明韩国市场 KOSPI 并非有效市场。而采用事前预测的方法进行期现套利研究时仍然能获得套利利润。同时 Jaehalee 还发现韩国市场(KOSPI)的期货市场存在到期日效应。

我国台湾在股指期货方面的研究要早于台股期货的发行。其中,林文政与臧大年(1996)在台股期货发行前就探讨了许多关于台指期货定价和套利机会的问题。但由于当时台股期货还没有发行,林和臧的研究并不能完全真正的阐述台湾台股期货的真实情况。

杨玉川等(2002)对于我国期货市场在制度方面的缺陷和业务方面的不足进行了研究分析,并针对此间种种缺陷,指出我国应当实现期货市场规范化、发展期货和进行期货创新治理等多方面的对策和相应的理论依据。并提出中国应当大力发展金融商品的现货市场和期货市场。

刘持彦 (2010)研究了我国沪深 300 指数股指期货合约在金融期货市场上的交易机制,研究分析了做空机制对于我国 A 股市场上投资者的投资形式和投资资金管理模式的改进作用。做空机制的出现使得原来只能被动做空状态得以改善,投资者可以合理的利用做空来管理投资资金。同时股指期货的出现使得 ETF 期现套利交易策略得以活跃,广大投资者纷纷采用 ETF 模拟沪深 300 指数来进行期现套利。

刘家兴 (2011)在他的论文《沪深 300 股指期货套利功能的实证研究》一文中,通过实证的方法,用现实数据得出了如下结论:套利投资者可以自行考虑影响无套利区间的参数,并在确定无套利区间之后,通过适当的套利操作来获取无风险利润。

1.3 全文结构和研究方法

本文从阐述股指期货各种类型套利交易的基本原理与我国金融市场实际情况开始,进而讨论几种现阶段可行的沪深股指期货合约套利交易策略,随后通过实证和仿真交易数据实例模拟进行相关策略模式的股指期货套利交易和分析。

从全文的结构来看,本文大体上可分为五大部分:

第一部分(第一章):绪论部分介绍全文的研究背景和目的以及进行文献综述。文献综述部分概括国内外一些有关股指期货套利交易策略或者实证研究的经典文献。

第二部分(第二章):第二章首先介绍有关股指期货的基本理论和股指期货套利的基本思想。第二章主要介绍股票指数期货的三种经典定价模型。

第三部分(第三章和第四章):第三章介绍本文进行股指期货期现套利的现货组合的构建模型和现货组合效果的实证分析。第四章则分析股指期货期限套利的相关参数的选择和无套利区间的确定。

第四部分(第五章):模拟股指期货的期现套利。其中,第五章使用 MATLAB 程序把期现套利的模型和确定的无套利区间程序化,并分析其模拟效果。第五章针对套利策略程式和套利机会的效果,分析套利策略的套利效果以及分析现货组合追踪沪深 300 指数的实际效果,分析现货组合对样本内的拟合效果以及对样本外的预测效果,并最后分析整个套利策略的总体套利效果。

第五部分(第六章):对全文进行大体的总结,并对现实的股指期货期现套利提出相应的建议。

本文设计的套利产品相对于其他的套利产品,具有以下一些特点:

本文先从股指期货的相关理论入手,分别介绍了股指期货和我国的沪深 300 指数期货的基本概况。采用部分复制的方法选取 30 支股票拟合沪深 300 指数,并把这 30 支股票打包成为一个现货组合,其对应的期货仍然是沪深 300 指数期货。基于无风险套利理论,在综合考虑交易成本、跟踪成本和持有成本的情况下确定无套利区间。

华中科技大学硕士学位论文

本策略特点：（1）采用部分股票跟踪复制沪深 300 指数，而不是全部复制沪深 300 指数，在交易成本和跟踪误差之间权衡使得总成本最小。（2）既有采用成分股复制跟踪沪深 300 指数，又有采用 ETF 指数复制跟踪沪深 300 指数（实证部分为了更好地比较说明套利策略同样适用了 ETF 跟踪指数进行期现套利而进行了 ETF 跟踪指数期现套利的效果分析），不会夹杂由于混合股票和 ETF 指数而导致的拟合现货的高风险。（3）不同于其他的策略设计研究者对于选股的简单处理，本文专门采用聚类分析法选取成分股票，可以选取到 50 只股票拟合沪深 300 指数，可尽可能地复制出沪深 300 指数的所有特征。（4）不同于其他的研究者常既采用遗传算法选择股票又使用遗传算法来模拟指数，本文采用在数值逼近和函数拟合方面表更为优秀的 BP 神经网络模型对历史数据进行学习训练，再对样本外数据进行预测，能很大程度的预测近期的沪深 300 指数的走势。

本文基于 MATLAB 软件使用聚类分析算法选取 50 只成分股，在最小化跟踪误差的原则下通过 50 只成分股来跟踪模拟沪深 300 指数，并同时得到各只成分股的权重；通过把样本划分为样本内外两部分，使用遗传算法利用样本内数据跟踪模拟得到权重，利用权重预测样本外的现货组合收益率走势，比较沪深 300 指数得到样本外误差以判断总体跟踪模拟效果。最后通过简单的程序化算法得到无套利区间和套利时机，从而实现期现套利。

2 股票指数期货的定价模型研究

股票指数期货是以股票市场上的某一个股票价格指数作为标的物的标准化期货合约。我国的沪深 300 指数期货合约是由中国金融期货交易所推出的，交易代码为 IF，该合约通常由四种：当月、下月和随后的两个季月；到期日为到期月份的第三周的周五，其最低交易保证金未交易金额的 15%。

股票指数期货是联系股票市场与期货市场的金融衍生产品，它既有股票的一些特点，同时还具备期货的特点，除了具有金融期货的一般特点之外，还具有自身的特点：（1）不同于其他的期货，股指期货的标的物并非实物，而是抽象意义上存在股票综合指数。（2）股指期货在到期日并不进行标的物的交割，而是采用现金结算的交割方式。（3）股指期货交易的另一个特点是引入了卖空机制，这也是股指期货进行套期保值规避风险所必须的。（4）股指期货具有交易成本低、以小搏大的高杠杆作用。（5）由于股指期货合约本身可以在期货交易所交易，期货合约的买卖双方都可以通过卖出或买进合约进行提前清算。（6）股指期货交易有较高的透明度。股指期货的所有交易都在有组织的交易场所内进行，随时可以获得不同种类，不同期限的股指期货的买入价、卖出价、成交价、成交量等准确的市场信息。（7）股指期货的流动性高。

2.1 股指期货定价的经典模型

指数套利策略是否能够成功的关键在于是否能确定股指期货的合理价格，从而可以确定出期现套利的无套利区间。从股指期货出现到目前的研究来看，研究期货定价的文献基本上都是以持有成本模型（cost of carry model）和预期理论（expectation theory）两个理论模型为基础，两种理论模型中当以持有成本模型被使用的更为广泛和更为实用。然而以这两个模型为基础放宽一些前提条件而构成的模型也逐步出现。本文主要研究期现套利，因此，本文主要介绍这两个期货定价模型。

2.1.1 持有成本模型

股指期货的持有成本定价模型基于完美市场。这里所说的完美市场是指在市场上交易没有摩擦，没有交易成本，各投资者可以完全自由的买卖任意数量的产品的市场。

为了更好的理解持有成本模型，投资者可以分别采用两种不同的交易策略来进行投资，完美市场给出的结论是：无论采用哪种交易策略，最终的投资收益都一样。

定义符号：

$F(t)$ ： t 时刻股指期货的价格； $F(T)$ ： T 时刻股指期货的价格； $S(t)$ ： t 时刻现货的价格； $S(T)$ ： T 时刻现货的价格。 r ：无风险收益率； d ：股息率； $D(t,T)$ ：时点 t 到时点 T 股利。

策略 1：买入股指期货、同时卖出国债。

投资者在 t 时刻买入一份价格为 $F(t)$ 的股指期货同时卖出价值为 $S(t)$ 的国债。因为买股指期货不需要保证金，买期货现金流为零，卖出国债的现金流为 $-S(t)$ ；所以，在 t 时刻的现金流为 $-S(t)$ ；在 T 时刻，股指期货的价格为 $F(T)$ ，投资者从国债投资中获得 $S(t)(1+r)^{T-t}$ ；综合考虑，投资者总的投资收益为：

$$S(t)(1+r)^{T-t} - S(t) + F(T) - F(t) \quad (2.1)$$

策略 2：买入一揽子股票。

投资者在 t 时刻购买总价值为 $S(t)$ 的一揽子股票， t 时刻的现金流为 $-S(t)$ ；并将其在时刻 T 卖出，假设 t 到 T 时间段内一揽子股票总的股利收入为 $D(t,T)$ ， T 时刻的现金流为 $S(T)+D(t,T)$ ；投资者总的投资收益为：

$$S(T) + D(t,T) - S(t) \quad (2.2)$$

因为在完美市场中，无论何种策略，只要初始投资一样，最终的投资收益也一定相等；而这两种策略初始的投资额都为 $S(t)$ ；因此在 T 时刻两种策略的投资收益应该相等。由此有：

$$S(T) + D(t,T) - S(t) = S(t)(1+r)^{T-t} - S(t) + F(T) - F(t) \quad (2.3)$$

由于到期日，股指期货期货价格收敛于现货价格，即 $F(T)=S(T)$ ，因此：

$$S(T) + D(t, T) = S(t)(1+r)^{T-t} + S(T) - F(t) \quad (2.4)$$

整理后, 可得到:

$$F(t) = S(t)(1+r)^{T-t} - D(t, T) \quad (2.5)$$

上述表达式在考虑利息和股息时都只考虑单期计息, 如果卡率多期的利息和股息计算时, 采用连续复利计算, 上式表达式可以变换为:

$$F(t) = S(t)e^{(r-d)(T-t)} \quad (2.6)$$

由持有成本模型计算得到的期货定价表达式 (2.5) 和 (2.6) 是在完美市场中的定价式。现实市场中, 考虑到各种交易成本和跟踪误差, 期货的定价表达式会更加复杂, 这在本文后面的无套利区间中将要介绍。按照持有成本模型, 如果期货价格高于理论价格, 投资者可以买入现货、同时卖出期货进行建仓, 并到期平仓来获取无风险收益; 称之为正向套利策略。同样, 当期货价格低于理论价格, 投资者可以卖出期货、同时买入现货进行建仓, 并到期平仓那个来获取无风险收益率; 称之为反向套利策略。

2.1.2 预期理论模型

期货价格预期理论基于理性预期, 认为期货价格是未来现货价格的预期, 如果预期为理性预期, 那么在未来现货的价格和期货价格应当相等。按照预期理论的观点来说, 股指期货的市场价格是反映未来股票指数现货价格的预期值, 所以随着期货合约到期日的逐渐临近, 期货价格和现货价格会趋于一致。

依照预期理论模型所描述, 在 T 时刻到期的期货合约在 t 时刻的价格为

$$F(t, T) = E_t(S_T) \quad (2.7)$$

其中, S_T 表示在在期货合约到期日 T 的现货市场价格, 它是一个随机变量, E_t 表示 t 时刻的数学期望。

2.2 股指期货定价模型的不足

上面两种期货定价模型都是在完美市场的假设下构建的, 而实际的期货市场总是存在各种导致市场不完美的因素, 因此, 直接套用这两种定价模型在不完美的市场环境就必然会到处错误的期货价格。从而得出错误的无套利区间, 致使整个套

利策略不能获取利润甚至会亏损。

持有成本模型的前提假设太过苛刻，在实际的市场中很难找到符合完美市场假定条件的现实市场，因而实务操作难以令人信服。而从预期理论来研究股指期货的定价，能给我们提供更多对于市场与投资者之间相互关系的认识，以及对于市场均衡关系和风险回报的分析，然而，对于市场上实际进行交易操作的投资者来说，这样的理论化模型比较难于在实践中量化运用。

3 构建期现套利指数现货组合

套利是投资者进入金融市场进行投资的三大策略（套期保值、套利和投机）之一，基于一价定律和无套利理论。套利的具体含义是指，理论上两种相关度很高的资产或者同一资产为标的物的两种金融衍生产品的价格会逐渐趋于接近或相同。如若不然，套利投资者就将买入价值被低估的一种资产，同时卖出价值被高估的资产，从中获取套利利润。然而大量的套利投资者会使得价值被低估的产品的价格趋于上升，价值被高估的产品的价格趋于下跌，逐渐使得两种产品的价格关系变得合理，套利机会就会逐渐的不复存在。

套利交易有两种基本类型：一种是在期货和现货之间套利，另一种是在不同的气体或合约之间套利。第一种套利，我们称之为期现套利，这种套利是在现货与期货市场上同步进行的；第二种套利，我们称之为价差交易，这种套利方式的交易行为都是发生在期货市场中的。价差套利进一步又可以分为：跨期套利、跨市套利和跨品种套利。而本文主要讨论股指期货的期现套利策略。为此，本文先介绍现货组合的构建模型。

3.1 现货组合的构建模型

进行期现套利需要买（卖）股指期货同时卖（买）期货，买卖股指期货合约在实际操作中比较容易直接实现，而买卖股指期货却在实际操作中较难实现，因为市场的股指期货为反映市场总体发展情况的指数，并不存在实际对应的指数现货，因此为了实现期现套利，投资者必须构建一个现货组合来模拟现货指数。所以期现套利成功的关键就在于是否能够构建出一个拟合效果良好的现货组合。常用的现货组合构建方式有：股票组合模拟指数、指数衍生品（主要指包括金和 ETF）模拟指数以及股票组合和衍生品混合模拟指数。

（1）股票组合模拟指数

利用股票组合模拟指数又可以分为两种：完全复制模拟指数、部分复制模拟指数。其中部分复制模拟指数方法更为常用。常用的部分复制模拟指数法有优化抽样

模拟法和分层抽样模拟法。

完全复制模拟指数法是指直接按照标的指数中各个成分股的权重来买卖成分股构建现货组合；并使用该现货组合与股指期货进行套利交易。在实际的套利交易中，完全复制模拟法实用性较低。

优化抽样法复制指数法是指首先把标的指数各个成分股按照一定的标准进行标示，然后使用计量模型拟合成分股和指数，选择是现货组合跟踪误差最小的成分股及其权重。在实际操作中，使用此种方法的投资者也较少。

分层抽样法复制指数法综合吸收了完全复制法和优化抽样法的优点。该方法首先，按照股票的某一明显的特征或者指标把成分股分成几类相关度极低的类别；然后，从各个类别中选取最具有代表性的股票用来跟踪拟合指数。

（2）指数衍生品模拟指数

在采用股票组合模拟指数时，无论是采用完全复制法还是部分复制法，都需要构建一定数目的股票组合，然而股票的交易成本都相对较高。考虑到基金和 ETF 的交易成本较低，以及 ETF 大多是某一板块的综合指数，因此使用 ETF 来跟踪拟合沪深 300 指数拟合效果从理论上来看会非常优良。

（3）股票加衍生品模拟指数

由于股票交易成本较高，而 ETF 的流动性风险远大于股票的流动性风险。综合考虑两个法制指数方法的优缺点。可以混合使用股票和 ETF 来跟踪拟合指数，各取其优点，可能会模拟出优于单纯使用股票或者 ETF 模拟的效果。

3.2 沪深 300 指数现货组合构建的实证分析

3.2.1 样本选取和分别样本内外检验拟合效果

在此使用的样本数据来自通达信金融终端的沪深 300 指数的收盘价、沪深 300 各个成分股和上证 50ETF、上证 180ETF、深证 100ETF、上证红利 ETF 的每日收盘价。样本空间取自时间段：2011 年 10 月 10 日到 2012 年 10 月 19 日，总共 252 个交易日，可覆盖模拟一年内的股指期货套利。

本节中，首先使用样本数据对沪深 300 成分股对沪深 300 指数的部分复制模拟

华中科技大学硕士学位论文

效果进行实证分析，以确定以后的实证分析中采用多少成分股对沪深 300 指数进行模拟以构建现货组合。

在实证模拟分析之前，本文首先确定沪深 300 成分股对沪深 300 指数进行跟踪模拟的效果优劣的判断标准。通常，利用沪深 300 成分股对沪深 300 指数进行跟踪模拟的效果优劣有两种判断标准：（1）超越指数；（2）跟踪误差最小化。而本文的最终目的为设定期现套利策略，构建现货组合的目的在于得到实际的现货，因此，本文模拟沪深 300 指数不要求超越指数，只需要最小化跟踪误差即可。

在此基础上，首先定义跟踪误差为构建的现货组合和沪深 300 指数的收益率之间的标准差。即如下式：

$$TE = \frac{1}{T} \sqrt{\sum_{t=1}^T (r_t - \sum_{i=1}^n w_i R_{it})^2} \quad (3.1)$$

约束条件为：

$$1) \alpha_i \leq w_i \leq \beta_i;$$

$$2) \sum_{i=1}^n w_i = 1。$$

其中，TE 表示跟踪误差，T 为样本空间时间段长度， r_t 表示沪深 300 指数的日收益率， R_{it} 表示现货组合中各成分股的日收益率， w_i 表示各成分股的权重， α_i 、 β_i 分别表示权重的上下界限制。且 w_i 可表示为：

$$w_i = \frac{p_i q_i}{\sum_{i=1}^N p_i q_i}$$

$$\sum_{i=1}^N p_i q_i = V, \quad V \text{ 表示投资者的投资金额总量。}$$

其中， p_i 表示第 i 只成分股的股价， q_i 表示第 i 只成分股需要买卖的数量。

同时为了更好地检验跟踪模拟的效果，本文把样本分为两部分，前一部分作为样本内数据以跟踪模拟，后一部分作为样本外数据以检验跟踪模拟效果的预测效果。

为了处理数据的方便，样本内数据区前 222 个数据，后 30 个数据作为样本外检验用数据。

在构建现货组合之后，除了需要进行样本内的跟踪拟合效果的分析外，还需要进行样本外的拟合效果的分析。目前对样本进行拟合效果分析的指标主要有：1) 日均跟踪误差；2) 总跟踪误差；3) 偏离度；4) 累积偏离度；5) 组合收益率与指数收益率之间的相关系数。又由于本文判断样本内数据的跟踪模拟效果时采用跟踪误差最小化原则，因此为了检验效果原则的一致性，检验样本外采用日均跟踪误差。即如下式：

$$ATE = \frac{1}{T} \sqrt{\sum_{t=1}^T (r_t - R_{pt})^2} \quad (3.2)$$

其中， R_{pt} 为现货组合的收益率，其表达式为： $R_{pt} = \sum_{i=1}^n w_i R_{it}$ 。

3.2.2 实证检验分析

构建现货组合是进行期现套利的重中之重，本文通过两种不同的算法分别对成分股和 ETF 跟踪模拟沪深 300 指数进行实证检测以寻找最优的构建现货组合的方式，因此，本文将构建四种现货组合：“成分股+BP 神经网络算法”、“成分股+遗传算法”、“ETF+BP 神经网络算法”、“ETF+遗传算法”。通过程序化实证来比较四种组合的跟踪误差大小和各个参与跟踪模拟的股票或者 ETF 的权重。

