

指数组合优化方法、模型与应用

—基于上证 180 指数的实证研究

联合课题组：

同济大学现代金融研究所
天同基金管理有限公司

负责人：

陈伟忠

课题协调人：

上海证券交易所 王建春 司徒大年

课题组成员：

朱顺 陈春锋 叶振飞 马玉林 施红俊

摘要

本文以上证 180 指数为标的，研究了从 2002 年 7 月 1 日上证 180 指数首次发布之日起，一直到 2004 年 3 月 15 日为止的这个时间段中的指数组合优化问题。文中分别研究了优化复制和抽样复制这两种指数组合优化的方法、模型及运用，而且为了比较，对完全复制问题也做了研究。每一种复制方法下都同时考查了四个组合（即创设问题以及随后的三次调整）的追踪效果（包括追踪误差、交易成本以及组合流动性等）。针对优化复制和抽样复制问题，文中构造了两类模型（即一般模型和流动性惩罚约束模型），并运用了两种算法（即二次序贯规划法（SQP）和遗传算法（GA））在 MATLAB6.5 环境下进行优化求解。抽样复制中先在 SAS8.2 环境下根据个股流动性和调整流通市值两个指标完成抽样，个股流动性的衡量是基于市场微观结构理论的研究成果利用高频数据完成的。此外对天同 180 指数基金和华安 180 指数增强型基金的追踪效果也做了简要评估。研究的主要结论如下：

1. 优化复制和抽样复制无论在样本内还是在样本外的追踪效果都相当不错。
2. 不过相对而言，完全复制的复制精度最高，其次是优化复制，最后是抽样复制。
3. 但从交易成本看，抽样复制最小，其次是优化复制，最后是完全复制。
4. 而从仓位变动对市场的冲击成本看，无论选择哪种复制方法，冲击成本都较小且可控。
5. 其它条件不变的情况下，样本内的时间窗越长，追踪效果越好，而且样本内追踪效果越好，样本外的表现也越好。
6. 资产配置时由于最小交易手数的限制而进行的个股买卖手数取整对现金误差的影响可以忽略不计，对组合样本内追踪误差的影响不能确定，可能加大，也可能减少组合样本内的追踪误差，而且因时间段不同而有所差异。
7. 从追踪组合中个股的集中度看，抽样复制较高，优化复制其次，而完全复制最分散。不过无论哪种方法，都没有突破模型中设定的比例上限。抽样复制中由于考虑了行业的分散性和代表性，因此尽管个股集中度略高一点，也不影响组合的分散化。
8. 流动性惩罚约束模型达到了预期的效果，即在没有严重恶化组合追踪误差的情况下，使得模型优化计算的结果能把更多的资金配置在比较有流动性的资产上。
9. 尽管 GA 在 SQP 计算结果的基础上都有所改进，但改进的幅度非常小，而且并没有比单纯应用 SQP 求解的结果好。可见 SQP 算法已经比较好地找到了模型的最优解，GA 算法能够改进的余地很小。
10. 本研究中优化后的指数组合的追踪效果比天同 180 指数基金、华安 180 指数增强型基金的实际追踪效果要好，这从侧面支持了本研究具有一定的实际应用价值。

关键词

优化复制 抽样复制 完全复制 一般模型 流动性惩罚约束模型 SQP 算法 GA 算法

目 录

0 . 引言	5
0 . 1 证券价格指数的演变	5
0 . 2 指数类金融产品创新	5
0 . 3 指数组合优化研究及其意义	7
1 . 文献回顾	9
1 . 1 问题的提出	9
1 . 2 指数复制方法	9
1 . 3 优化模型与算法	11
1 . 4 参数选择	15
1 . 5 研究的不足与改进	16
2 . 模型	18
2 . 1 一般模型	18
2 . 1 . 1 优化复制问题	18
2 . 1 . 2 抽样复制问题	22
2 . 1 . 3 完全复制问题	26
2 . 1 . 4 样本外绩效考察	27
2 . 2 流动性惩罚约束模型	27
2 . 2 . 1 基于市场冲击成本考虑的流动性惩罚约束	28
2 . 2 . 2 改进后的模型	29
3 . 数据及计算	32
3 . 1 数据来源	32
3 . 2 参数设定	32
3 . 3 数据处理	33
3 . 4 计算结果	33
3 . 4 . 1 一般模型下的计算结果	34
3 . 4 . 1 . 1 优化计算的迭代信息	34
3 . 4 . 1 . 2 追踪组合资产配置	35
3 . 4 . 1 . 3 组合分散化	36
3 . 4 . 1 . 4 市场冲击成本	37
3 . 4 . 1 . 5 追踪误差与相关系数	38
3 . 4 . 1 . 6 追踪组合的表现	39
3 . 4 . 1 . 7 追踪组合构成比较	42
3 . 4 . 2 流动性惩罚约束模型下的计算结果	43
3 . 4 . 2 . 1 优化计算的迭代信息	43
3 . 4 . 2 . 2 追踪组合资产配置	44
3 . 4 . 2 . 3 组合分散化	45
3 . 4 . 2 . 4 市场冲击成本及流动性改进	46
3 . 4 . 2 . 5 追踪误差与相关系数	48
3 . 4 . 2 . 6 追踪组合的表现	50
3 . 4 . 2 . 7 追踪组合构成比较	52
4 . 算法改进	54
4 . 1 遗传算法	54
4 . 2 基于 MATLAB 的遗传算法工具箱	55