在进行程序化实证之前，本文首先把样本数据进行分类，本文选取的样本数据来自沪深 300 指数的 300 只成分股、上证 50ETF、上证 180ETF、深证 100ETF、红利 ETF 以及股指期货当月连续的市场价格，时间段为 2011 年 10 月 10 日到 2012 年 10 月 19 日，共 252 个交易日的日交易数据。本文把 2011 年 10 月 10 日到 2012 年 9 月 28 日（考虑到国庆放假的七天，这段时间刚好一个整的交易年度）的共 242 个交易日的日交易数据作为样本内数据，也就是说，这 242 组数据用来进行跟踪拟合得到跟踪误差和权重。而最后 10 个交易日的数据作为样本外数据，用来检验样本外跟踪误差和样本外的跟踪拟合效果。

而本节使用到的三种算法分别为：聚类分析算法、BP 神经网络算法和遗传算法。

华中科技大学硕士学位论文

而为了读者更好的理解本文程序化套利的思路，在此对本文使用的三种基本算法进行阐述。

1、聚类分析算法

聚类分析算法主要用来把 300 只成分股分类，此算法对于大多数的分类问题都能给出很好的结果。

聚类分析算法的具体步骤如下：

Step1: 确定初始聚类中心，可以自定义初始聚类中心，也可以随机产生聚类中心。本文采用随机产生聚类中心的作法。在此说明一下，初始聚类中心对于最终的聚类中心有较大的影响，因此，使用随即产生聚类中心的作法在一定程度上能避免由于人为因素导致的不合理影响。

Step2: 计算除了聚类中心外的所有元素到各个聚类中心的距离，并比较每个元素到各个聚类中心的距离，将其分类到距离最近的聚类中心所在的类当中。待所有元素都进行分类之后，对每个类别进行重新定义聚类中心（选择最中心的元素作为新的聚类中心）。

Step3: 重复 Step2，直到聚类中心不再发生变化（聚类中心不发生变化也即所有元素的分类不会再发生变化）。

Step4: 选取出所有的聚类中心，算法终止。

2、BP 神经网络算法

神经网络算法多用在进行数据拟合和函数逼近的问题当中，在本文使用神经网络算法主要是看重其数据拟合方面的好表现。

其具体步骤如下：

Step1: 数据归一化。因为神经网络算法所处理的数据必须处于-1 到 1 之间，所有要进行神经网络算法，首先必须对原始数据进行归一化已得到符合算法要求的初始数据。

Step2: 创建 BP 神经网络。创建的 BP 神经网络通常包括输入层、隐含层和输出层。并且理论上只需要两个隐含层就可以保证 BP 神经网络算法得到最优的拟合优度。因此，本文创建的 BP 神经网络包含两个隐含层。

Step3: 训练创建的神经网络。完成之后得到跟踪误差、权重和阈值。

Step4: 数据反归一化, 并进行仿真以检测拟合效果以及进行仿真预测。

Step5: 算法终止。

3、遗传算法

遗传算法为常用的一种进化算法, 它的基本思路来源为达尔文的进化论, 并且算法也充分刻画了进化的思想。

它的具体步骤如下:

Step1: 产生初始种群。根据具体问题产生一定的初始种群, 算法默认的初始种群为 20 个, 考虑到本文研究的问题的初始数据较多, 本文定义初始种群为 100。

Step2: 通过交叉 (可以自定义交叉概率和交叉函数) 产生下一代, 并计算每一个后代个体的适应值 (通过适应性函数得到, 适应性函数通常为附加惩罚因子的目标函数, 本文中的目标函数就是最小化跟踪误差)。根据适应值来进行判断后代的优劣, 淘汰适应值较小的后代, 并对高适应值的后代给予一定的优待, 优良的后代将被赋予更高的交叉概率, 也就是说优良的有待参与交叉的机会会更多。

Step3: 重复 Step2, 并且每一代中找到最优的个体都需要对比上代的最优个体, 直到后代中的最优个体在一段时间和代数 (时间和代数可以自定义, 默认为 20 秒和 50 代) 中都没有得到进化为止。

Step4: 给出最优个体、最优个体对应的适应值和个体中各组成元素对应的权重。到此算法终止。

首先进行股票组合的选取, 本文使用的算法为聚类分析算法, 具体的程序化操作均在 MATLAB 软件上进行, 在文章中不一一列举算法的进行过程, 在此文章中, 由聚类分析算法得到的结果如下所示。

当然在进行构建现货组合之前, 需要先选取出参与跟踪模拟的成分股。本文使用聚类分析对样本数据进行分类, 并从中选取 50 只成分股 (当然也可以选取其他数量的成分股, 具体多少数量的成分股最优, 以后的研究者可以具体深入的分析) 用来模拟沪深 300 指数。

首先, 通过聚类分析从样本中选取 50 只成分股用来模拟沪深 300 指数, 得到如

华中科技大学硕士学位论文

下 50 只成分股：

表 3-1：50 只成分股列表

股票代码股票名称	股票代码股票名称	股票代码股票名称	股票代码股票名称
SH600000 浦发银行	SH600123 兰花科创	SH600550 天威保变	SZ000069 华侨城 A
SH600005 武钢股份	SH600125 铁龙物流	SH600583 海油工程	SZ000625 长安汽车
SH600010 宝钢股份	SH600132 重庆啤酒	SH600595 中孚实业	SZ000703 恒逸石化
SH600011 华能国际	SH600143 金发科技	SH600770 综艺股份	SZ000800 一汽轿车
SH600029 南方航空	SH600150 中国船舶	SH600887 伊利股份	SZ000961 中南建设
SH600030 中信证券	SH600153 建发股份	SH601186 中国铁建	SZ000969 安泰科技
SH600031 三一重工	SH600170 上海建工	SH601633 长城汽车	SZ002081 金螳螂
SH600037 歌华有线	SH600183 生益科技	SH601888 中国国旅	SZ002106 莱宝高科
SH600048 保利地产	SH600316 洪都航空	SZ000001 平安银行	SZ002146 荣盛发展
SH600058 五矿发展	SH600376 首开股份	SZ000046 泛海建设	SZ002493 荣盛石化
SH600062 华润双鹤	SH600415 小商品城	SH600109 国金证券	SH600519 贵州茅台
SH600068 葛洲坝	SH600456 宝钛股份	SH600111 包钢稀土	SH600549 厦门钨业
SH600085 同仁堂	SH600516 方大炭素		

得到 50 只成分股之后，本文即将开始四种不同构建现货组合的方式的分别程序化实证检验：

(1) 成分股+BP 神经网络算法

选取到 50 只成分股之后，得到 50 个 252 组数据，把这 252 组数据分为两部分，前 242 组数据用来进行神经网络拟合沪深 300 指数，同时可以得到各只成分股的权重；再使用后 10 组数据进行样本外的拟合预测效果检验。首先来看下样本内数据的拟合效果以及样本外的模拟效果，从图 3-1 来看，跟踪误差在前期非常小，而到后期跟踪误差一直保留着较大，这主要原因在于，图 3-1 所示的是累积收益率，而算法进行拟合时计算的每日收益率，所以累积收益率会把前期的所有误差进行累积，因此后期的跟踪误差看起来很大，其实剔除前期的误差累积后，每日的跟踪误差亦不是很大。累积收益率跟踪误差图如下所示：

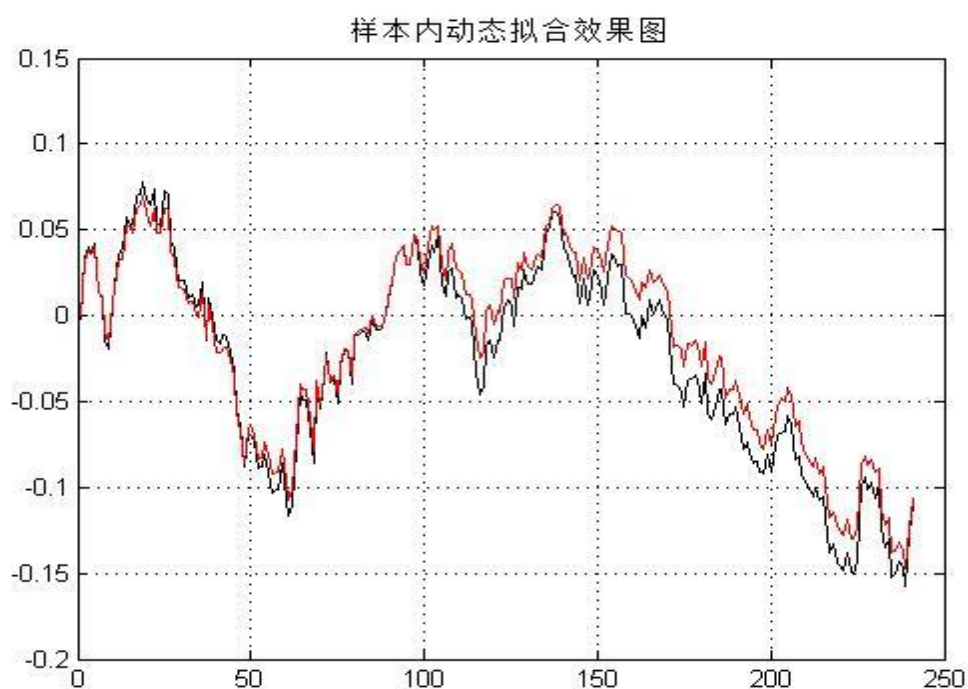


图 3-1: 样本内平均拟合效果图: 累积收益率

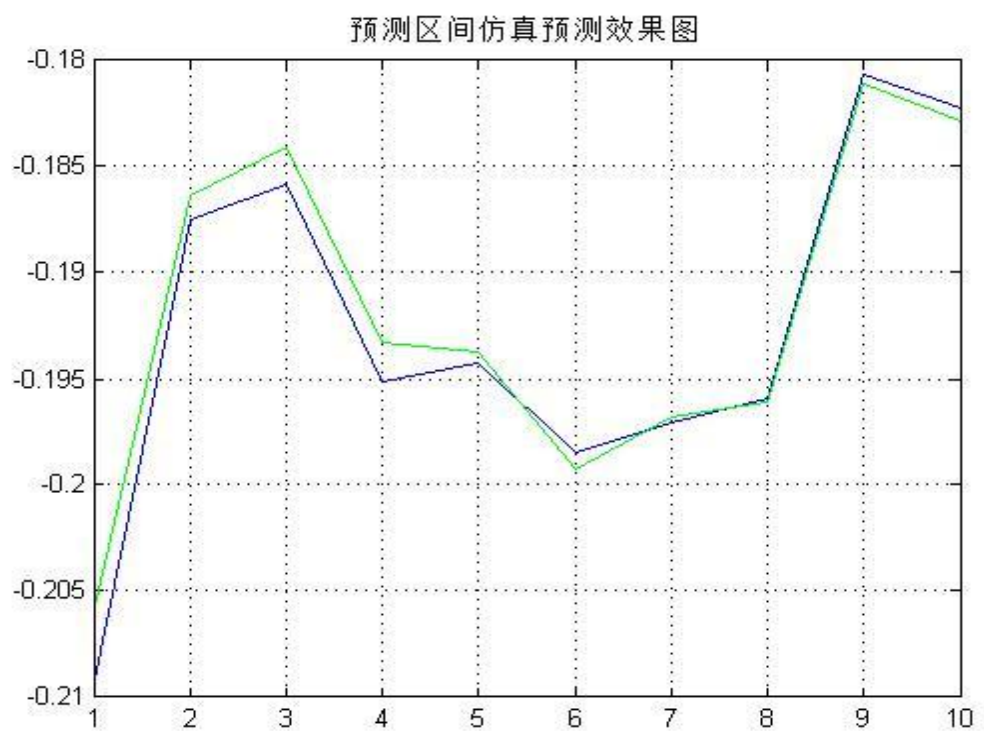


图 3-2: 样本外仿真模拟效果图: 累积收益率

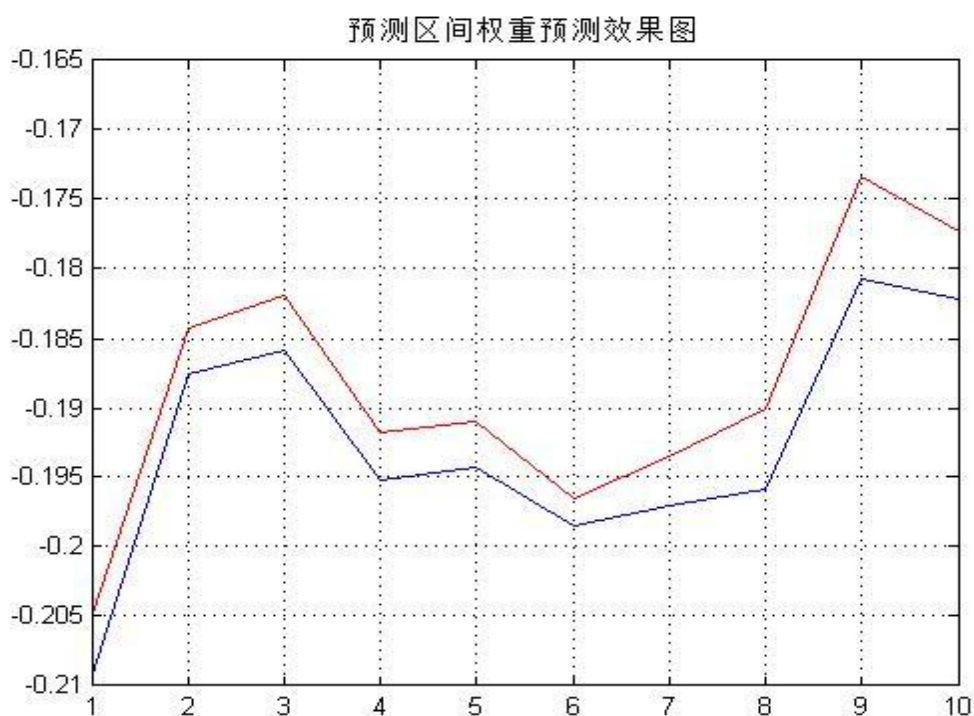


图 3-3: 样本外权重拟合效果图: 累积收益率

三个图中, 图 3-1 中的黑线代表沪深 300 指数累积收益率的真实走势 (偏下), 红线代表拟合的现货组合的累积收益率的走势 (偏上); 图 3-2 中的蓝线代表沪深 300 指数累积收益率的真实走势 (前面偏下, 后面偏上), 绿线代表现货组合的累积收益率的走势 (前面偏上, 后面偏下); 图 3-3 中的蓝线代表沪深 300 指数累积收益率的真实走势 (偏下), 红线代表现货组合的累积收益率走势 (偏上)。

从样本内动态平均拟合效果图来看, 在样本内, 通过神经网络对沪深 300 指数的收益率走势的拟合效果非常好, 几乎可以拟合到绝大部分的指数的变化。

由此得到的样本内跟踪误差为 0.0067, 化为百分比为 0.67%。样本外跟踪误差为 0.0055, 化为百分比为 0.55%。如此小的跟踪误差说明选取的 50 只成分股对沪深 300 指数的收益率的跟踪拟合效果非常令人满意。

注意到样本外预测效果图的横坐标为日期, 纵坐标为指数点数, 从效果图不难看出, BP 神经网络算法通过仿真模拟对样本外的收益率的预测的效果非常好, 几乎可以预测出未来 10 天的收益率的真实水平。相比之下, 通过权重计算得到的对未来 10 天的收益率的预测效果较差一些, 但也能较好的预测未来收益率的大体走势。

华中科技大学硕士学位论文

而通过 BP 神经网络算法同时可以得到各只成分股的权重如下（假设初始投资金额为一亿元人民币）：

表 3-2：BP 神经网络算法得到的 50 只成分股的权重

成分股名称	权重	成交股数	成交手数
SH600000 浦发银行	0.153618733	2026632.36	20266
SH600005 武钢股份	0.096973549	3926054.63	39261
SH600010 宝钢股份	0.006925258	61722.44	617
SH600011 华能国际	0.020706239	86528.37	865
SH600029 南方航空	0.024568803	692078.97	6921
SH600030 中信证券	0.010786287	171483.10	1715
SH600031 三一重工	0.016327946	208530.60	2085
SH600037 歌华有线	0.039040853	856159.05	8562
SH600048 保利地产	0.01478125	44189.09	442
SH600058 五矿发展	0.008381395	159950.28	1600
SH600062 华润双鹤	0.015240556	92535.25	925
SH600068 葛洲坝	-0.00561023	-53635.10	-536
SH600085 同仁堂	0.008508154	50855.67	509
SH600109 国金证券	0.05455118	21561.73	216
SH600111 宝钢稀土	0.018490276	259694.89	2597
SH600123 兰花科创	0.013077332	255916.47	2559
SH600125 铁龙物流	0.017365139	306804.57	3068
SH600132 重庆啤酒	0.00099127	15659.88	157
SH600143 金发科技	0.003337799	49375.72	494
SH600150 中国船舶	0.012855418	196566.02	1966
SH600153 建发股份	0.010549213	87983.43	880
SH600170 上海建工	0.009625314	48223.02	482
SH600183 生益科技	0.006005376	51196.73	512
SH600316 洪都航空	0.019416945	349854.87	3499
SH600376 首开股份	0.037998173	780249.97	7802
SH600415 小商品城	0.00708483	73190.39	732
SH600456 宝钛股份	0.030957979	517691.96	5177
SH600516 方大炭素	-0.01202041	-227228.86	-2272
SH600519 贵州茅台	0.016203976	88065.09	881
SH600549 厦门钨业	0.046533433	879649.01	8796
SH600550 天威保变	0.006485049	36888.79	369
SH600583 海油工程	0.025949107	87194.58	872
SH600595 中孚实业	0.008338258	41196.93	412
SH600770 综艺股份	0.065150871	485114.45	4851
SH600887 伊利股份	0.009644831	265697.82	2657
SH601186 中国铁建	0.018486999	312280.39	3123
SH601633 长城汽车	0.028433549	158757.95	1588
SH601888 中国国旅	0.002096336	22043.49	220
SZ000001 平安银行	0.02723038	481102.12	4811

SZ000046 泛海建设	0.014959711	124353.37	1244
SZ000069 华侨城 A	0.024222239	154873.65	1549
SZ000625 长安汽车	0.0082195	119817.78	1198
SZ000703 恒逸石化	0.019502304	91560.11	916
SZ000800 一汽轿车	0.010174873	96352.96	964
SZ000961 中南建设	0.009337494	73350.30	734
SZ000969 安泰科技	0.008873195	23107.28	231
SZ002081 金螳螂	0.012695886	130885.42	1309
SZ002106 莱宝高科	-0.00649923	-37786.25	-378
SZ002146 荣盛发展	0.011412927	30605.86	306
SZ002493 荣盛石化	-0.00798631	-66002.56	-660

其中，各只成分股的权重有正有负，正的权重表明构建现货组合需要买入相应的该成分股，负的权重表明构建现货组合需要卖出相应的该成分股。

(2) 成分股+遗传算法

同样使用聚类分析算法选取 50 只成分股之后，得到 50 个 252 组数据，把这 252 组数据分为两部分，前 242 组数据用来进行神经网络拟合沪深 300 指数，同时可以得到各只成分股的权重；再使用后 10 组数据进行样本外的拟合预测效果检验。首先来看下遗传算法的全过程图示：

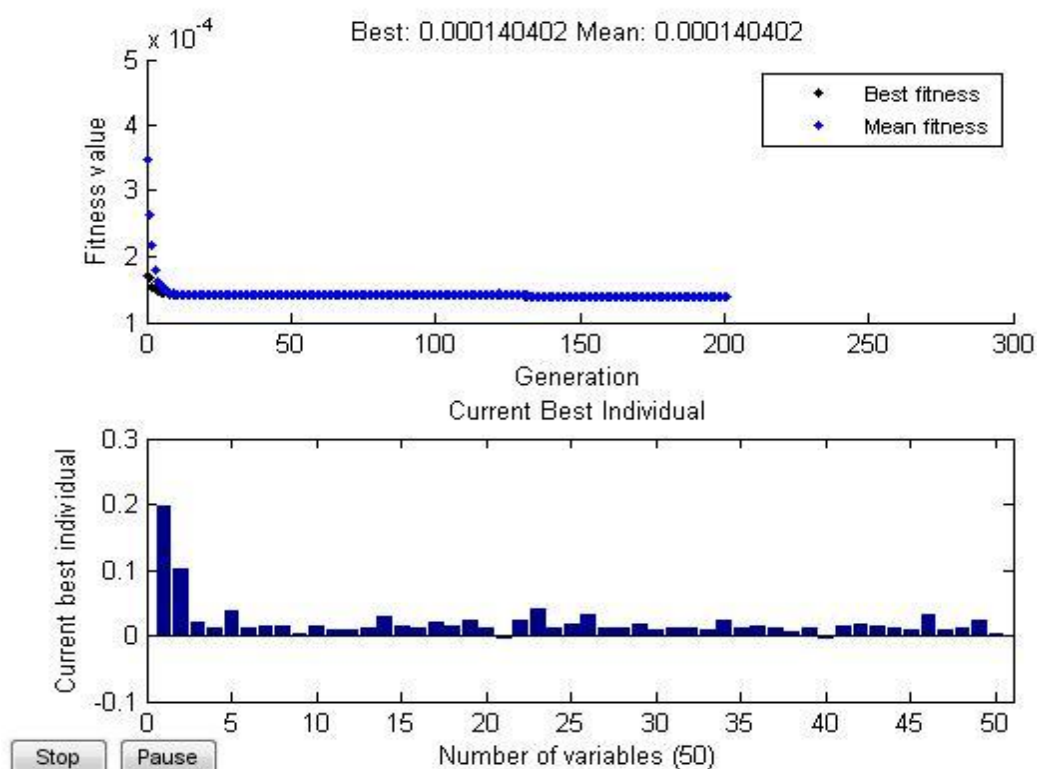


图 3-4：成分股+遗传算法的全过程图示

图 3-4 中可以看出，图的上半部分表示遗传算法得到最优种群的过程，下半部分表示最优种群对应的各只成分股的权重。

接下来可以对比查看样本内数据的拟合效果以及样本外的模拟效果，如下图：

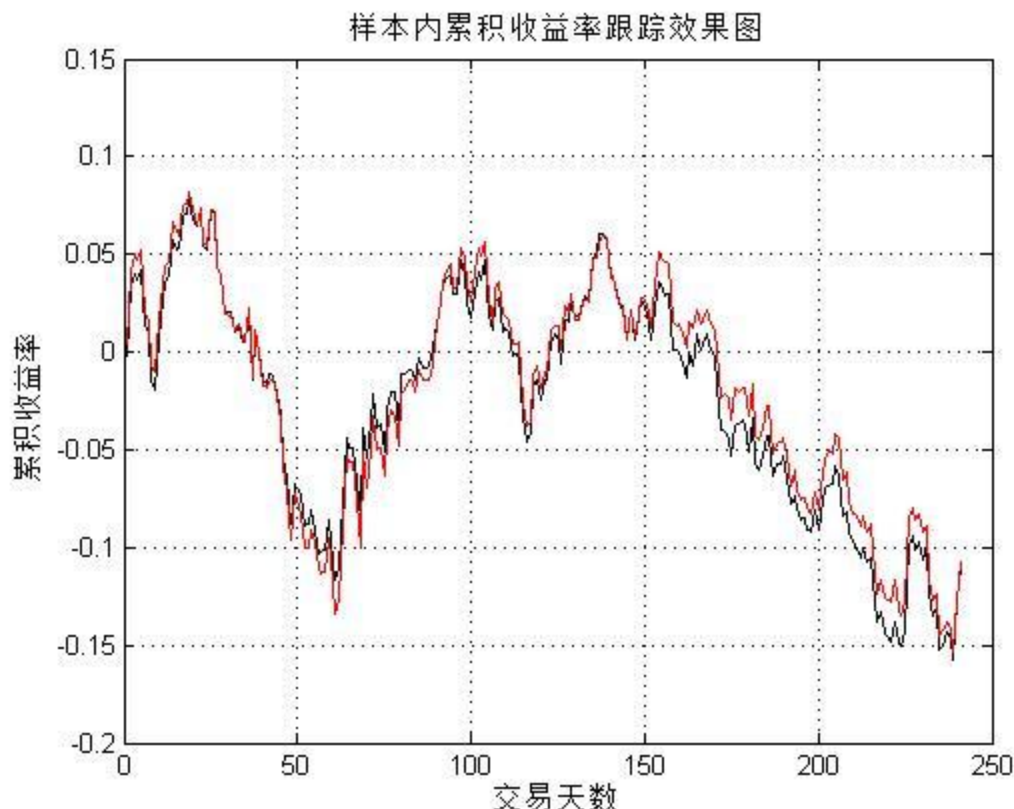


图 3-5：样本内跟踪拟合效果图：累积收益率

图 3-5 中，黑线代表沪深 300 指数的收益率走势（偏下），红线代表现货组合的收益率走势（偏上）。不难发现使用遗传算法对成分股进行模拟拟合的效果也非常好。同样的，图中给出的累积收益率，所以前期的跟踪误差会被累积到后期，导致从图形来看后期的跟踪误差较大，实际上每日的跟踪误差应当是每日累积跟踪误差剔除所有前期的累积跟踪误差之后的剩余，这样就不难看出，每日的跟踪误差实际很小。

图 3-6 中，黑线代表沪深 300 指数的收益率走势（偏下），红线代表现货组合的收益率走势（偏上）。对比使用 BP 神经网络算法的拟合效果，遗传算法的样本外拟合效果比神经网络算法的权重拟合效果要好的多，但相比于神经网络的仿真拟合效

果则不相上下。

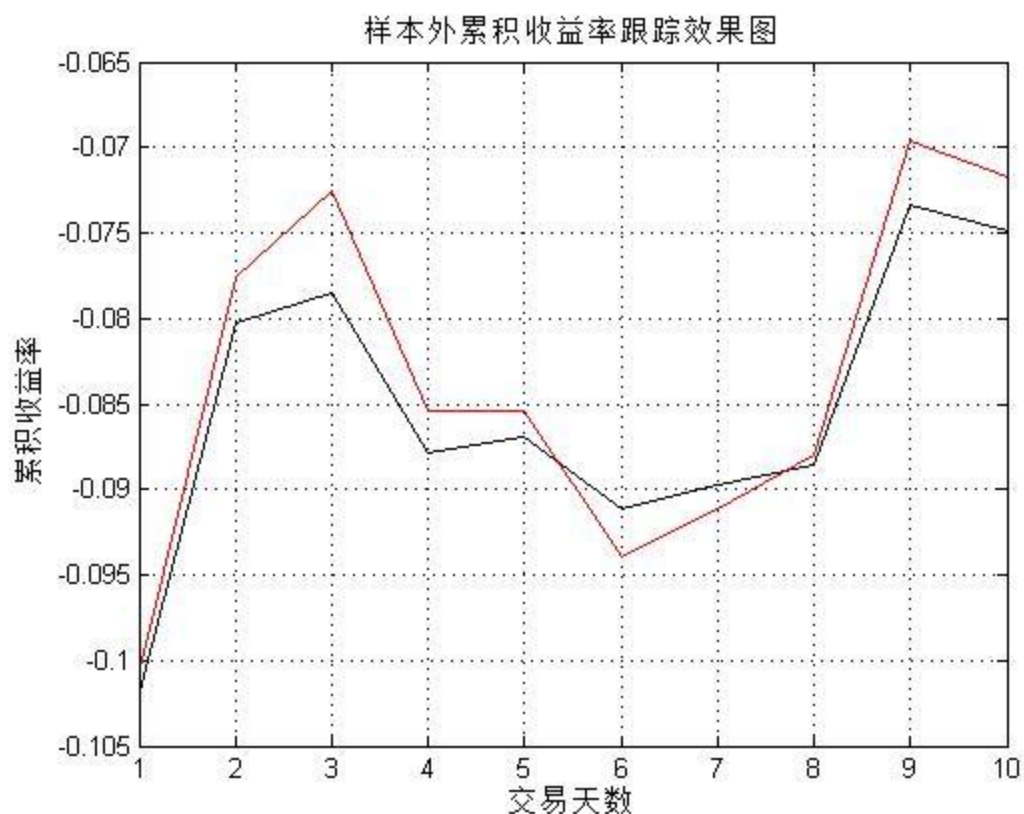


图 3-6: 样本外拟合效果图: 累积收益率

遗传算法得到的样本内的跟踪误差为 1.4040×10^{-4} , 近似为 0.014%; 样本外的跟踪误差为 7.9434×10^{-4} , 近似为 0.079%。对比于神经网络算法的跟踪误差, 遗传算法得到的跟踪误差无论样本内还是样本外都要更小。

遗传算法得到的权重如下 (同样假设投资者的初始投资金额为一亿元人民币):

表 3-3: 遗传算法得到的成分股的权重

成分股名称	权重	成交股数	成交手数
SH600000 浦发银行	0.19817506	2614446.65	26144
SH600005 武钢股份	0.10124142	4098842.95	40988
SH600010 宝钢股份	0.02046186	182369.51	1824
SH600011 华能国际	0.01018162	42547.52	425
SH600029 南方航空	0.03803741	1071476.29	10715
SH600030 中信证券	0.01167501	185612.26	1856
SH600031 三一重工	0.013873	177177.58	1772
SH600037 歌华有线	0.01355837	297332.61	2973
SH600048 保利地产	0.00397181	11873.87	119

华 中 科 技 大 学 硕 士 学 位 论 文

SH600058 五矿发展	0.01505571	287322.62	2873
SH600062 华润双鹤	0.00985581	59841.00	598
SH600068 葛洲坝	0.00900875	86125.76	861
SH600085 同仁堂	0.01167501	69784.88	698
SH600109 国金证券	0.02929257	11578.09	116
SH600111 包钢稀土	0.01578747	221734.14	2217
SH600123 兰花科创	0.01147553	224570.03	2246
SH600125 铁龙物流	0.02111771	373104.50	3731
SH600132 重庆啤酒	0.01453883	229681.36	2297
SH600143 金发科技	0.02222098	328712.70	3287
SH600150 中国船舶	0.01167501	178516.99	1785
SH600153 建发股份	-0.0059853	-49919.00	-499
SH600170 上海建工	0.02370428	118758.91	1188
SH600183 生益科技	0.04158811	354544.81	3545
SH600316 洪都航空	0.01181834	212943.12	2129
SH600376 首开股份	0.01732068	355660.76	3557
SH600415 小商品城	0.03310127	341955.28	3420
SH600456 宝钛股份	0.01189359	198889.47	1989
SH600516 方大炭素	0.01167501	220699.64	2207
SH600519 贵州茅台	0.01723229	93653.74	937
SH600549 厦门钨业	0.00772009	145937.35	1459
SH600550 天威保变	0.01167501	66410.76	664
SH600583 海油工程	0.01075777	36148.42	361
SH600595 中孚实业	0.00903804	44654.34	447
SH600770 综艺股份	0.02287416	170321.34	1703
SH600887 伊利股份	0.01144871	315391.41	3154
SH601186 中国铁建	0.01490623	251794.46	2518
SH601633 长城汽车	0.01167501	65187.11	652
SH601888 中国国旅	0.00602937	63400.36	634
SZ000001 平安银行	0.01179955	208472.61	2085
SZ000046 泛海建设	-0.0064543	-53651.30	-537
SZ000069 华侨城 A	0.0142081	90844.64	908
SZ000625 长安汽车	0.01693675	246891.41	2469
SZ000703 恒逸石化	0.01414046	66387.12	664
SZ000800 一汽轿车	0.01167501	110558.82	1106
SZ000961 中南建设	0.00963553	75691.53	757
SZ000969 安泰科技	0.03150761	82051.07	821
SZ002081 金螳螂	0.00854175	88059.25	881
SZ002106 莱宝高科	0.01167501	67877.97	679
SZ002146 荣盛发展	0.02193333	58818.27	588
SZ002493 荣盛石化	0.00304954	25202.80	252

对比两种算法的权重，虽然成分股的权重值不尽相同，但是同样的成分股在神经网络算法中得到高权重在遗传算法中同样会得到高权重。可见两种算法都能很好的

跟踪模拟沪深 300 指数。

（3）ETF+BP 神经网络算法

本文选取四只 ETF：上证 50ETF、上证 180ETF、深证 100ETF 和红利 ETF 来跟踪模拟沪深 300 指数，其原因在于这四只 ETF 跟沪深 300 指数的相关度非常高，具体的相关系数可参看第五章的相关系数表。

对于样本的划分方法跟成分股的划分方法一样，前 242 个交易日的数据作为样本内数据，最后 10 个交易日的数据作为样本外检验用数据。

首先，来看样本内的跟踪拟合效果和样本外的模拟拟合效果。

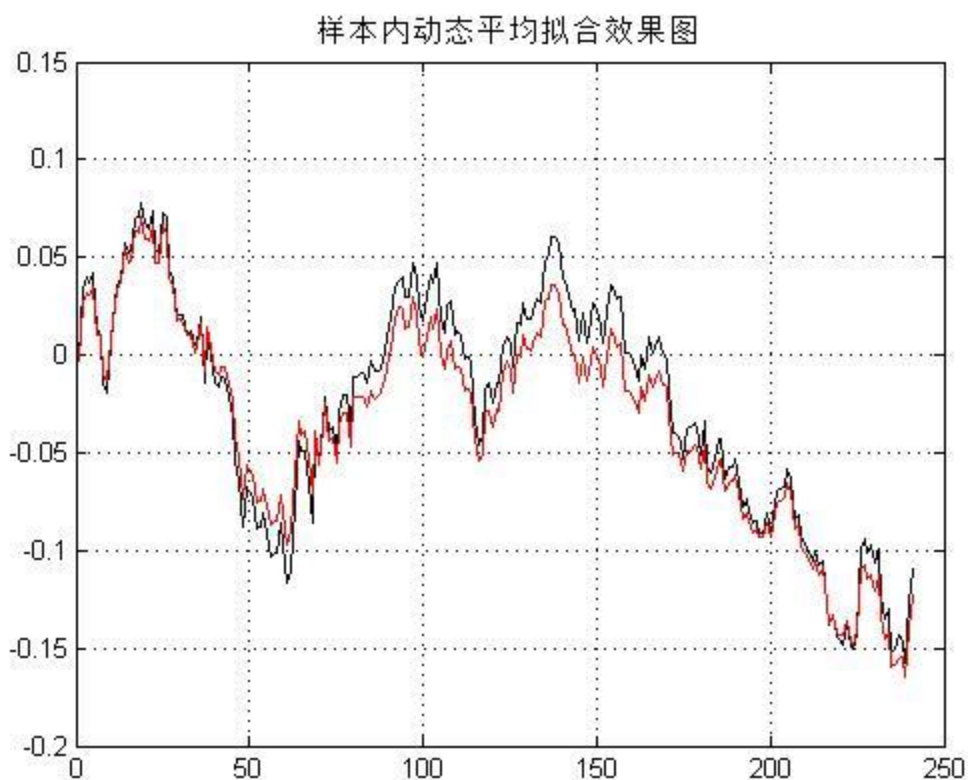


图 3-7：样本内跟踪拟合效果图：累积收益率

其中，黑线代表沪深 300 指数累积收益率的真实走势（偏上），红线代表现货组合的累积收益率的走势（偏下）。从图示来看，使用四只 ETF 对沪深 300 指数的跟踪拟合效果较好，基本上能拟合到沪深 300 指数的全部走势。