4.3 流动性惩罚约束模型的 GA 求解	55
4.4 计算结果	57
4.4.1 基于 SQP 的 GA 计算	57
4.4.2 追踪组合资产配置	58
4.4.3 组合分散化	60
4.4.4 市场冲击成本及流动性改进	60
4.4.5 追踪误差与相关系数	61
4.4.6 追踪组合的表现	63
4.4.7 追踪组合构成比较	66
5. 实践评估	68
5.1 评估目的及数据来源	68
5.2 天同 180 指数基金	69
5.3 华安 180 指数增强型基金	72
6. 主要结论	76
主要参考文献	81
附录	83
致谢	100

0. 引言

0.1 证券价格指数的演变

1884 年 7 月 3 日，美国道琼斯公司的创始人查尔斯·道（Charles Dow）首次公开发表了人类有史以来的第一只证券价格指数。一百多年过去了，随着时代的变迁和金融市场的发展，今日的证券价格指数已经发生了很大的变化。

1. 证券价格指数的体系演变。今日的证券价格指数已经不仅仅是股票价格指数这种单一的类型了，而是由股价指数、债券指数、商品指数、房地产指数以及对冲基金指数等多种类型并存的、多层次、多角度的证券价格指数体系（indices family）。这些指数还可以根据市场空间范围的不同分为行业、国家、区域乃至全球指数，而且还可以根据风格定位的不同分为价值、成长以及大盘、中盘、小盘等风格指数。目前仅道琼斯公司一家每天要计算和发布的指数数目超过了 20,000 只，其中包括需要实时计算发布的 800 多只实时指数（real-time indices）。这些指数构成的体系包含了二十几个系列，每一个系列下又有好几只指数，比如最近刚刚开发完成的道琼斯全球指数系列（global indices series）就包含了 3,000 多只独立的指数，范围涵盖了全球 36 个国家的 5,000 多家上市公司。而标准普尔、摩根斯坦利资本国际（MSCI）、富时（FTSE）以及罗素（Russell）、威尔希尔（Wilshire）等几家金融公司也是很重要的指数服务商，由其编制的指数数目也相当可观。不同的指数代表不同的风险收益结构，因此指数体系的日益细化和多样化正好可以满足不同的投资者或不同风险收益偏好的投资需求。

2. 证券价格指数的功能演变。当年查尔斯·道首创道琼斯工业平均指数（DJIA），其目的是为了给《每日通讯》的读者们提供一个了解股市整体动态的窗口，因此人类历史上的第一只证券价格指数是作为证券市场乃至整个国家经济社会“晴雨表”的这种以股市标尺的身份出现的，这是证券价格指数的初始功能。然而到了今天，随着全球风险管理的日益盛行以及金融产品创新的急速发展，证券价格指数的功能开始发生了很大的变化。以强调标的证券的交易量、换手率等流动性指标和成长性、低成本可复制性、编制的透明性以及指数与投资需求的对应性等特征的证券价格指数的可交易性（tradable）和可投资性（investable）功能变得日益重要，指数编制者专职负责指数的编制和维护工作，并通过收取授权使用费（license fee）从而第一次成为了指数服务商。

0.2 指数类金融产品创新

证券价格指数的演变也宣告了金融产品创新新纪元的到来。在最近的十几年中，全球范围内围绕着证券价格指数的金融产品创新可谓如火如荼，蔚为壮观。从指数基金、

指数期货、指数期权到上市交易基金以及指数联动型债券、指数联动型存款等不一而足，层出不穷。

1. 指数组合与指数基金。这些产品的背后就是指数化投资思想，这种思想的核心就是构建指数组合，从而达到复制指数的目的。1971 年美国威弗银行(Wells Fargo Bank)为其客户萨姆索纳养老基金(Samsonite Pension Fund)设计了世界上第一个指数组合。1976 年美国前卫基金管理公司(Vanguard Fund)推出了世界上第一只真正意义上的指数基金——追踪标准普尔 500 指数的 Vanguard 500 指数基金。80 年代末美国全国只有 43 只指数基金在运作，资产规模不过百亿美元。而目前全球约有 3 兆美元的资金是采用指数作为投资标的的，仅追踪标准普尔系列指数的基金资产就超过了 1.5 万亿美元，指数基金在美国和英国基金市场中的占有率分别占到了 30%和 21%。

2. 指数期货、指数期权及指数权证。指数期货、指数期权及指数权证(index warrant)与指数组合和指数基金不同，其不是作为一种指数化投资工具被引入市场的，而是出于对风险规避的考虑而加以创设的。1982 年 2 月美国密苏里州堪萨斯农产品交易所(KCBT)首次推出了股票指数期货交易。同年 4 月，芝加哥商品交易所(CME)开展指数期货业务。5 月，纽约期货交易所(NYFE)也开展了指数期货业务。此后股票指数期货交易成为金融期货市场上最热门和发展最快的一种期货交易。据美国期货业协会(FIA)统计，目前全球股指期货的交易量已经超过了 25.2 万亿美元，占金融期货交易总量的 1/4 左右。指数期权交易则由芝加哥期权交易所(CBOE)于 1983 年首先推出。今日美国芝加哥期权交易所的标准普尔 500 指数期权、费城股票交易所的 TOPIX 指数期权等都是国际上著名的股指期货交易品种。指数权证则是一种与证券价格指数相关联的认股权证，持有者有权在特定日期结束之前根据事先约定的价格水平买入或卖出一定数量的证券价格指数。在过去的十年中，指数权证的交易日趋活跃，如日经卖出指数权证(Nikkei put warrants)的交易量就占到了美国股票交易所(AMEX)交易总量的 15%。

3. 交易所交易基金(Exchange-Traded Funds ,简称 ETFs)。从美国的实践看，绝大多数 ETFs 都以追踪某一证券价格指数为目标，通过分散化投资和被动式管理来取得市场的平均收益。因此同指数组合和指数基金一样，其背后也是指数化投资思想，只不过是其为投资者提供了一二级市场交易同时并存的双重交易模式(dual trade pattern)，这种交易模式的存在还产生了独特的套利机制，这些都保证了 ETFs 比传统的指数组合和指数基金具有更高的市场流动性和价格稳定性。实际上 ETFs 是在融合了指指数基金、开放式基金和封闭式基金之所长，又创造了自身的一些独特优势的一种混合型金融创新产品，因此被称为是第三代投资基金而更受投资者的欢迎。自 1993 年美国 SSgA 发行并由美国证券交易所(AMEX)上市交易的第一只严格意义上的 ETF——追踪标准普尔 500 的 SPDRs 以来，短短几年时间，全球 ETFs 市场无论在资产规模上，还是在产品数量上，都获得了极大的发展。据美国投资公司协会(ICI)统计，截止 2003 年 4 月，全球 ETFs 的资产总额已经超过 1,500 亿美元，10 年间总计成长了 300 多倍。

4. 其它与指数挂钩的种种结构性金融产品 (structured finance products)。如指数联动型债券 (index-linked bond)、指数联动型存款 (index-linked term deposits) 等。如果把服务也看作是一种产品的话, 还有指数定制服务 (customized index service) 和指数追踪服务 (index tracking service)。后两种盈利模式最近一两年刚刚发展起来, 但已经给指数编制商带来了可观的回报。Indextracker.ch 就是一家为投资者提供指数追踪服务的网上服务商, 此外美国著名的中介服务机构 Indexinvestor.com 等也是该行业的杰出代表。

0.3 指数组合优化研究及其意义

然而, 无论是指数基金、指数期货、指数期权还是 ETFs 等都要涉及到的一个共同的东西就是指数组合 (index portfolio)。对指数基金、ETFs 来讲, 其产品本身实际上就是一个指数组合 (只是其交易模式和产品架构有所不同), 而对指数期货、指数期权等其它指数类金融产品来讲, 通过构建指数组合来进行以对冲交易为主要目的的指数套利 (index arbitrage) 操作是规避金融风险的必不可少的操作手法。所谓的指数套利就是指当指数衍生品的价格贴现值大于现货指数时, 投资者通过买入现货指数 (即在现货市场构建指数组合以复制指数) 并卖出指数衍生品 (反之, 则进行反向操作), 从而获得无风险利润的交易活动。

因此某种程度上说, 指数组合的构建及其优化问题是指数类金融产品包括指数基金、ETFs 产品设计、指数套利以及实施指数化投资策略中的一个核心环节。换句话说, 指数组合的构建及其优化在指数类金融产品 (包括指数基金、ETFs) 的产品设计、指数套利以及实施指数化投资策略中有着重要而广泛的应用。