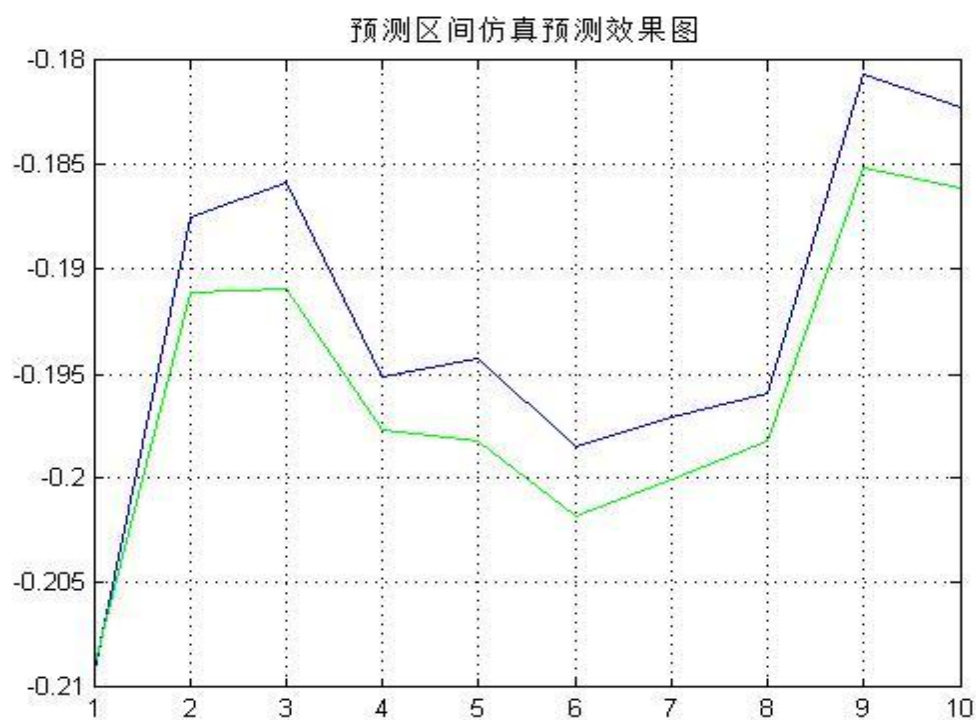


图 3-8: 样本外仿真拟合效果图: 累积收益率

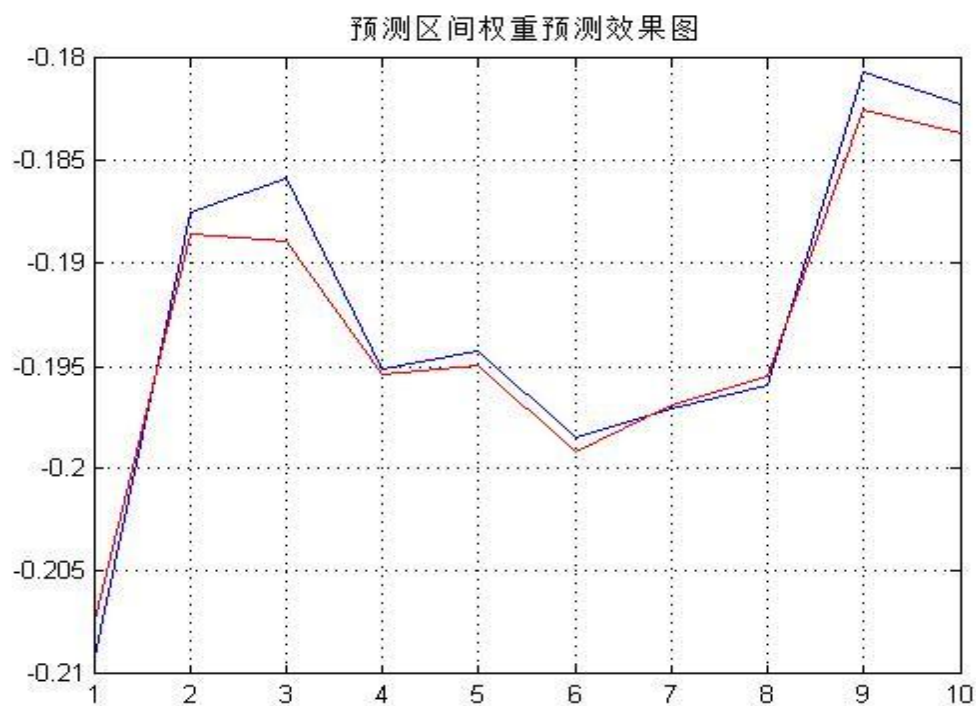


图 3-9: 样本外权重拟合效果图: 累积收益率

其中，图 3-8 和图 3-9 中的蓝线代表沪深 300 指数的累积收益率的真实走势（均为偏上），图 3-8 中的绿线代表仿真拟合的现货组合的累积收益率的走势（偏下），图 3-9 中的红线代表权重拟合的现货组合的累积收益率的走势（偏下）。

从两种样本外拟合效果图来看，不同于成分股拟合的情况，ETF 拟合时采用权重对样本外的收益率的走势进行拟合预测能得到更好的效果。

ETF+BP 神经网络算法得到的样本内的跟踪误差为 0.0057，化为百分比为 0.57%，样本外的跟踪误差为 0.0040，化为百分比为 0.40%。相比于成分股拟合的情况，跟踪误差在样本内外都要稍小一些。

ETF+BP 神经网络算法得到的权重如下（假设初始投资金额为一亿元人民币）：

表 3-4：BP 神经网络算法得到的各只 ETF 的权重

ETF	权重	成交份数	成交手数
深证 100ETF	0.143593	25459718.96	25460
上证 50ETF	0.012744	759920.64	760
上证 180ETF	0.444311	86610299.63	86610
红利 ETF	0.399352	22820141.65	22820

（4）ETF+遗传算法

ETF 及其样本划分方法不变，首先来看遗传算法的全过程图示和各只 ETF 的最优权重图示，其次列示样本内的跟踪拟合效果和样本外的模拟拟合效果图：

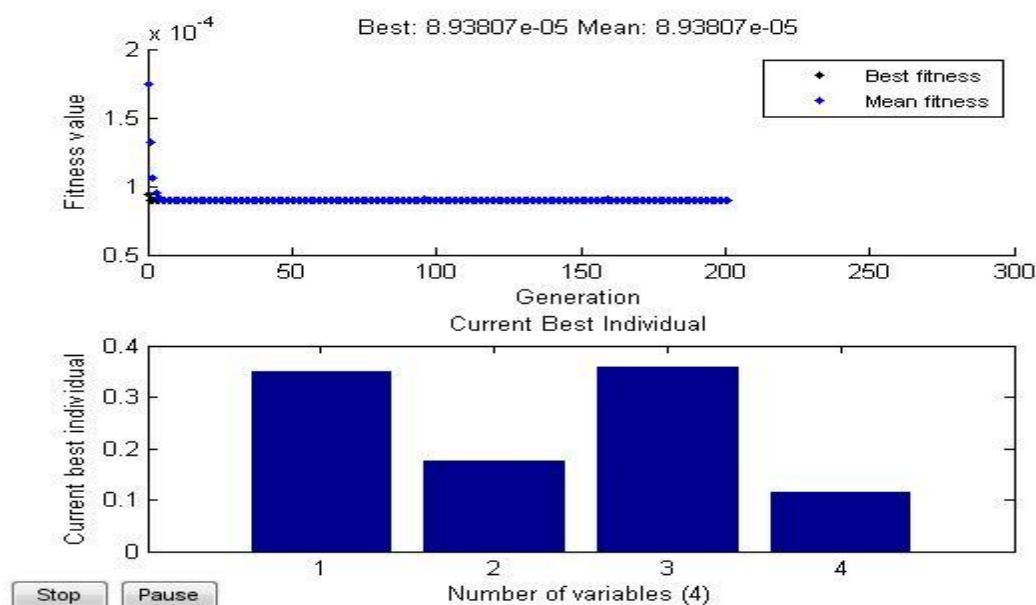


图 3-10：遗传算法的全过程图示和最优权重图示

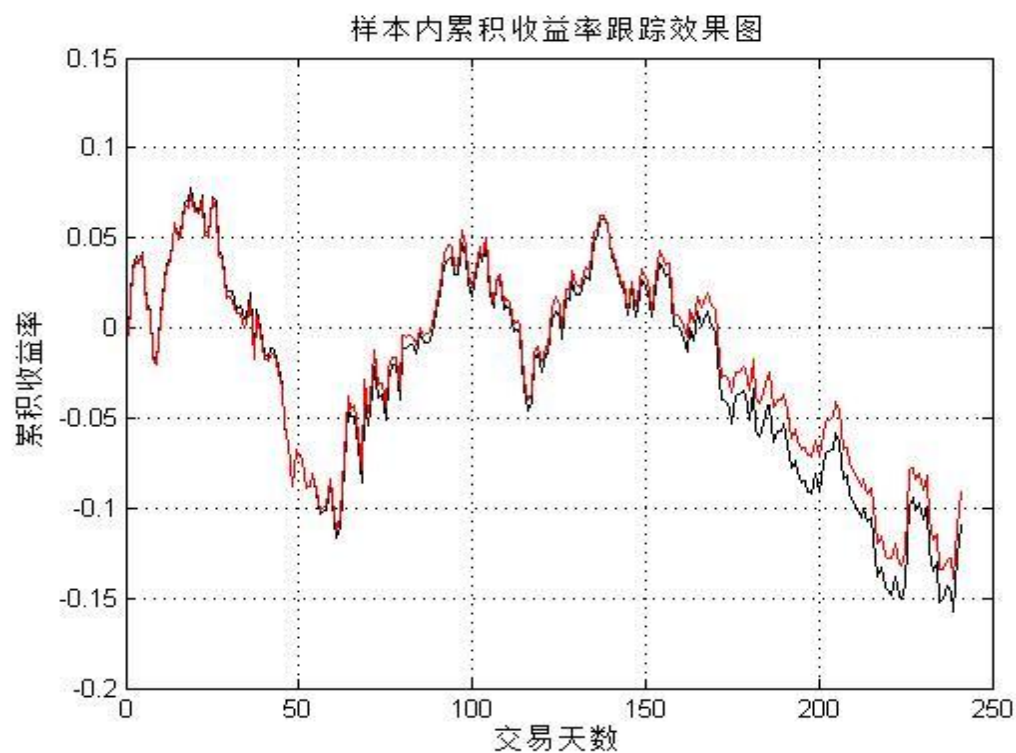


图 3-11: 样本内跟踪拟合效果图: 累积收益率

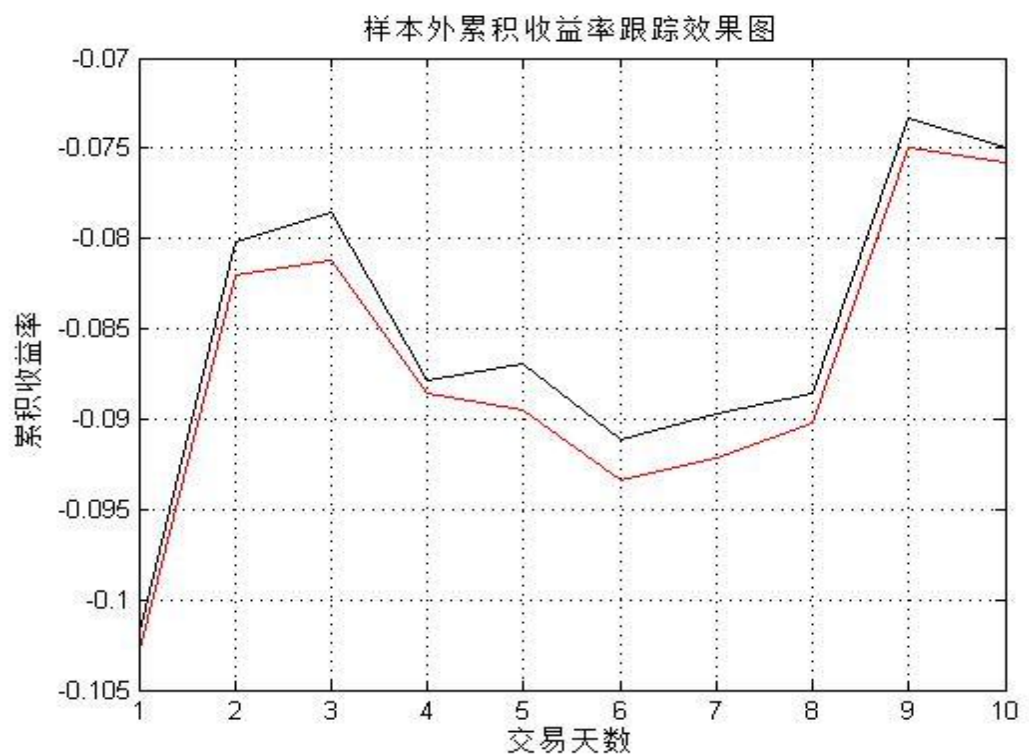


图 3-12: 样本外权重拟合效果图: 累积收益率

华中科技大学硕士学位论文

其中,图 3-11 中黑线代表沪深 300 指数累积收益率的真实走势(偏下),红线代表现货组合的累积收益率的走势(偏上);图 3-12 中黑线代表沪深 300 指数的累积收益率的真实走势(偏上),红线代表现货组合的累积收益率的走势(偏下)。

ETF+遗传算法得到的样本内跟踪误差为 8.9381×10^{-5} ,近似为 0.00894%,样本外跟踪误差为 3.2463×10^{-4} ,近似为 0.03246%。相比于成分股+遗传算法得到的跟踪误差,采用 ETF+遗传算法得到的跟踪误差无论样本内还是样本外都要更小。

最后查看得到的权重(假设初始投资金额为一亿元人民币):

表 3-5: 遗传算法得到的 ETF 权重

ETF	权重	成交份数	成交手数
深证 100ETF	0.349894	62038000.73	62038
上证 50ETF	0.174931	10431192.80	10431
上证 180ETF	0.358764	69934497.98	69934
红利 ETF	0.115411	6594894.08	6595

从权重来看,遗传算法得到的权重跟神经网络得到的 ETF 的权重完全不同,就连权重大小排序都不同,通过多次运行神经网络算法和遗传算法发现,遗传算法的到的权重值很稳定,每次运行得到的权重都相差不多,而通过神经网络算法得到的全权重值波动较大,每次得到的结果都不尽相同。由此可见,遗传算法在得到 ETF 的权重方面表现更为优秀。

(5) 对比分析四种方式的优劣

通过对比四种方式的跟踪误差可得到表:

跟踪误差	成分股+神经网络	成分股+遗传算法	ETF+神经网络	ETF+遗传算法
样本内	0.0067	1.4040×10^{-4}	0.0057	8.9381×10^{-5}
样本外	0.0055	7.9434×10^{-4}	0.0040	3.2463×10^{-4}

对比四种方式的跟踪误差不难发现,无论是使用成分股来跟踪沪深 300 指数还是使用 ETF 来跟踪沪深 300 指数,通过遗传算法来拟合数据,都能在最小化跟踪误差方面表现更优。

而如前文所说,神经网络得到的权重不太稳定,同样的数据,每次运行得到的权重相差较大,而使用遗传算法得到权重较稳定,每次运行的结果都相差无几。

由此可见,在最小化跟踪误差和得到稳定权重的原则下,构建现货组合的最佳方式为使用遗传算法,所以在本文后面的程序化实证效果分析中采用遗传算法。

4 期现套利相关参数及套利区间

4.1 期现套利相关参数

期现套利基于一价定律和无套利原则；由持有成本模型，在完美市场下期货价格和现货价格有如下关系式：

$$F(t) = S(t)e^{(r-d)(T-t)} \quad (4.1)$$

其中， $S(t)$ 为现货价格， $F(t)$ 为期货价格， r 为无风险利率， d 为股票股息率， T ， t 分别为到期时间和当前时间。

然而，实际的股指期货市场并不是完美市场，市场上存在市场交易摩擦，期货交易需要保证金，借贷利率不同，卖空限制等等限制条件下，期货价格和现货价格的关系就不能简单的由如上的关系式表示。因此，在进行期现套利时，投资者必须要考虑影响期现套利的各种成本和相关参数。

4.1.1 期现套利中的现金股利

在持有成本定价模型中，持有成本误差主要来自于借贷利率和股息率。因此，当 r 大于 d 时，股指期货价格高于现货价格，投资者可以进行正向套利获取价差收益；反之，当 r 小于 d 时，股指期货价格低于现货价格，投资者可以进行反向套利获取价差收益。

在股指期货套利的过程中，如果使用成分股来构建现货组合，则由于成分股会不定期的发放股利而现货指数并没有股利的发放，从而现货组合的价值中必然含有现金股利的价值，所以现金股利必然影响到套利利润的大小。

我国沪深 300 指数的成分股的股利发放呈现集中在 5、6、7、8 几个月份，根据 2008 年到 2010 年的沪深 300 指数成分股的股利发放频率和比率，可以得到如下的图示：