指数组合优化的内涵包含广义与狭义两个层面: 广义上的指数组合优化指的是“指数组合的优化管理”, 包括如何选择最优的标的指数、如何复制指数、如何持有最佳的现金比例以及如何实施风格指数化 (style indexing) 和增强型指数化 (enhanced indexing) 策略、如何融合主动管理型优势实施积极指数化 (active indexing) 策略等; 狭义上的指数组合优化仅仅指“标的指数的优化复制”, 包括如何构建优化模型和选择优化算法、如何控制追踪误差和追踪成本等。本课题所要研究的指数组合优化方法、模型与应用就是从狭义上的指数组合优化的内涵出发, 以上证 180 指数为例, 来研究标的指数给定下的指数优化复制问题 (包括下文所要说的优化抽样复制和分层抽样复制两种方法)。因此在本研究中“指数组合优化”与“上证 180 指数的优化复制”是一个概念。

我国证券市场从成立至今, 已经有了十几年的历史了。在这十几年的发展历程中, 随着我国经济体制改革的深化和经济、金融的进一步发展, 我国证券市场无论是市场规模、投资者数量、结构, 还是投资理念、资产管理模式等都发生了巨大的变化。然而其中一个很重要的变化是, 随着 1999 年 7 月 8 日基金普丰和基金兴和同时登台亮相, 指数化投资方式开始被引入到国内证券市场中来。经过短短的几年时间的发展, 目前我国

共有各类“准指数基金”(由于这些基金的产品架构与交易模式等诸多方面与国外真正的指数基金比还有很大的差异,而且目前国内对这些基金的投资品种和投资比例还有政策上的限制,因此称之为“准指数基金”)9只,包括天同180、华安180、融通100、易方达50、基金景福、博时裕富和长盛中信全债等。尽管以上几只指数基金还不是纯粹意义上真正的指数基金,然而仅以2003年前8个月就发行4只指数基金的速度看,指数化投资在我国急剧升温的趋势已经很明显,那些追求低风险、长期性、有稳定收益的投资者尤其是社保基金管理者、境外合格的机构投资者(QFII)对指数基金的需求尤为强烈。目前沪深交易所都在研究ETFs在国内金融市场的推出问题。此外,2004年2月1日国务院发布了关于推进资本市场改革开放和稳定发展的九点意见,为中国股市的发展提供了一个纲领性的、具有历史意义的重要文件,其中引人注目的一点是强调要健全资本市场体系、丰富证券投资品种。因此可以预见,在不远的将来,指数化投资方式在我国必将更为普及。

指数化投资方式在我国的引入以及指数类金融产品创新在我国国内的急剧升温给理论界与实务界提出了很多非常重要而又实际的问题,比如指数化投资需要一些什么样的前提条件?指数化投资在我国是否有优势?什么样的证券价格指数适合作为资产跟踪的标的亦或作为股指期货、期权等其它指数类金融产品创新的基础?可喜的是,这些问题已经引起了研究者的注意,沪深交易所和一些实力较强的券商以及国内高校的一些研究机构都已经就这些问题展开了研究,也取得了一些重要的研究成果。然而围绕着这些问题的另外一个核心,就是指数组合的优化问题,尤其是基于我国国内证券市场实际的指数组合优化问题目前国内还没有引起足够的重视。为此,本课题的研究力图弥补这一欠缺,同时也为推动指数化投资在我国的应用与发展以及进行包括指数、指数基金、ETFs等产品创新方面提供一些理论指导。这也是本课题研究的意义所在。

1. 文献回顾

1.1 问题的提出

应当说国外不仅很早就对指数化投资的科学性与可行性进行了深入的研究,而且对指数追踪方法、指数追踪效果等众多方面都有很深入的研究。而且随着最优化理论的发展和计算机技术的革新,更多、更复杂的新理论与新方法被应用到指数组合的优化研究中来,这也正是当前国外研究的一个热点和趋势。纵览指数组合优化研究的有关文献,研究者试图在回答这样的一些问题,即:

1. 如何设定优化目标:是战胜指数(beating index)还是追踪指数(tracking index)?
2. 如何实现优化目标:除了要考虑追踪误差(tracking error)之外,要不要考虑超额收益(excess return),或者是重点考虑有正偏差的追踪误差?
3. 如何基于证券市场的实际情况来限定不同的约束条件:包括投资资金规模、交易成本、冲击成本以及最少交易手数等?
4. 如何构建优化模型和选择优化算法?这是研究者关注的重点,多数文献都把重心放在这个问题的研究上。下面文献回顾中的“优化模型与算法”部分比较详细地对此加以综述。

1.2 指数复制方法

本部分是基于投资实践和理论研究两个方面所做的综述,它界定了指数组合优化研究的边界,同时也回答了指数组合优化的必要性这一问题,因此有必要对此先予以回顾。

复制指数的方法有两大类:即完全复制(full replicate)和不完全复制(即优化复制, optimized replicate)。顾名思义,完全复制就是购买标的指数中的所有成份证券,并且按照每种成份证券在标的指数中的权重确定购买的比例以构建指数组合从而达到复制指数的目的。以标准普尔 500 指数为例,按市值比重购入全部 500 种成分股就可以完全复制指数。当然,实际情况要复杂的多,因为指数是一个“纸面上的组合”(paper portfolio),每种成份证券在标的指数中的权重时时刻刻在发生变化,以某一时刻的相对权重值来确定组合的结构显然不能保证组合的走势与指数完全一致,因此实务中即便是完全复制也要根据追踪误差的偏离状况对组合进行动态调整。不过,相对于其它复制方法来讲,这种方法的思想还是比较简单明了,而且构建的指数组合较好地继承了标的指数所具有的代表性和投资的分散性,较容易获得比较小的追踪误差(tracking error)。然而这种方法有其优势,但也并非成本—收益最优,其至少有以下不足(也就是优化复制的必要性):

1. 完全复制指数，特别是成分股较多的指数要占用很大的资金。比如完全复制标准普尔全球 1200 指数(S&P Global 1200 Index)或者威尔希尔 5000 全市场指数(Wilshire 5000 Total Market Index) 等，所需资金量巨大，对于一般的投资者，根本无此实力来完全复制指数。即便是有实力的大机构投资者，这么大量资金的占用也意味着其所面临的风险加大；

2. 完全复制指数的指数组合所承受的冲击成本 (impact cost) 较高。通常这个组合的规模巨大，如果市场容量较小，市场深度不足，短时间内买入或抛出整个指数组合必然会对市场造成很大的冲击，导致股票价格向不利于投资者的方向变动，也使得构建指数组合以及随后的组合调整所承受的冲击成本较高；

3. 较多的成分证券交易将导致下单及撮合成交的时间延长，加大了指数组组合所承担的价格波动风险；

4. 完全复制指数还可能面临很大的流动性风险。当标的指数的某个或某些成分证券的流动性严重不足时，如果再加上价格被操纵 (price manipulation) 的可能，则选择完全复制指数而没有剔除掉这些 “ 问题证券 ” 就可能面临很大的损失。

5. 完全复制指数还可能导致较高的调整频率和追踪成本。标的指数内任何一只成分证券发生调整，指数组合也随之发生变动，因此调整频率和追踪成本相对较高。

所谓的优化复制指的是根据预先设定的标准剔除掉部分成分证券并对其在组合中的相对权重进行优化再配置，从而使得构建出来的指数组合的追踪成本及其与标的指数之间的追踪误差控制在可以接受的范围之内。优化复制的方法又可以进一步细分为优化抽样复制(optimized sampling replication)和分层抽样复制(stratified-sampling replication)两种。前者是单阶段优化法，即把抽样和权重优化再配置同时进行，通过对标的指数中各成份证券的历史数据和相关指标进行分析，并利用最优化算法模型来寻找一个可控数量和一定误差范围内的投资组合，使该组合相对标的指数的追踪误差最小。如果优化求解出来的某种成分证券的权重值为零，就剔除该成分证券；不为零，就保留该证券并按计算出来的权重值配置资产，这种做法使得抽样和权重优化再配置能同时进行。可见，优化抽样复制中的抽样是计算机自动完成的，也是优化计算的一个结果，不需要额外的抽样工作，因此优化抽样复制通常就简称为 “ 优化复制 ”。与之对应，分层抽样复制就简称为 “ 抽样复制 ”。本研究中也沿用这种称法，下文中的 “ 优化复制 ” 即指 “ 优化抽样复制 ”，而 “ 抽样复制 ” 即指 “ 分层抽样复制 ”。

与优化复制不同，抽样复制则是两阶段优化法，即第一阶段是抽样，一般做法是先将市场上所有证券按某一特征分成几个或几十个互不相交的行业类别，以不同行业类别在目标指数中所占的权重来决定在该行业类别内的投资量，然后分别从每个行业类别中按照与行业整体相关性、总市值及市盈率水平等有关指标的综合评价结果，选出最能代表该行业类别的样本证券；样本选出之后的第二阶段则是权重的优化再配置，通过对最

优化算法模型的应用，求出追踪组合内各样本证券的最优权重，使得组合的表现与标的指数相一致，即追踪误差最小，同时保证较小的调整频率和追踪成本。

1.3 优化模型与算法

可见在指数组合优化中，无论是优化复制还是抽样复制都要用到最优化模型与算法来进行求解，这是剔除部分成分证券之后进行权重优化再配置的必经步骤。以下部分就对国外众多学者如何构建最优化模型并应用最优化算法来对指数组合进行优化求解的相关研究进行简要回顾。模型与算法之间是统一的，在选择合适的算法之前更关键的是构建科学合理的数学模型，当然有时为了算法也需要适当地简化数学模型。因此下文以算法为序，重点回顾每种算法下所构建的优化模型，限于篇幅，具体的求解过程从略。此外，为了理解上的方便，文中的数学符号不再沿用原模型的符号，而是做了统一。