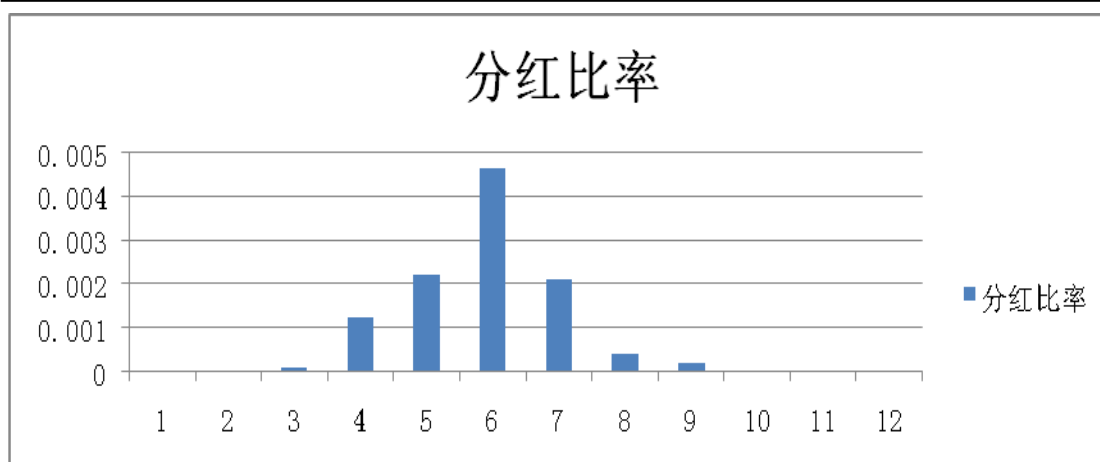


图 4-1: 沪深 300 指数成分股股利发放平均比率

4.1.2 期现套利中的交易成本

交易成本是指在进行股票和期货交易过程中产生的交易费用。

我国沪深 300 指数成分股交易成本主要包括印花税和佣金等。⁹印花税税率为成交金额的 1%，向卖方单边征收。券商交易佣金：最高为成交金额的 3%，最低 5 元起，单笔交易佣金不满 5 元按 5 元收取。上海证券交易所规费为千分之 0.18，深圳证券交易所规费为千分之 0.2175。本文选取的股票绝大多数来自上海证券交易所，因此，本文采用的现货的交易费用为：印花税 0.1%，佣金 0.018%。期货的固定费用为：佣金 0.01%。而我国 ETF 基金的交易成本相对于成分股的交易成本则更低，通常主要包括交易佣金，没有印花税，而本文选取的 ETF 的交易佣金比例为 0.05%。

4.1.3 期现套利中的其他成本

其他成本主要指冲击成本和跟踪误差等。

冲击成本又称流动性成本，是指一定数量的委托（订单）迅速成交时对价格的影响。股票组合冲击成本较高。以我国为例，经过测算，沪深 300 成分股组合总冲击成本一般在 5-6% 左右，该冲击成本在 80% 的概率下很难被突破。冲击成本最小的 20 只股票总冲击成本为 2-3%。¹⁰金元证券的研究小组利用 level-1 行情的成交明细数据，计算出每手当月连续合约的冲击成本为 0.005%，本文直接使用该冲击成本来

⁹数据结果来自百度百科：股票的交易费用。

¹⁰数据来自百度百科：冲击成本。

进行研究分析。为了简单处理方便，本文直接取现货的冲击成本为 0.25%。期货的冲击成本为 0.005%。

跟踪误差主要来自构建的现货组合和真实标的指数之间的误差大小。跟踪误差主要用来衡量构建的现货组合对现货指数的跟踪效果的好坏。如果跟踪误差小，则说明构建的投资组合能很好的反映现货组合。所以跟踪误差的大小关系到整个套利策略的成败。

4.2 期现套利无套利区间

在确定无套利区间的上下限之前，为了方便叙述和推导去套利区间的上下限，本文给出如下的记号：

S_t ：现货指数在 t 时刻的市场指数点； S_T ：现货指数在 T 时刻的市场指数点；

F_t ：期货指数在 t 时刻的市场指数点； F_T ：期货指数在 T 时刻的市场指数点；

C_{ss} ：卖出指数现货组合的交易成本； C_{bs} ：买入指数现货组合的交易成本；

C_{sf} ：卖出指数期货的交易成本； C_{bf} ：买入指数期货的交易成本；

r_b ：借入资金的成本； r_f ：借出资金的成本；

M_f ：期货初始保证金比例； M_s ：融券保证金比例；

$D(t,T)$ ： t 时刻到 T 时刻指数成分股的现金股利折算成的指数点在 T 时刻的值。

(1) 正向套利策略确定无套利区间的上限

当指数期货合约价格被高估时，投资者可以采用正向套利策略，在考虑交易成本之后，获取套利利润，由此可以推导出无套利区间上限，也就是说通过套利交易能获得正的无风险利润。

表 4-1: 正向套利的现金流

t 时刻	现金流	T 时刻	现金流
买入指数现货	$-S_t$	卖出指数现货	S_T
买入指数现货成本	$-S_t C_{bs}$	卖出指数现货成本	$-S_T C_{ss}$
		股息收入	$D(t, T)$
卖出股指期货	0	买入股指期货平仓	$F_t - F_T$
卖出期货成本	$-F_t C_{sf}$	买入期货成本	$-F_T C_{bf}$
支付期货保证金	$-F_t M_f$	收回期货保证金	$F_t M_f$
借款	$S_t(1 + C_{bs}) + F_t(M_f + C_{sf})$	还款	$-[S_t(1 + C_{bs}) + F_t(M_f + C_{sf})]e^{r_b(T-t)}$
净现金流	0	投资收益	$F_t + F_t M_f - F_T C_{bf} - S_T C_{ss} + D(t, T) - [S_t(1 + C_{bs}) + F_t(M_f + C_{sf})]e^{r_b(T-t)}$

通过现金流量表（表 4-1），我们不难看出 t 时刻的现金流净值为 0，由无套利理论可知，在 T 时刻的投资收益也应当为零，如果上述的投资策略在 T 时刻得到的投资收益大于零，那么我们就可以进行正向套利，得到大于零的无风险利润；因此，是投资收益等于零的 F_t 就是无套利区间的上限，根据预期理论，T 时刻现货价格和期货价格在理论上是相等的，于是简单的推导就可以得到股指期货无套利区间的上限为下式：

$$F_U = \frac{S_t[(C_{ss} + C_{bf}) + (1 + C_{bs})e^{r_b(T-t)}] - D(t, T)}{(1 + M_f) - (C_{sf} + M_f)e^{r_b(T-t)}} \quad (4.2)$$

(2) 反向套利策略确定无套利区间的下限

同样的道理，当指数期货合约价格被低估时，利用反向套利策略并考虑交易成本，可以求出无套利区间的下限。

表 4-2: 反向套利的现金流

t 时刻	现金流	T 时刻	现金流
支付融券保证金	$-M_s S_t$	收回融券保证金	$M_s S_t$
融券卖空	S_t	买入股指期货	$-S_T$
卖出股指期货	0	买入现货成本	$-S_T C_{bs}$
		支付股息	$-D(t, T)$
买入期货	0	卖出期货	$F_T - F_t$
买期货成本	$-F_t C_{bf}$	卖期货成本	$-F_T C_{sf}$
支出期货保证金	$-F_t M_f$	收回期货保证金	$F_t M_f$
余额借出	$-[S_t(1 - M_s - C_{ss}) - F_t(M_f + C_{bf})]$	收回借出余额	$[S_t(1 - M_s - C_{ss}) - F_t(M_f + C_{bf})]e^{r_f(T-t)}$
净值	0	投资收益	$M_s S_t - S_T C_{bs} - D(t, T) - F_t - F_T C_{sf} + F_t M_f + [S_t(1 - M_s - C_{ss}) - F_t(M_f + C_{bf})]e^{r_f(T-t)}$

同样根据无套利理论以及预期理论认为到期日期货价格收敛于现货价格，即 $S_T = F_T$ 。T 时刻投资收益理论上应该等于零，而当 T 时刻投资收益大于零时，满足无风险套利，从而由 T 时刻的投资收益等于零可以推导出无套利区间的下限：

$$F_L = \frac{S_t [M_s - C_{bs} - C_{sf} + (1 - M_s - C_{ss})e^{r_f(T-t)}] - D(t, T)}{(1 - M_f) + (C_{bf} + M_f)e^{r_f(T-t)}} \quad (4.3)$$

(3) 套利策略确定无套利区间

确定无套利区间的上下限之后，直接可以得到无套利区间为：

$$[F_L, F_U] \quad (4.4)$$

其中无套利区间的上下限见前述表达式。

而在考虑到现货组合的跟踪误差的情况下，无套利区间的上限应上浮一个单位的跟踪误差；无套利区间的下限下浮一个单位的跟踪误差；由此来确定更贴近实际的无套利区间。令跟踪误差为 TE，则新的无套利区间为：

$$[F_L + TE, F_U - TE] \quad (4.5)$$

当股指期货的市场价格落在这个无套利区间内，则说明市场价格是合理的，不存在套利机会；如果市场价格高于区间上限，说明市场价格被高估，此时投资者可

以采取正向套利策略，从而获取无风险利润，正向套利策略使得市场上对现货的需求增加，多头增加拉动现货价格上涨，同时对期货的空头增加，迫使期货价格下跌，两者同时作用下，期货价格和现货价格趋于收拢，无风险套利空间逐渐消失；反之，如果市场价格低于区间下限，说明市场价格被低估，此时投资者可以采取反向套利策略，从而获取无风险利润，反向套利使得现货价格下跌、期货价格上升，使得期现价格趋于靠拢，无风险套利空间亦会逐渐消失。由此不难发现，套利有利于平稳市场价格。

5 期现套利策略、程序化和套利效果分析

5.1 期现套利策略流程简介

沪深 300 指数期现套利策略可分为五个主要步骤，这些步骤分别是：

一、构建现货组合模拟沪深 300 指数

要进行沪深 300 指数期现套利，那么投资者必须同时拥有现货指数组合和指数期货产品。指数期货在市场上存在并且可以自由买卖，而沪深 300 指数虽然存在，但沪深 300 指数仅仅是用来衡量沪深交易所的总体趋势的全局指标，并非真正的现货组合，因此，投资者必须自己选择适当的方法来构建沪深 300 指数现货组合，之后根据现货组合和沪深 300 指数期货的价差关系进行套利。

二、根据参数估计股指期货合约的无套利区间

期货合约的无套利区间在第四章中已通过表达式阐述出来，现在进行实务操作，则只需要确定无套利区间上下限表达式中的各个参数的具体值即可确定无套利区间。

三、根据期货价格判断是否存在套利空间

确定无套利区间之后，要进行套利以获取利润，必须要综合考虑期货价格和无套利区间的上下限，判断套利机会和套利空间的大小，以方便投资者判断是否进入进行套利交易。

四、确定建仓时机，建立套利头寸

一旦投资者确定进行期现套利，首先应该建立套利头寸，进行建仓，并关注期现价格的价差水平，一旦价格回归到处于无套利区间以内，投资者可以及时的停止建仓，等待价格再次处于无套利区间以外时再进行继续建仓。

五、到期平仓或者策略性提前平仓

套利头寸建立之后，等到价格恢复到正常水平，投资者应该及时的进行平仓实现获利。

5.2 期现套利策略流程的程序化

上一节介绍了本文进行期现套利的主要步骤。即首先，构建现货组合以拟合跟踪沪深 300 指数；其次，确定期现套利的无套利区间的上下限；第三，根据期货的市场价格判断套利空间是否存在以及套利空间大小；第四，如果存在套利空间，则应该及时的建立套利头寸以及时获取瞬息万变的套利利润；第五，根据期货价格和现货价格的走势分析平仓的时机。

然而在实际的市场上，为了及时的把握套利机会，投资者尤其是机构投资者通常会借助于交易程序，通过交易程序自动的完成上述的五个步骤。因此，本文主要就上述五个步骤的具体程序化进行研究，并通过程序软件把各个步骤转换为实际的程序语言，以此来及时获取套利利润。

结合 MATLAB 的编程，本文分别把实际的期现套利策略安排分解如下：第一，选取合适数量的成分股或者 ETF 基金，本文使用分层抽样复制指数法来选取成分股，通过聚类分析算法来实现；第二，跟踪拟合沪深 300 指数，通过遗传算法在最小化跟踪误差的原则下来拟合沪深 300 指数；第三，确定无套利区间上下限，以持有成本模型为基础，再考虑到市场上的各种交易成本和跟踪误差来确定无套利区间的上下限，直接简单编程即可实现；第四，捕捉套利机会，建立套利头寸进行开仓套利，同时按照到期平仓原则，等待到期平仓。为此，按上述的策略进行程序化操作分别编程即可进行期现套利。

（1）选取成分股程序化算法

正如本文前面所述，利用成分股来复制拟合沪深 300 指数有许多的方法，有完全复制指数法，有优化抽样复制法，有分层抽样复制法。考虑到沪深 300 成分股有 300 只之多，完全复制需要买入所有的 300 只成分股，交易成本过大，因此，本文采用分层抽样复制法。即，按照一定的规律把 300 只成分股分为几个或者几十个大类，然后从各个大类中抽取最具有代表性的股票用来拟合沪深 300 指数。

而本文抽取股票参与跟踪拟合沪深 300 指数的程序化算法就是聚类分析算法。聚类分析可以自定义程序把 300 只成分股分为几个大类，然后程序从各个大类中选

取聚类中心参与跟踪拟合，如此就能使得在使用尽可能少的股票的情况下得到尽可能好的拟合效果。

（2）构建现货组合程序化算法

确定好参与跟踪拟合沪深 300 指数的成分股之后，需要进一步确定各只成分股拟合沪深 300 指数的权重（以最小化跟踪误差为原则），进而可以预测未来一段时间的沪深 300 指数走势，通过比较期货指数的走势和现货组合的走势可以预测未来的套利空间的可能大小。

本文使用遗传算法来拟合各只成分股。遗传算法的原理来源于达尔文的进化论，认为种群通过交叉来产生下一代，同时算法对每一代的各个个体进行评价（通过适应性函数），淘汰劣质的后代，保留优质的后代，再重复对优质的后代进行下一次的交叉，同时算法还考虑了交叉产生下一代可能会发生变异的可能性；如此不断重复，最终得到的最优的后代个体即为需要的最优解。遗传算法的具体介绍本文不进行详细的阐述。本文使用遗传算法来把成分股拟合沪深 300 指数，最小化跟踪误差，同时可以计算出各只成分股的权重，给出跟踪误差，以方便下文确定无套利区间的上下限。

（3）确定无套利区间上下限的程序化

前文中通过聚类分析和遗传算法已经得到了各只成分股的权重和相应的跟踪误差；在第六章中本文已经推导出期现套利的无套利区间上下限的表达式。因此，在进行程序化编程时，只需要把相应的表达式转化为程序代码即可得到程序化的无套利区间上下限。

（4）判断套利开仓时机和平仓时机

套利开仓时机的确定比较简单，通常情况下只需要期货市场价格超过无套利区间的上限，就表明存在正向套利机会；反之，期货市场价格低于无套利区间的下限，表明存在反向套利机会。而实际操作中，如果程序表明套利机会仅仅是瞬间存在，按照上述套利方式，程序会进行套利；然而投资者并不希望程序为了微薄的利润而把资金投入其中，因此在判断开仓时机时希望程序向前预测未来的套利机会和套利空间大小，来综合判断是否进行套利。

理论上认为期货价格和现货价格应当一致，因此，如果短期内期现价格存在差异（成为基差），那么长期中期现价格很可能会向着使期现价格一致的方向各自变动，最终期货价格和相应指数的现货价格必然趋于一致或者相近。按照理论所述，投资者完全可以在开仓套利之后不必花多余的时间去关注已投资的期现合约的价格变化，只需等到合约到期自动平仓即可获得无风险套利利润。然后，在实际中，期货价格和现货价格很可能在合约到期前两者相等或者趋于很接近，此时，考虑到利润的时间价值，投资者应该提前平仓了结套利。

考虑到提前平仓在实际操作中并不常见，并且为了更好地理解投资者进行期现套利的策略，在实证部分的期现套利策略中，本文采用简单的到期平仓策略。

5.3 期现套利效果分析

按照前文确定的套利步骤，使用 MATLAB 软件对沪深 300 指数的成分股进行拟合之后确定无套利区间，最后进行套利操作。

（1）成分股跟踪沪深 300 指数的实证效果

<1> 选取成分股拟合沪深 300 指数，计算跟踪误差

正如前文所述，通过聚类分析算法，首先选取 50 只成分股用来拟合跟踪沪深 300 指数的走势。50 只成分股见表 3-1。选取 50 只成分股之后，本文通过遗传算法拟合数据，得到 50 只成分股的权重和跟踪误差。

为了更好的利用全部的数据，本文对于样本的处理方法跟第三章一样，即使用前 242 组数据进行样本内拟合，得到成分股的权重和跟踪误差，仅仅使用最后 10 组数据用来预测未来的期现货价格的走势。得到如下的样本内拟合效果图和样本外预测效果图。

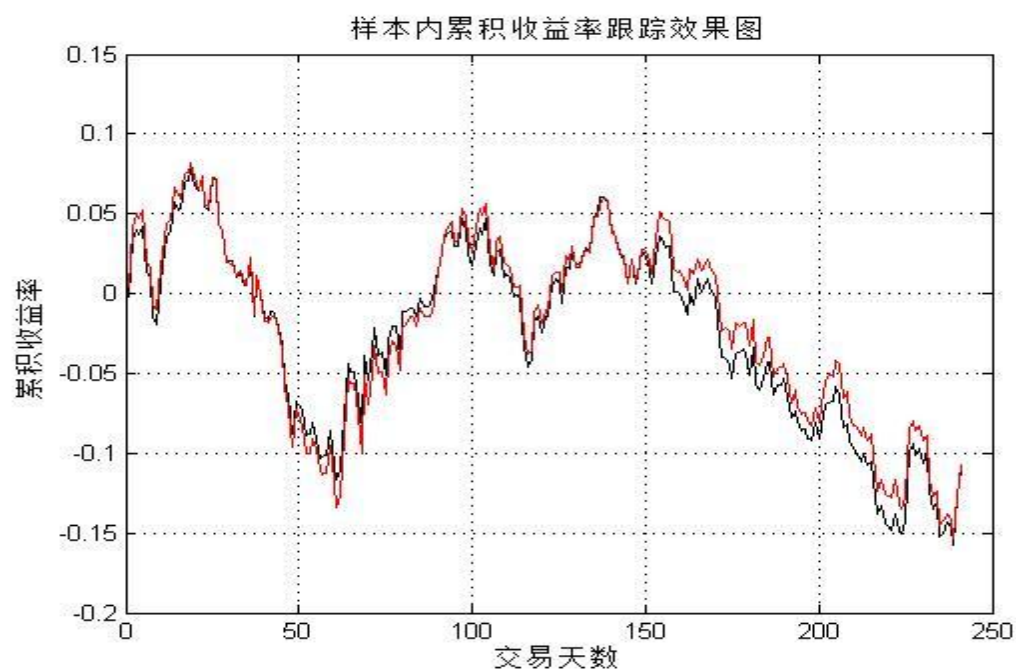


图 5-1：样本内拟合效果图：累积收益率

其中，图 5-1 中，黑色为沪深 300 指数的实际走势（偏下的那条线）；红色为现货组合的走势（偏上的那条线）。

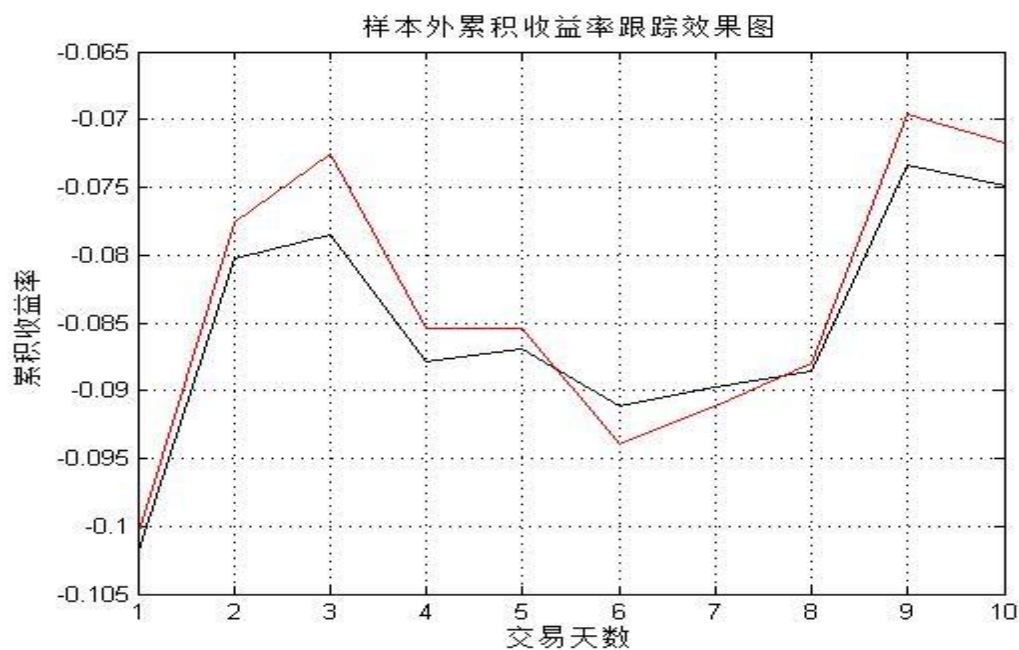


图 5-2：样本外权重拟合效果图：累积收益率

图 5-2 中，黑色为沪深 300 指数（收益率）的实际走势（偏下）；红色为现货组

华中科技大学硕士学位论文

合（收益率）的走势（偏上）。

从上述两个图来看，遗传算法在样本内对沪深 300 指数走势的跟踪拟合效果非常好；同时遗传算法得到的权重对于未来的指数现货的走势的拟合也能起到较好的预测效果。

遗传算法得出的各只成分股的权重如下表 3-2 所示，在此不再重复列示。

从选取的成分股的权重来看，各成分股的权重占有大小跟成分股在沪深 300 指数中占有的权重大小（本文已列出 300 只成分股的名称和代码，见附录；沪深 300 指数各成分股的具体权重可以自行从各种股票软件中查询得到）多数一致（沪深 300 指数中权重大的在跟踪拟合中得到的权重也较大）。由此可见，跟踪拟合的效果较为令人满意。

通过遗传算法得到样本内跟踪误差为 1.4040×10^{-4} ，近似为 $TE=0.014\%$ 。样本外跟踪误差为 7.9434×10^{-4} ，近似为 0.079% 。

<2> 无套利区间上下限

由第四章的期现套利相关参数可知，无套利区间的确定取决于诸多的参数的实际值。为此，首先确定参数的值；各个参数的值见下表：

表 5-1：无套利区间参数值汇总表

参数名称	参数值
C_{ss} （卖出现货的交易成本）	0.118%（交易费用）+0.25%（冲击成本）
C_{bs} （买入现货的交易成本）	0.018%（佣金）+0.25%（冲击成本）
C_{sf} （卖出期货的交易成本）	0.01%（佣金）+0.005%（冲击成本）
C_{bf} （买入期货的交易成本）	0.01%（佣金）+0.005%（冲击成本）
M_f （期货投资保证金比例）	45%
M_s （融券保证金比例）	50% ¹¹

表 5-2：存贷款利率

年利率（%）	2012 年 7 月 6 日	2012 年 6 月 8 日	2011 年 7 月 7 日
存款利率	2.6	3.1	3.1
贷款利率	5.6	5.85	6.1

设定无套利区间的相关参数之后，通过编程运算即可得到无套利区间，以及判断确定套利机会的存在与否。无套利区间上下限及套利机会效果如下图所示：

¹¹证券法规定融券保证金比例不得低于 50%，为了研究方便取最低保证金比例

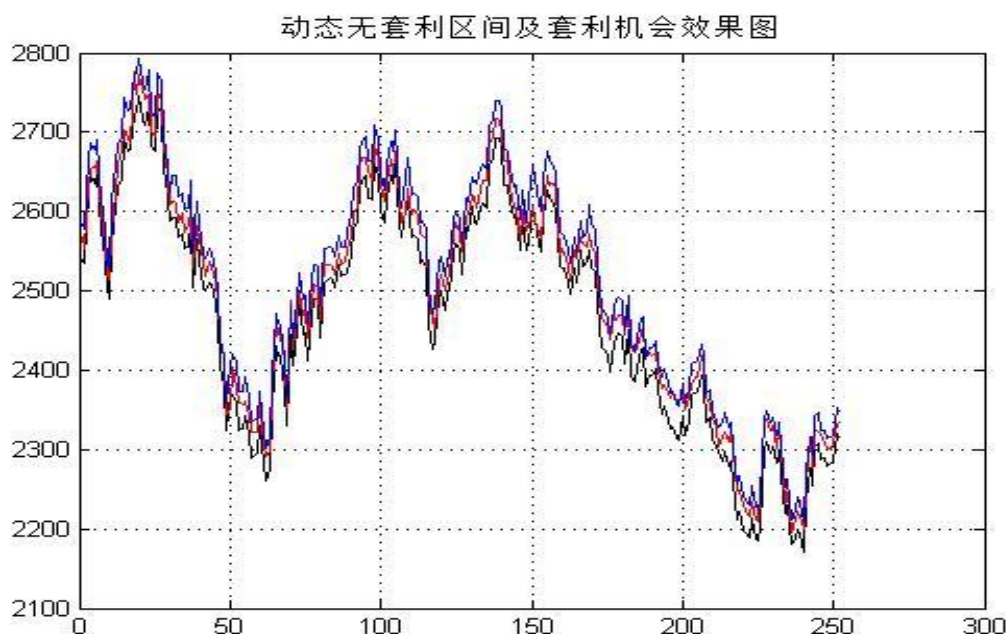


图 5-3：无套利区间及套利机会效果图

其中，黑线代表无套利区间的下边界，蓝线代表无套利区间的上边界，红线表示期货价格的走势。

当红线介于黑线和蓝线之间时，表明市场上期货的价格相对合理，投资者不能通过套利获取无风险利润；当红线高于蓝线时，表明期货价格超过无套利区间的上限，期货价格过高，此时，投资者可以通过进行正向套利获取无风险利润；当红线低于黑线时，表明期货价格过低，此时，投资者可以通过进行反向套利获取无风险利润。

由于上图日期过长，单纯从图上难以看出何时存在套利机会，为了方便套利分析，使用表格方式来展示套利时机，同时计算套利收益率和年化收益率。

以正向套利为例，收益率公式推导如下：

考虑交易成本的现货收益率：

$$R_s = \ln(S_T(1 - C_{ss} - TE)) - \ln(S_t(1 + C_{bs} + TE)) \quad (5.1)$$

考虑交易成本的期货收益率：

$$R_f = \ln(F_t(1 - C_{sf})) - \ln(F_T(1 + C_{bf})) \quad (5.2)$$

从而套利收益率为：

华中科技大学硕士学位论文

$$R_{it} = \ln(S_T(1 - C_{ss} - TE)) - \ln(S_t(1 + C_{bs} + TE)) + \ln(F_t(1 - C_{sf})) - \ln(F_T(1 + C_{bf})) \quad (5.3)$$

由套利平仓时考虑交易成本之后，期货和现货的市场价格趋于相等，从而套利收益率可以简化为：

$$R_{it} = \ln(F_t(1 - C_{sf})) - \ln(S_t(1 + C_{bs} + TE)) \quad (5.4)$$

又因为 $\ln(1 \pm \alpha) \approx \pm \alpha$ ，其中 α 极小。

则 R_{it} 介于 $[\frac{F_t(1 - C_{sf}) - S_t(1 + C_{bs} + TE)}{F_t(1 - C_{sf})}, \frac{F_t(1 - C_{sf}) - S_t(1 + C_{bs} + TE)}{S_t(1 + C_{bs} + TE)}]$ 之间；为了

数据处理的方便，本文取两者之间的中间值： $\frac{F_t(1 - C_{sf}) - S_t(1 + C_{bs} + TE)}{(S_t + F_t)/2}$ 。则套利收

益率公式为：

$$R_{it} = \frac{F_t(1 - C_{sf}) - S_t(1 + C_{bs} + TE)}{(S_t + F_t)/2} \quad (5.5)$$

不难发现，上述公式的分子刚好为每次套利的套利利润，从而收益率公式进一步简化为：

$$R_{it} = \frac{\text{Profit}}{(S_t + F_t)/2} \quad (5.6)$$

得到每次套利的套利收益率计算公式之后，进一步可以计算得出每次套利的年化收益率 $R = 12 * R_{it}$ 。

表 5-3：套利时机和收益率

日期	沪深 300 指数 收盘价	期货当月连续 收盘价	距离到期日	套利利润	年化收益 率	正反向 套利
11/02/2011	2742.39	2777	13	3.33888	1.45%	1
12/14/2011	2397.48	2419	3	2.79243	1.39%	1
12/16/2011	2390.13	2352	1	21.77725	11.02%	-1
01/17/2012	2460.6	2487.4	4	6.58319	3.19%	1
03/16/2012	2623.52	2598.2	1	7.20100	3.31%	-1
05/23/2012	2616.87	2587.8	18	5.77824	2.66%	-1
07/16/2012	2399.73	2417	5	2.10349	1.05%	1
07/30/2012	2335.79	2363.8	15	5.63235	2.88%	1
07/31/2012	2332.92	2363.2	14	8.69189	4.44%	1
09/06/2012	2217.82	2241.2	12	0.07872	0.04%	1
09/07/2012	2317.18	2341.2	11	0.43118	0.22%	1

从得到的套利时机和套利收益率来看,2011 年 10 月到 2012 年 10 月一年中套利机会并不多,每次的套利收益率参差不齐,为了能更好的表示一年中的平均套利收益率,本文取全部套利收益率的平均值来进行分析,简单计算可得平均套利年收益率为 2.88%。由平均收益率来看,虽然 2011 年到 2012 年这一年度中套利机会并不多,但总体的平均套利年化收益率为 2.88%,对于投资者(尤其是机构投资者)来说可谓仍然有利可图。

从各个套利时机来看,不难发现,套利时机近半数出现在距离股指期货到期日较近的时间;这一点很好的应证了股指期货的“到期日效应”,到期日的临近刺激了投机交易者操纵结算价格的欲望,使得到期日临近时,股指期货的成交量和波动率显著提高,从而导致大量的套利机会的出现。

在年度末,套利机会也较多,并且套利空间相对于其他时点特别显著;此现象的出现可能的原因在于年度末是企业一整个年度业务的整合时点,企业的各项业务在年末进行综合处理,投机交易者在此可以利用企业与投资者的信息的不对称来操纵结算价格,使得股指期货的价格波动率显著增加,导致产生套利机会。

除了到期日效应和年末效应外,研究还发现,正向套利机会(表 5-3 中最后一列,1 表示正向套利机会,-1 表示反向套利机会)远多于反向套利机会,并且正向套利时机多数出现在股价走势趋于下降的阶段,可能原因在于股价下降阶段为股票市场熊市,投资者对于股票的投资信心下降,转向投资与其他金融工具,如投资期货,导致期货价格和现货价格之间的关联不如牛市那么紧密,从而容易产生套利机会。

(2) ETF 跟踪沪深 300 指数的实证效果

前面使用 50 只成分股跟踪拟合沪深 300 指数,并以此构建的现货组合来跟股指期货进行期现套利。由于成分股交易成本和其他各个影响因素较大,使得最终的套利机会和套利利润并不十分理想,为此,特在本文最后使用深证 100ETF、上证 50ETF、上证 180ETF 和红利 ETF 四个 ETF 来跟踪沪深 300 指数。之所以使用这四个 ETF 来跟踪拟合沪深 300 指数,原因在于这四个 ETF 跟沪深 300 指数的相关度非常之高,使用它们来跟踪沪深 300 指数能达到十分理想的效果。

为了跟纯成分股的跟踪效果比较,本文选取自 2011 年 10 月 10 日到 2012 年 10

月 19 日的股指期货和四只 ETF 的全部交易数据, 共计 252 组数据。如在第三章的 3.2.2 的实证检验分析部分所提及的, 选择这四只 ETF 来跟踪模拟沪深 300 指数, 因为它们跟沪深 300 指数的相关度非常高, 为了更好的说明它们之间的相关度大小, 在此首先对 ETF 和沪深 300 指数的相关度大小进行分析。

下表给出了这四个 ETF 跟沪深 300 指数的相关系数。

表 5-4: ETF 和沪深 300 指数价格相关系数

价格相关系数	沪深 300	深 100ETF	上证 50ETF	上证 180ETF	红利 ETF
沪深 300 指数	1.0000	0.9852	0.9199	0.9893	0.9792
深证 100ETF	0.9852	1.0000	0.8751	0.9691	0.9365
上证 50ETF	0.9199	0.8751	1.0000	0.9615	0.9240
上证 180ETF	0.9893	0.9691	0.9615	1.0000	0.9685
红利 ETF	0.9792	0.9365	0.9240	0.9685	1.0000

表 5-5: ETF 和沪深 300 指数的收益率相关系数

收益率相关系数	沪深 300	深 100ETF	上证 50ETF	上证 180ETF	红利 ETF
沪深 300 指数	1.0000	0.9751	0.9648	0.9850	0.9559
深证 100ETF	0.9751	1.0000	0.9119	0.9496	0.9147
上证 50ETF	0.9648	0.9119	1.0000	0.9753	0.9571
上证 180ETF	0.9850	0.9496	0.9753	1.0000	0.9559
红利 ETF	0.9559	0.9147	0.9571	0.9559	1.0000

类似于成分股跟踪拟合沪深 300 指数的数据处理方法, 留取最后 10 个交易日的数据最为预测分析之用, 前面的数据用来跟踪拟合, 所以前面的 242 组数据使用来跟踪拟合沪深 300 指数, 最后 10 组数据以供预测效果分析之用。跟踪效果和预测效果图示如下:

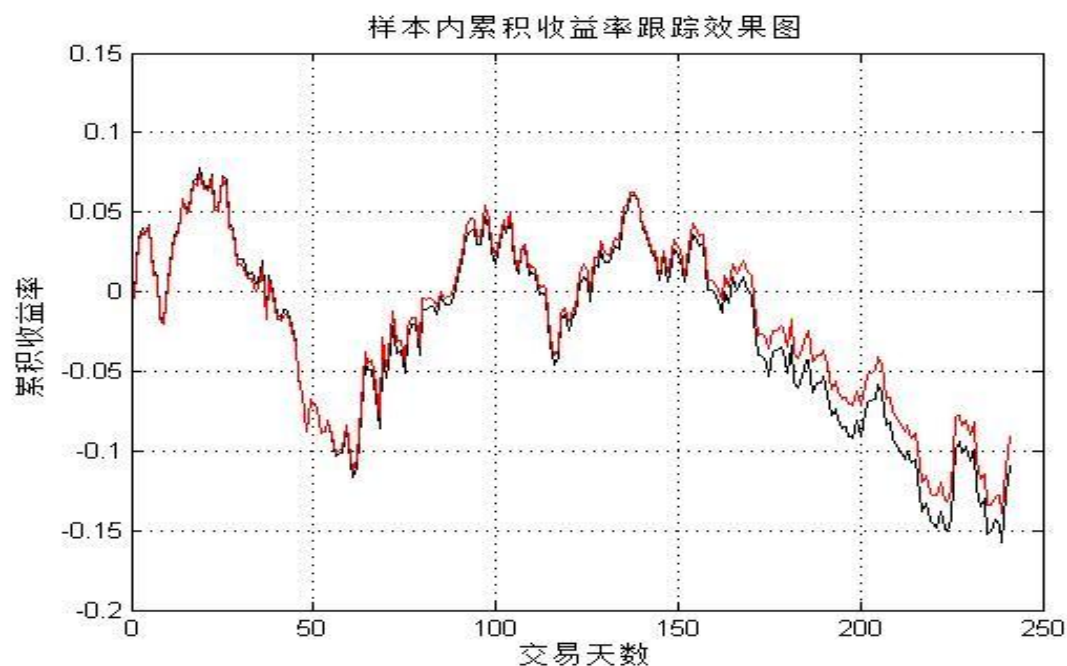


图 5-4: ETF 跟踪效果: 累积收益率 (沪深 300 指数真实走势线偏下)