1. 二次规划 (quadric programming)。 Roll (1992) 在 Markowitz 建立的均值 - 方差模型的基础上研究了组合追踪误差最小化问题，其通过最小化组合收益率与标的指数收益率之差的平方和最小化，即追踪误差波动率的最小化来改进组合对标的指数的复制效果，见下式 (a) 至 (e)。设组合中包含有 n 中资产（如股票和债券），令 X 为 n 种资产的收益矩阵， w 为组合中各资产的权重矢量， Y 代表组合所要跟踪的标的指数的收益矢量。则均值 - 方差模型把组合的跟踪误差定义为：

$$\varepsilon = Y - Xw \quad \text{..... (a)}$$

$$Y \in R^T, X \in R^{T \times n}, w \in R^n, \varepsilon \in R^T$$

其中 T 为观测值数目， n 为组合中资产的数量， R 代表实数空间。

定义追踪误差为组合收益率与标的指数收益率之差的平方和 $\varepsilon' \varepsilon$ ，通过 w 的变化使得追踪误差最小，即目标函数为：

$$\min \varepsilon' \varepsilon = \min (Y - Xw)' (Y - Xw) \quad \text{..... (b)}$$

$$\text{求解得最优权重为：} w = (X'X)^{-1} X'Y \quad \text{..... (c)}$$

此外可以在模型中加入组合资产不能卖空（权重不能为负）以及组合资产权重之和为 1 的约束：

$$(1) Aw \geq b \quad \text{..... (d)}$$

其中 $A \in R^{k \times n}, w \in R^n, b \in R^k$ (k 为约束个数)；

$$(2) \sum_{w=1} \dots\dots\dots (e)$$

有时为了保证组合的充分分散化,还可能要求每只股票或债券在组合中的最大权重不超过某一预先给定的值。因此约束条件的多少依实际情况而定。

Y. Tabata 和 E. Takeda (1995) 也在 Markowitz 的有效边界和均值 - 方差模型的基础上研究了指数基金的优化问题。其对追踪误差的定义与 Roll 的一致,不同的是 Y. Tabata 和 E. Takeda 在 Markowitz 的有效边界的基础上发展出了一套优化算法(算法内容略)。其经过一系列的推理和简化,最小化追踪误差最终由下式给定:

$$E\{Y-X\}^2 = w_n' A_{nm} w_n - w_n' A_{nm} A_{mm}^{-1} A_{nm}' w_n + \frac{[w_n' A_{nm} A_{mm}^{-1} 1_m - 1]^2}{[1_m' A_{mm}^{-1} 1_m]}$$

其中, Y 和 X 分别代表有 n 种成分证券的指数收益矢量和有 m 种证券的组合收益矢量, A_{nm} 、 A_{nm} 和 A_{mm} 分别为对应维数的特征根实对称矩阵, w_n 为指数成分证券的权重矢量。特别地,如果组合与标的指数一致($m=n$),则有待求解权重矢量等于 w_n , 并且有 $E\{Y-X\}^2 = 0$ 。

2. 线性规划 (lineal programming)。在 Konno (1991) 和 Clarke (1994) 等人研究的基础上, Markus Rudolf、Hans-Jurgen、Wolter 和 Heinz Zimmermann (1999) 把追踪误差定义为组合收益率与标的指数收益率之间的差值,并提出了以下 (f) 至 (i) 4 个线性追踪误差最小化模型,分别是均值绝对离差 (mean absolute deviations, MAD) 模型、均值绝对向下偏差 (mean absolute downside deviation, MADD) 模型、最小最大化 (minimize maximum negative deviation, MinMax) 模型、向下最小最大化 (downside MinMax, DMinMax) 模型。主要的数学表达式如下:

$$TE_{MAD} = \min_w l'(|Xw - Y|) \dots\dots\dots (f)$$

$$TE_{MADD} = \min_w l'(|\bar{X}w - \bar{Y}|), \text{ 其中 } \bar{X}_t w < \bar{Y}_t \dots\dots\dots (g)$$

$$TE_{MinMax} = \max_w |Xw - Y| \dots\dots\dots (h)$$

$$TE_{DMinMax} = \max_w |\bar{X}w - \bar{Y}|, \text{ 其中 } \bar{X}_t w < \bar{Y}_t \dots\dots\dots (i)$$

其中, X 为 n 种资产(如股票和债券)的收益矩阵, w 为组合中各资产的权重矢量, Y 代表组合所要跟踪的标的指数的收益矢量; \bar{X} 和 \bar{Y} 分别是仅包含指数收益率低于组合收益率时的组合收益矩阵和指数收益矢量。

Markus Rudolf、Hans-Jurgen、Wolter 和 Heinz Zimmermann 还把以上 4 个线性模型应用于美国、日本、英国、德国、法国、瑞士六国的股票市场指数和摩根斯坦利资本国际指数 (MSCI) 的指数追踪实验,结果表明线性规划模型优于二次规划模型,把追踪

误差定义为线性关系的组合收益率与标的指数收益率之间的差值也比传统的二次平方和的定义更接近现实。

3. 鲁棒回归 (robust regression)。 Gunter Bamberg 和 Niklas Wagner (2000) 在一项研究中首先对应用线性回归法来求解最优的指数组合问题进行了探讨,发现无论是使用复权调整后的股价还是未复权调整的股价以及使用收益率还是累积收益率建立的线性回归模型,都将违反古典线性回归最小二乘法的一些经典假设。然而 Gunter Bamberg 和 Niklas Wagner 认为这不影响回归方法在求解最优指数组合问题中的应用,他们采用鲁棒回归法来避开最小二乘法的假设限制,并应用德国 DAX 指数进行了实证检验。为此,他们构建了如下模型来估计权重参数:

$$\varepsilon_t = Y_t - X_t w_t \dots\dots\dots(j)$$

其中 ε_t 为误差值, Y_t 为因变量(指数收益矢量), X_t 为自变量(成分证券收益矢量), w_t 为权重矢量, $t = 1, 2, \dots, T$;

引入规模参数(scale parameter) $\sigma > 0$, 把总损失函数(aggregate loss function)定义为:

$$Q(X, \sigma) = \sum_{t=1}^T \rho((Y_t - X_t w_t) / \sigma) \dots\dots\dots(k)$$

$$\text{最大似然估计为: } (\hat{X}, \hat{\sigma}) = \arg \min_{(X, \sigma) \in Z} Q(X, \sigma) \dots\dots\dots(1)$$

其中 $Z = \{(X, \sigma) : X \in R^N, \sigma > 0\}$

应用 Huber 最大似然回归估计, 则损失函数可以变换为:

$$\rho_H(u_t) = \begin{cases} \frac{1}{2} u_t^2 + \frac{1}{2} \beta & \text{if } |u_t| \leq c \\ c |u_t| - \frac{1}{2} c + \frac{1}{2} \beta & \text{if } |u_t| > c \end{cases} \dots\dots\dots(m)$$

其中 $\beta = E[\psi_H(Z^2)]$, $Z \sim N(1, 0)$, u_t 为标准残差 ($u_t = \varepsilon_t / \sigma$), c 为常数;

应用 Turkey 最大似然回归估计, 则损失函数可以变换为:

$$\rho_T(u_t) = \begin{cases} \frac{1}{2} d^4 u_t^2 - \frac{1}{2} d^2 u_t^4 + \frac{1}{6} u_t^6 & \text{if } |u_t| \leq d \\ \frac{1}{6} d^6 & \text{if } |u_t| > d \end{cases} \dots\dots\dots(n)$$

其中 d 为常数。

在此模型的基础上, 通过最小化上述损失函数((k) \(\lambda\)(m) 或 (n)), 从而得到待求变量 w_t 。Gunter Bamberg 和 Niklas Wagner 应用德国 DAX 指数进行的实证检验表明,

当样本区间的市场指数波动较大时，应用鲁棒回归法来计算最优指数组合效果较好。

4. 蒙特卡洛模拟 (Monte Carlo simulation)。 Francesco Corielli 和 Massimiliano Marcellino (2002) 基于动态因素模型 (dynamic factor model) 考虑了指数成分股动态变化的特征，也通过最小化样本期内损失函数的方法来解决指数组合的最优化问题。他们以 EuroSTOXX50 指数为样本做了实证研究，在具体的算法上使用了蒙特卡洛模拟来求解最优的成分股权重。蒙特卡洛模拟是用一系列随机数来寻找近似解的一种方法。Francesco Corielli 和 Massimiliano Marcellino 构建的模型如下：

假设以市值加权的标的指数的价值矩阵为： $I_t = wp_t$ ，并且价格基于以下因素模型实现进化：

$$p_t = \Lambda f_t + e_t \quad \text{..... (o)}$$

其中 w 为 $1 \times N$ 权重矢量， p_t 为 $N \times 1$ 价格矢量， f_t 为 $r \times 1$ 的因素 (factors) 矢量， Λ 为 $N \times r$ 的因素权重 (loadings) 矩阵， e_t 为 $N \times 1$ 的随机扰动项；

此外，定义追踪指数的组合市值为 \bar{I}_t ，则有：

$$\bar{I}_t = wSp_t \quad \text{.....(p)}$$

其中 w 为 $1 \times q$ 的权重矢量， S 为 $q \times N$ 的选择矩阵 (即从 N 个价格 p_t 中选出 q 个组成的矩阵)，可见就价格而言， \bar{I}_t 是 I_t 的一个子集，但同一资产在两者中各自占有的比重却不同。