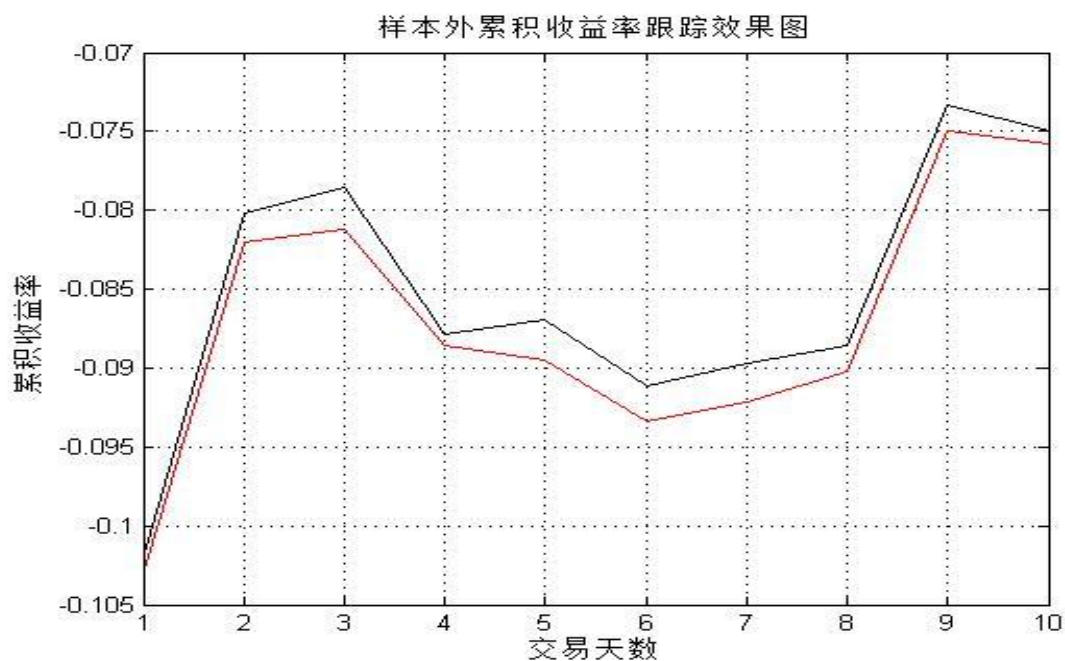


图 5-5: 样本外权重预测效果 (沪深 300 指数真实走势线偏上)

上述四个图中, 图 5-4 中的黑线代表沪深 300 指数的实际 (收益率) 走势, 红线代表跟踪拟合出来的现货组合 (收益率) 的走势。图 5-5 中的黑线代表样本外沪深 300 指数收益率的实际走势, 绿线代表仿真预测的现货组合收益率的走势。

华中科技大学硕士学位论文

使用 ETF 跟踪沪深 300 指数的样本内跟踪误差为 8.9381×10^{-5} ，近似为 0.0089%。样本外的跟踪误差为 3.2463×10^{-4} ，近似为 0.0325%。相比于使用成分股跟踪拟合沪深 300 指数，跟踪误差更小，说明使用相关度高的 ETF 跟踪拟合沪深 300 指数的效果非常好。

同样，通过遗传算法可以直接得到四只 ETF 的权重如表 3-5 所示，在此不再重复列示。

接下来便可以确定无套利区间的上下限。然而跟成分股跟踪沪深 300 指数来套利不同，ETF 不存在分红的考虑，所以无套利区间的上下限需要重新调整。

无套利区间上限为：

$$F_U = \frac{S_t [(C_{ss} + C_{bf}) + (1 + C_{bs})e^{r_b(T-t)}]}{(1 + M_f) - (C_{sf} + M_f)e^{r_b(T-t)}} \quad (5.7)$$

无套利区间下限为：

$$F_L = \frac{S_t [M_s - C_{bs} - C_{sf} + (1 - M_s - C_{ss})e^{r_f(T-t)}]}{(1 - M_f) + (C_{bf} + M_f)e^{r_f(T-t)}} \quad (5.8)$$

其中，各个参数的含义仍然不变，如第四章中的参数描述。

同样无套利区间的确定还需要考虑跟踪误差，从而无套利区间为：

$$[F_L + TE, F_U - TE] \quad (5.9)$$

由第三章的 3.2.2 的实证检验分析的 ETF+遗传算法；得到的样本内跟踪误差为 8.9381×10^{-5} ，近似为 0.0089%，样本外跟踪误差为 3.2463×10^{-4} ，近似为 0.0325%。无套利区间的上下限的相关参数的实际取值却不同于成分股跟踪模拟沪深 300 指数的参数值。存贷款利率如表 5-2 所示，其他参数的实际取值如下表：

表 5-6: ETF 跟踪指数无套利区间的参数值汇总表¹²

参数名称	参数值
C_{ss} (卖出现货的交易成本)	0.05% (佣金) +0.1249% (冲击成本)
C_{bs} (买入现货的交易成本)	0.05% (佣金) +0.1249% (冲击成本)
C_{sf} (卖出期货的交易成本)	0.01% (佣金) +0.005% (冲击成本)
C_{bf} (买入期货的交易成本)	0.01% (佣金) +0.005% (冲击成本)
M_f (期货投资保证金比例)	45% ¹³
M_s (融券保证金比例)	50% ¹⁴

无套利区间及套利时机效果图如下:

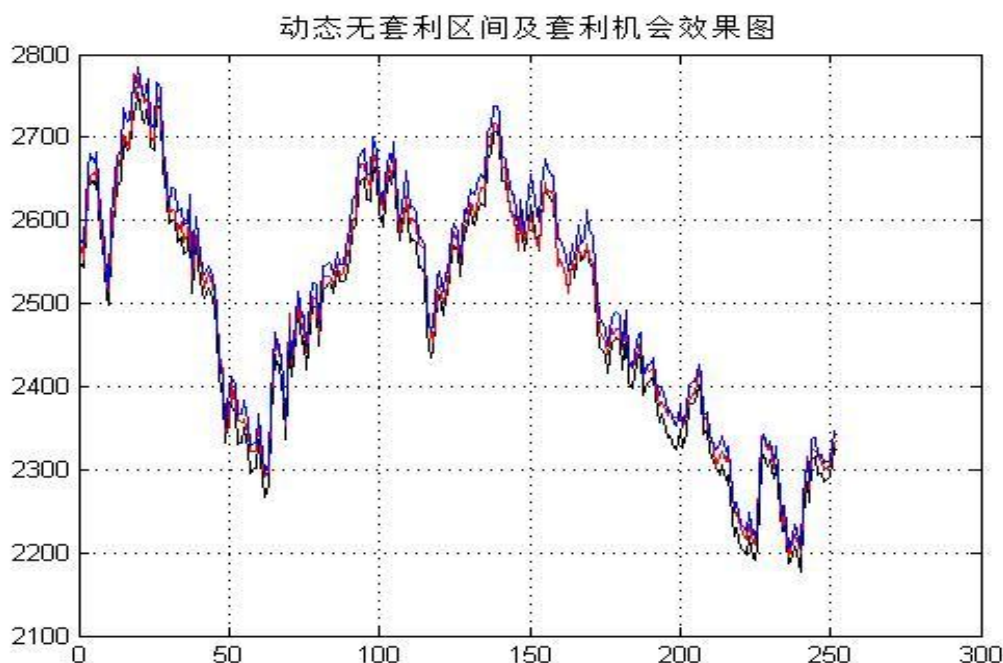


图 5-6: 无套利区间及套利时机

其中,黑线代表无套利区间的下限,蓝线代表无套利区间的上限,红线代表股指期货的实际走势。

而实际的存在套利机会的时机可列表如下:

¹²参数值直接选择取自:申璐娟《沪深 300 股指期货套利策略研究》,山西财经大学,2012.10,硕士学位论文

¹³通常期货保证金需要确保 15%,同时初始保证金水平最好为保证金水平的 3 倍。

¹⁴选择法规规定的最低的融券保证金比例。

华中科技大学硕士学位论文

表 5-7：套利时机和收益率

日期	沪深 300 指数	期货当月连续	距离到期日	套利利润	年化收益率	正反向套利
11/01/2011	2697.53	2686.6	14	0.684498	0.31%	-1
11/02/2011	2742.39	2777	13	11.32064	4.92%	1
12/13/2011	2421.93	2438.8	4	4.033883	1.99%	1
12/14/2011	2397.48	2419	3	9.661916	4.81%	1
12/16/2011	2390.13	2352	1	28.8609	14.61%	-1
01/13/2012	2394.33	2410.2	6	1.484316	0.74%	1
01/17/2012	2460.6	2487.4	4	13.75893	6.67%	1
03/16/2012	2623.52	2598.2	1	15.14579	6.96%	-1
05/14/2012	2615.53	2603.8	5	1.651133	0.76%	-1
05/15/2012	2617.37	2604.2	4	3.067943	1.41%	-1
05/16/2012	2574.65	2563.2	3	1.496989	0.70%	-1
05/17/2012	2613.94	2603.2	2	0.619024	0.28%	-1
05/22/2012	2627.53	2607.6	19	10.03119	4.60%	-1
05/23/2012	2616.87	2587.8	18	19.19525	8.85%	-1
05/24/2012	2595.26	2582.2	17	3.250829	1.51%	-1
05/28/2012	2614.69	2598.8	15	5.975219	2.75%	-1
06/04/2012	2559.03	2545.4	10	3.847564	1.81%	-1
06/08/2012	2524.33	2513	6	1.618098	0.77%	-1
06/13/2012	2580.64	2568	3	2.663833	1.24%	-1
06/15/2012	2568.05	2556	1	2.09091	0.98%	-1
07/16/2012	2399.73	2417	5	4.050163	2.02%	1
07/27/2012	2349.11	2372.2	16	1.730059	0.88%	1
07/30/2012	2335.79	2363.8	15	7.533813	3.85%	1
07/31/2012	2332.92	2363.2	14	10.59035	5.41%	1
08/13/2012	2351.93	2367.8	5	2.913488	1.48%	1
08/17/2012	2313.48	2303.4	1	1.105856	0.57%	-1
09/06/2012	2217.82	2241.2	12	6.108487	3.29%	1
09/07/2012	2317.18	2341.2	11	6.729988	3.47%	1
09/28/2012	2293.11	2315.6	11	5.37959	2.80%	1
10/11/2012	2302.53	2320	7	3.288124	1.71%	1

其中，套利收益率及年化收益率的计算公式跟成分股跟踪沪深 300 指数套利收益率及年化收益率的计算公式相同。

前文中使用成分股跟踪沪深 300 指数进行期现套利在一整个年度中，得到的套利时机非常少，而使用 ETF 跟踪拟合沪深 300 指数进行期现套利得到的同样一年中的套利时机相对而言较多，套利时机亦有 30 个之多，由此可见使用 ETF 跟踪沪深 300 指数进行套利相对于使用成分股跟踪沪深 300 指数套利能把握更多的套利机会。而 ETF 套利的平均套利年化收益率为 3.07%。ETF 跟踪沪深 300 指数进行期现套利的收

益率为 3.07%对于投资者而言是有利可图的。

从套利时机点来看,套利时机多出现在期货合约到期日临近时,进一步说明股指期货的“到期日效应”的存在对期现套利的影响。并且在使用 ETF 跟踪模拟沪深 300 指数进行期现套利时,发现的套利时机点相比于使用成分股跟踪模拟沪深 300 指数得到的套利时机点更多的出现在临近与期货合约到期日的时点,由此更进一步证实了期货合约到期日效应的存在。

同时,类似于使用成分股跟踪沪深 300 指数,套利机会在年度末出现较多,同时套利空间相对于其他时点来说都是十分显著的。可称为“年末效应”。

对应了图 5-6,不难发现,股价走势趋于上升阶段时,反向套利机会较多,而当股价走势趋于下降阶段时,正向套利机会占据绝大多数。可能原因在于,股市牛市时,投资者大量投资股票,股价泡沫较大,股价指数相应较高,导致反向套利机会较多;反之,股票熊市时,投资者对股票信心不足,转而投资于其他金融工具,其中较多投资于期货,导致期货价格被高估的机会增加,正向套利机会较多。这一结论与使用成分股跟踪模拟沪深 300 指数进行套利得到的结论一致。

6 结论和建议

6.1 结论

沪深 300 指数期货推出以来,我国基于股指期货进行套利越来越受到广大的投资者的关注。使用成分股跟踪沪深 300 指数时,本文采用基于 2011 年 10 月到 2012 年 10 月共计一整年度的沪深 300 指数和股指期货的日度交易数据,分析了股指期货的定价模型和股指期货套利的现货组合的构建模型,再者确定无套利区间的参数,最后通过程序化的方法把股指期货套利策略展现出来。使用 ETF 跟踪沪深 300 指数时,为了更好的进行比较本文采用基于 2011 年 10 月到 2012 年 10 月的四只 ETF 和沪深 300 指数、股指期货的交易数据分析了使用 ETF 跟踪沪深 300 指数再进行期现套利的套利机会及分析了最终的套利效果。通过实证和编程得到如下的结论:

第一,在确定股指期货的定价模型方面,常用且成熟的模型有持有成本模型和与其理论模型;本文在综合考虑了模型的适应性、可操作性等方面的因素后,选择使用持有成本模型。以持有成本模型为原型,考虑实际市场上可能存在的影响股指期货定价的因素,为股指期货的定价确定了初步的模型,这为下文的无套利区间的确定打好了基础。

第二,构建期现套利的现货组合通常有完全复制指数法和部分复制指数法。本文拟使用完全的成分股来跟踪拟合沪深 300 指数,并在其后又使用 ETF 来复制沪深 300 指数,主要目的在于对比两种策略的优劣。这其中主要考虑到,ETF 跟沪深 300 指数有较好的相关性,在跟踪拟合沪深 300 指数方面容易出现权重不稳定的可能性;并且 ETF 类型远没有股票丰富,存在流动性不足的特点,从而即使理论上确定的 ETF 组合对沪深 300 指数的跟踪拟合效果很好,也很难保证实际操作中由于流动性不足会导致产生难以预料的成本和风险。即使完全的成分股跟踪你和沪深 300 指数交易成本比 ETF 复制要大,但是流动性风险就能很好的避免。通过实证分析,本文选取了 50 只成分股来跟踪拟合沪深 300 指数,实证得到的跟踪误差仅为 0.01404%,使用 ETF 跟踪拟合沪深 300 指数得到的跟踪误差为 0.00894%,较小的跟踪误差很好的保

证了利用成分股拟合沪深 300 指数的效果，同时为无套利区间的确定提供了必要的参数值。

第三，确定无套利区间之后，利用 MATLAB 软件运行发现套利机会，由于本文考虑了尽可能多的交易成本，无套利区间几乎覆盖了期货市场价格的全部可能走势，得到的套利机会非常少，尽管如此，得到的几个套利机会也能给投资者提供丰厚的收益率。因为本文使用的交易数据覆盖了一个完整的交易年度，从而最终的平均套利收益率可以用来表示 2011 年 10 月到 2012 年 10 月这一周年内的套利机会大小的平均水平。同时成分股跟踪套利的 2.88% 的平均年化收益率以及 ETF 跟踪套利的 3.07% 的平均年化收益率同比也其他研究者的结果，处于合理的水平，说明本文的套利策略能在实际的操作中发挥一定的作用。此外，通过分析套利时机发现股指期货的“到期日效应”对期现套利有显著的影响；并且年度末，期现套利的套利利润（或称套利空间）相对于其他时点的套利利润特别的显著。

第四，套利与套期保值、投机并成为金融产品的三大投资工具。我国沪深 300 指数期货推出之后，市场上各大投资机构纷纷开发自己的包括套利产品在内的各种金融产品。然而规模较小的投资机构和普通投资者，开发套利产品对于他们而言成本太高，过渡期太长，没有动力去研究开发自己的套利产品，因此失去了抢占金融市场的大好工具，本文对于股指期货期现套利策略的研究重点在于具体的程序化算法，本文的研究在一定程度上是通过对套利的实际操作策略的研究向广大投资者展示期现套利的整个过程，投资者对于期现套利策略或者产品的开发可以较好的基于本文所介绍的几种算法进行。

6.2 不足之处及建议

在实际的套利实施的操作过程中，市场上出现的任何不利的因素或者事件都可能侵蚀掉套利利润，投资者及研究者都不可能规避所有的不利因素，因此，投资者和研究人员只有不断优化套利实施步骤，尽可能全面的考虑所有的因素和事件的发生，才能保证套利收益的最大化。因此，在具体套利操作方面，本文提出几点建议：

第一，在选取成分股跟踪拟合沪深 300 指数时，本文不加考虑数量的直接使用

50 只成分股来跟踪指数，虽然本文的跟踪拟合效果较好，但是仍然存在隐藏的缺点，因此投资者如果采用纯成分股来跟踪指数，可以考虑使用不同数目的成分股来跟踪指数，选取跟踪误差最小的数目的成分股跟踪指数才能更好的进行后续的套利策略。

第二，在构建现货组合方面，完全复制法和部分复制法（本文选择的是分层抽样法）各有优缺点，选择哪种方法取决于投资者对收益和成本的权衡。同时，在选择使用成分股跟踪指数还是选择使用 ETF 跟踪指数方面，不同的选择有其优点和缺点，两个方法都能较好的构建现货组合，当然如果投资者希望能兼得两者的优点，也不妨考虑综合使用成分股和 ETF 来跟踪指数。

第三，本文没有考虑投资者的资金规模和资金风险管理水平，因此，本文的套利策略适用的投资者有限。投资者应当综合考虑自己获取资金的难易程度以及自己对风险的偏好程度来选择使用不同的套利策略，以免承担不必要的风险。

第四，本文研究套利开仓时机和平仓时机时提到，对于开仓时机的把握在于套利空间的大小。而对于平仓时机的确定，本文直接选择到期平仓策略，而市场期现价格瞬息万变，如果期现价格在到期前已经相近或者相等，投资者应当当机立断，提前平仓相对于到期平仓来言更好。

第五，本文在跟踪拟合现货组合时，考虑到了使用当前的数据预测未来的现货指数的走势，但限于数据的有限以及限于作者软件使用能力，并没有完全使用预期未来现货指数走势来发现未来的套利机会。因此，以后的研究者在研究套利策略方面，如果对于软件的使用和对当前数据的利用能力所及，希望能够尽可能考虑充分利用当期的交易数据预测现货指数的走势，从而发现未来可能的套利，以便投资者提前进行套利获取收益。

致谢

光阴如梭，不知不觉美好的两年的研究生生活即将画上句号，紧接着研究生生活的结束，我也将要离开美丽的华中科技大学踏上人生崭新的旅途。论文也将要定稿，借此时机向所有支持我、帮助过我的老师们、同学们和朋友们表示衷心的感谢。

在此，我首先要感谢我的研究生阶段的指导导师刘方池老师。感谢刘老师在我研究生阶段对我学习上的悉心指导和生活上的无微不至的关怀，令我在专业知识和生活知识上都得到了很大的进步；让我感觉到亲人般的温暖。特别是在我选和论文写作过程中，老师的关注都是我一往无前的无尽动力源泉。

其次，我要感谢我的室友和其他的同学，在你们的陪伴下，我严格按照坚持论文写作安排完成毕业论文，这使得我论文思路清晰、思考细微，同时感谢他们在资料收集方面对我的帮助。

最后，要感谢我的家人，是你们的支持、包容和肯定让我更加安心在校学习，顺利完成学业。

在此，相关所有关心、支持和帮助过我的老师们、同学们、朋友们表示最诚挚的谢意。谢谢你们！

参考文献

- [1] Brener,Subrahanyam,Uno.Arbitrage. Opportunities in the Japanese stock and future smarkets[J].Finacial Analysts,1990(46):14-24
- [2] Brennan M , E.Schwartz. Arbitrage in stock index futures[J].The Journal of Business,1990 (63):7-31
- [3] Cornell B and K.R French.The pricing of stock index futures[J].Journal of FuturesMarkets,1983(3):1-14
- [4] Francesco Corielli, Massimiliano Marcellino. Factor Based Index Tracking [R].Working Paper,2002(2):6-41.
- [5] Hans R.Stoll, Robert E.Whaley.The Dynamics of Stock Index and Stock Index FuturesReturns[J].TheJournal of Financial and Quantitative Analysis:1990
- [6] Harris.The October1987 S&P500 stock-futures basis[J].The Journal of Finance ,1989(44): 77-99
- [7] Hsu Hsinan ,Janchung Wang. The Pricing Model of Stock Index Futures in Imperfect Markets and Analysis of Price Expectation[J].Journal of NationalCheng Kung University,1998(33):355-381.
- [8] J.E.Beasley,N.Meade.An Evolutionary Heuristic for the Index Tracking Problem[J].Working Paper, 2001(10):2-34.
- [9] Klemksky,R.C. and J.H.Lee.The intraday ex post and ex ante profit ability of index arbitrage[J].Journal of Futures Markets,1991(11):291-311.
- [10] Markus R, Hans-Jurgen W, Heninz Z. A Linear model for tracking errorminimization[J]. Journal of Banking& Finance,1999(23):85-103.
- [11] MacKinlay,C,and K.Ramaswamy.Index-futures arbitrage and the behavior of stockindex futures prices[J].Review of Financ Studies,1988(1):137-158.
- [12] Marseilles. A Hybrid Genetic Algorithm for Passive Management [C]. Second Conference Computing in Economics and Finance Society of Computational Economics,1996(6): 1-21.
- [13] Meade N,Salkin G R.Developing and maintaining an equity index fund[J].Journal of the Operational Research Society,1990(41):599-607.

华中科技大学硕士学位论文

-
- [14] Modest,D.M.On the Pricing of Stock Index Futures[J].Journal of PortfolioMangement,1984:51-57
- [15] Neal,R.Direct tests of index arbitrage models[J].Journal of Financial andQuantitative Analysis,1996(31):541-562.
- [16] Roll R. A mean/variance analysis of tracking error[J]. The Journal of Portfolio Managemen,1992,(18):13-22.
- [17] Y.PeterChung. A transactions data test of stock index futures market efficiencyand index arbitrage profitability[J].The Journal of Finance1991(146):1791-1809.
- [18] 陈守东,刘英华.最优化模型下的股指期货套利[J].税务与经济, 2009
- [19] 陈守东,韩一哲,荆伟.主要股票市场指数与我国股票市场指数间的协整分析.数量经济技术经济研究.2003.05
- [20] 陈晓静,陈钟.从仿真交易看沪深 300 股指期货的期限套利.证券市场导报.2007,(10):12-17.
- [21] 崔建军.用上证 50ETF 与沪深 300 期指进行套利可行[N].期货日报,2007.
- [22] 党剑,方世圣.股指期货实战解析:沪深指数期货投资实务.上海财经大学出版社.2007.5,初版,第 202 页
- [23] 丁玉洁,周圣武. 股指期货套利与现货组合模型的构建 [J]. 现代经济,2008(7):16-18.
- [24] 方世圣.股指期货套保与套利.上海远东出版社.2010.01
- [25] 方卫东,张晓峰.股指期货套利中跟踪指数投资组合的构建[J].科学技术与工程,2008 (17) : 3-5.
- [26] 高辉,赵进文,2007,沪深 300 股指套期保值及投资组合实证研究[J].管理科学,4:82-92.
- [27] 郭德明.股指期货套利应用与实证研究.安徽大学硕士学位论文.2010.05
- [28] 高能斌.关于股指期货定价及其套利区间的研究.当代经济 2008,(01):44-46.
- [29] 何晓彬.股指期货套期保值策略理论与应用研究.厦门大学博士学位论文.2008.06
- [30] 雷泽,秦昌贵.股指期货:套利空间大利润高[J].资本市场,2007(1):68-69.
- [31] 李传峰.沪深 300 股指期货期限套利模型及实证分析.广东金融学院学
-

华中科技大学硕士学位论文

- 报.2011.01,Vol26,No.1
- [32] 梁斌.股指期货套期保值和套利策略研究.中国科学技术大学博士学位论文.2010.04
- [33] 林文正,臧大年.台湾股指期货定价与套利实务问题探讨[J].证券市场发展季刊,1996(3):1-31.
- [34] 刘菩,王彩虹.中国金融期货交易所结算业务及细则介绍 6.中国金融期货交易所保险机构培训专家组编委,2007 年 7 月,初版,第 1 页
- [35] 刘冰.沪深 300 股指期货期限套利中现货选择研究.时代经贸.2007,87(12):77.
- [36] 罗利.沪深 300 指数期货套期保值及套利应用研究.复旦大学硕士学位论文.2010.05
- [37] 陆珩琪,陈伟忠.指数期货套利交易策略设计.同济大学现代金融研究所.2004.04
- [38] 申璐娟.沪深 300 股指期货套利策略研究.山西财政大学.硕士学位论文.2012.02
- [39] 史瑛.股指期货套利与指数复制[J].信阳师范学院学报,2009(5):62-65.
- [40] 苏蝉媛.股指期货定价:沪深 300 股指期货期限套利分析.证券与保险.2007,(19):48-51.
- [41] 谭海东.指数基金如何利用沪深 300 股指期货进行套期保值.西南财经大学硕士学位论文.2011.04
- [42] 陶估.沪深 300 股指期货理论和实务[M].中国发展出版社,2008:221.
- [43] 王红兵.金融工程:股指期货全攻略,套利、套保与产品设计.联合证券.2010.01
- [44] 王鹏飞.论股票指数期货的定价原理与模型.金融观察,2006(3):46-52.
- [45] 王伟峰,刘阳.股指期货的跨期套利研究:模拟股指市场实证.金融研究.2007.12
- [46] 吴育华,徐忠东.股指期货定价方法研究.天津财经学院,2006(6):12-15.
- [47] 肖禄梅.沪深 300 指数及股指期货的研究[D].对外经济贸易大学博士学位论文.2007
- [48] 辛何庚.保底性金融产品的设计与定价.湖南大学硕士学位论文.2009.04
- [49] 徐国祥,檀向球.指数期货合约定价模型及其实证研究[J].统计研究,2003(8):50-53.(5):23-28.
- [50] 曾忠东,彭菊.ETF 在股指期货期现套利中的跟踪误差风险分析[J].西南民族大

华中科技大学硕士学位论文

- 学学报, 2010 (6) :197-200.
- [51] 张冠华.沪深 300 股指期货套利策略及风险管理研究.中国海洋大学.硕士学位论文.2009.06
- [52] 张敏.股指期货套利与套期保值研究.华中科技大学博士学位论文.2007.11
- [53] 张敏,黄冠钥.股指期货合约间套利研究.当代经济,2006(12):85-86
- [54] 张敏,徐坚.ETF 在股指期货期现套利的现货组合中的应用.技术经济与管理研究,2007(3):31-32.
- [55] 张学东.股价指数期货理论与实践研究.[M]中国社会科学出版社.2005(6)
- [56] 张宗成.卖空限制下股指期货投资组合最优模型[J].投资与证券,2002,(8)
- [57] 赵曙东.期货投资和期权.南京大学出版社,2001.

附录：沪深 300 成分股

本文选取使用的 300 只成分股为 2012 年 10 月时沪深 300 指数的 300 成分股，其中 300 只成分股的代码和名称如下表：

股票代码	股票名称	股票代码	股票名称	股票代码	股票名称
SH600000	浦发银行	SH600066	宇通客车	SH600169	太原重工
SH600005	武钢股份	SH600068	葛洲坝	SH600170	上海建工
SH600009	上海机场	SH600085	同仁堂	SH600177	雅戈尔
SH600010	宝钢股份	SH600089	特变电工	SH600183	生益科技
SH600011	华能国际	SH600096	云天化	SH600188	兖州煤业
SH600015	华夏银行	SH600100	同方股份	SH600196	复星医药
SH600016	民生银行	SH600104	上汽集团	SH600208	新潮中宝
SH600019	宝钢股份	SH600108	亚盛集团	SH600216	浙江医药
SH600022	山东钢铁	SH600109	国金证券	SH600219	南山铝业
SH600028	中国石化	SH600111	包钢稀土	SH600221	海南航空
SH600029	南方航空	SH600115	东方航空	SH600252	中恒集团
SH600030	中信证券	SH600118	中国卫星	SH600256	广汇能源
SH600031	三一重工	SH600123	兰花科创	SH600259	广晟有色
SH600036	招商银行	SH600125	铁龙物流	SH600266	北京城建
SH600037	歌华有线	SH600132	重庆啤酒	SH600267	海正药业
SH600048	保利地产	SH600143	金发科技	SH600271	航天信息
SH600050	中国联通	SH600150	中国船舶	SH600276	恒瑞医药
SH600058	五矿发展	SH600153	建发股份	SH600307	酒钢宏兴
SH600062	华润双鹤	SH600160	巨化股份	SH600309	烟台万华
SH600664	哈药股份	SH600166	福田汽车	SH600315	上海家化
SH600674	川投能源	SH601101	昊华能源	SH600508	上海能源
SH600690	青岛海尔	SH601106	中国一重	SH600516	方大炭素
SH600694	大商股份	SH601111	中国国航	SH600518	康美药业
SH600703	三安光电	SH601117	中国化学	SH600519	贵州茅台
SH600718	东软集团	SH601118	海南橡胶	SH600528	中铁二局
SH600739	辽宁成大	SH601158	重庆水务	SH600535	天士力
SH600741	华域汽车	SH601166	兴业银行	SH600546	山煤国际
SH600770	综艺股份	SH601168	西部矿业	SH600547	山东黄金
SH600779	水井坊	SH601169	北京银行	SH600549	厦门钨业
SH600783	鲁信创投	SH601179	中国西电	SH600550	天威保变
SH600795	国电电力	SH601186	中国铁建	SH600583	海油工程
SH600804	鹏博士	SH601216	内蒙君正	SH600585	海螺水泥
SH600809	山西汾酒	SH601233	桐昆股份	SH600588	用友软件
SH600811	东方集团	SH601258	庞大集团	SH600595	中孚实业
SH600812	华北制药	SH601268	二重重装	SH600598	北大荒
SH600827	友谊股份	SH601288	农业银行	SH600600	青岛啤酒
SH600832	东方明珠	SH601299	中国北车	SH600642	申能股份
SH600837	海通证券	SH601318	中国平安	SH600649	城投控股

华 中 科 技 大 学 硕 士 学 位 论 文

SH600839 四川长虹	SH601328 交通银行	SH600655 豫园商城
SH600859 王府井	SH601333 广深铁路	SH600660 福耀玻璃
SH600863 内蒙华电	SH601336 新华保险	SH600316 洪都航空
SH600873 梅花集团	SH601369 陕鼓动力	SH600320 振华重工
SH600875 东方电气	SH601377 兴业证券	SH600331 宏达股份
SH600881 亚泰集团	SH601390 中国中铁	SH600348 阳泉煤业
SH600887 伊利股份	SH601398 工商银行	SH600352 浙江龙盛
SH600893 航空动力	SH601555 东吴证券	SH600362 江西铜业
SH600895 张江高科	SH601558 华锐风电	SH600369 西南证券
SH600900 长江电力	SH601566 九牧王	SH600372 中航电子
SH600970 中材国际	SH601600 中国铝业	SH600376 首开股份
SH600971 恒源煤电	SH601601 中国太保	SH600383 金地集团
SH600997 开滦股份	SH601607 上海医药	SH600395 盘江股份
SH600999 招商证券	SH601618 中国中冶	SH600406 国电南瑞
SH601001 大同煤业	SH601628 中国人寿	SH600415 小商品城
SH601006 大秦铁路	SH601633 长城汽车	SH600418 江淮汽车
SH601009 南京银行	SH601666 平煤股份	SH600432 吉恩镍业
SH601018 宁波港	SH601668 中国建筑	SH600456 宝钛股份
SH601088 中国神华	SH601669 中国水电	SH600489 中金黄金
SH601098 中南传媒	SH601688 华泰证券	SH600497 驰宏锌锗
SH601099 太平洋	SH601699 潞安环能	SH600498 烽火通信
SH601718 际华集团	SH601717 郑煤机	SH600500 中化国际
SH601766 中国南车	SZ000422 湖北宜化	SZ000933 神火股份
SH601788 光大证券	SZ000423 东阿阿胶	SZ000937 冀中能源
SH601808 中海油服	SZ000425 徐工机械	SZ000960 锡业股份
SH601818 光大银行	SZ000527 美的电器	SZ000961 中南建设
SH601857 中国石油	SZ000528 柳 工	SZ000968 煤 气 化
SH601866 中海集运	SZ000538 云南白药	SZ000969 安泰科技
SH601888 中国国旅	SZ000559 万向钱潮	SZ000983 西山煤电
SH601898 中煤能源	SZ000562 宏源证券	SZ000999 华润三九
SH601899 紫金矿业	SZ000568 泸州老窖	SZ002001 新 和 成
SH601901 方正证券	SZ000581 威孚高科	SZ002007 华兰生物
SH601918 国投新集	SZ000623 吉林敖东	SZ002024 苏宁电器
SH601919 中国远洋	SZ000625 长安汽车	SZ002038 双鹭药业
SH601928 凤凰传媒	SZ000629 攀钢钒钛	SZ002069 獐 子 岛
SH601933 永辉超市	SZ000630 铜陵有色	SZ002073 软控股份
SH601939 建设银行	SZ000651 格力电器	SZ002081 金 螳螂
SH601958 金钼股份	SZ000680 山推股份	SZ002092 中泰化学
SH601988 中国银行	SZ000686 东北证券	SZ002106 莱宝高科
SH601989 中国重工	SZ000703 恒逸石化	SZ002122 天马股份
SH601991 大唐发电	SZ000709 河北钢铁	SZ002128 露天煤业
SH601992 金隅股份	SZ000718 苏宁环球	SZ002142 宁波银行
SH601998 中信银行	SZ000725 京东方 A	SZ002146 荣盛发展
SZ000001 平安银行	SZ000728 国元证券	SZ002155 辰州矿业
SZ000002 万 科 A	SZ000729 燕京啤酒	SZ002202 金风科技

华 中 科 技 大 学 硕 士 学 位 论 文

SZ000009 中国宝安	SZ000758 中色股份	SZ002241 歌尔声学
SZ000012 南 玻 A	SZ000768 西飞国际	SZ002299 圣农发展
SZ000021 长城开发	SZ000776 广发证券	SZ002304 洋河股份
SZ000024 招商地产	SZ000778 新兴铸管	SZ002310 东方园林
SZ000039 中集集团	SZ000780 平庄能源	SZ002344 海宁皮城
SZ000046 泛海建设	SZ000783 长江证券	SZ002353 杰瑞股份
SZ000059 辽通化工	SZ000792 盐湖股份	SZ002378 章源钨业
SZ000060 中金岭南	SZ000800 一汽轿车	SZ002385 大北农
SZ000061 农 产 品	SZ000807 云铝股份	SZ002399 海普瑞
SZ000063 中兴通讯	SZ000825 太钢不锈	SZ002405 四维图新
SZ000069 华侨城 A	SZ000839 中信国安	SZ002415 海康威视
SZ000100TCL 集团	SZ000858 五 粮 液	SZ002422 科伦药业
SZ000157 中联重科	SZ000869 张 裕 A	SZ002431 棕榈园林
SZ000338 潍柴动力	SZ000876 新 希 望	SZ002493 荣盛石化
SZ000401 冀东水泥	SZ000878 云南铜业	SZ002500 山西证券
SZ000402 金 融 街	SZ000895 双汇发展	SZ002570 贝因美
SZ000898 鞍钢股份	SZ002594 比亚迪	SZ002603 以岭药业