通常损失函数定义为：

$$g(\eta, q) = \sum_{t=0}^T d(\eta_t) + c(q) \quad \text{.....(q)}$$

其中 $\eta_t = I_t - \bar{I}_t$ ， $c(q)$ 为惩罚因子；

通常在损失函数中加入该项以修正参数的变换，最后应用最小二乘法求解最小化损失函数，从而得出待求的最优权重 w 。

5. 遗传算法 (genetic algorithm, GA)。 遗传算法是模仿自然界的生物进化过程而设计的一个求解程序首先对问题的所有可能解 (candidate solutions) 随机选取一个固定数量的族群 (populations) 进行迭代 (iterative)，每次迭代进行时选择、交叉与变异三个基因操作子就发生作用，以便产生新的后代 (offspring)，然后这些新的染色体会经由适应度函数 (fitness function) 来评估出染色体的适应能力，适应力较强之染色体将被保留再进行迭代，直到问题的解被找到为止。基于这种思想，J. Shapcott (1992)、Dirk

Eddelbuttel 和 Marseilles (1996) 等人利用遗传算法来寻找最优指数组合问题。Dirk Eddelbuttel 和 Marseilles 用到了混合遗传算法 (Hybrid Genetic Algorithm), 他们构建的模型如下:

$$\text{目标函数: } \min_i (w_p - w_m)' V (w_p - w_m) \dots\dots\dots (r)$$

约束条件:

$$w_{pi} \geq 0, \text{当且仅当 } j_i = 1 \dots\dots\dots (s)$$

$$w_{pi} = 0, \text{当且仅当 } j_i = 0 \dots\dots\dots (t)$$

$$\sum_{i=1}^N j_i \leq h \dots\dots\dots (u)$$

$$\sum_{i=1}^N w_{pi} = 1 \dots\dots\dots (v)$$

其中 w_p 为待求的组合内各证券的权重, w_m 为标的指数内各证券的权重, V 为样本协方差对称矩阵, j_i 用于标识标的指数内各成分证券是否被选入了组合, 如果选入, 则 $j_i = 1$; 未选入, 则 $j_i = 0$, 因此 $j_i \in \{0,1\} \forall i = 1,2,\dots,N$ 。(u) 式约束了选入组合的成分证券的数量, 而 (v) 式要求权重之和为 1。通过最小化目标函数, 求出组合各成分证券的权重 w_{pi} 。他们对德国 DAX 指数进行实证研究的结论表明, 通过选择少量的成分股并通过追踪误差的最小化来寻找最优权重方面涉及到的复杂的算法问题完全可以由混合遗传算法加以解决。

此外, 研究者还尝试使用其它一些更复杂的方法来进行建模和求解, 如 J. E. Beasley、N. Meade 和 T. J. Chang (2001) 以及 M. Gilli 和 E. K llezi (2001) 等人采用启发式算法 (Heuristic Algorithm, HA) Paolo Dai Pra、Wolfgang J. Runggaldier、和 Marco Tolotti (2002) 等人采用利用随机控制 (stochastic control) 和顺向优化法 (pathwise optimality) David D. Yao、Shuzhong Zhang、和 Xun Yu Zhou (2003) 等人采用随机线性二次控制法 (stochastic linear quadratic control, SLQ) 并利用半限量规划 (semidefinite programming, SDP) 等方法求解。尽管这些方法并不常用, 但其运用无疑使得指数复制问题的研究更为深入。

1.4 参数选择

除了优化模型与算法之外, 还要考虑模型中的各种参数选择问题。这些参数包括:

1. 模型的构建期间：比如是基于前 6 个月的历史数据还是基于前 3 个月或 1 个月的历史数据进行优化计算比较合理？所谓的“合理”就是能更好地追踪指数，即样本外追踪误差较小。这是样本内（in-sample）时间窗口的设定问题。

2. 模型的考察期间：比如是考察模型之外 6 个月的指数组合追踪效果还是考察 3 个月、1 个月或 10 个交易日的追踪效果？这是样本外（out-of-sample）时间窗口的设定问题，这与优化计算的用途有关。比如，如果优化计算的目的是为短期指数套利决策服务，则比较注重 1 个月或 10 个交易日这样的短期考察期间的指数组合表现；而如果是为指数化投资决策服务，则往往在一个较长的时间期间内比如 3 个月或 6 个月内考察指数组合的追踪效果。样本外时间窗口的长度往往与指数组合调整的时间长度（进而与指数组合的调整频率）有关。

3. 抽样数目：根据复制精度的要求不同，成分股抽样数目也有所不同，所以抽样数目的多少首先取决于要求的追踪精度的高低。当然抽样方法、抽样的目的以及指数组合的结构要求等因素也会影响最终的抽样数目。

台湾的谢文良、李进生、谢素娟等人（2002）对此进行了研究，实证的结论表明：一般而言，构建期间越长越能更好地追踪指数，即样本外追踪误差越小，但过长的构建期间可能包含过时的价格信息，严重的还可能将不同结构的状况包含在同一构建期间内从而得出错误的结论；另外，一般情况下，追踪误差随着考核期间的变长而放大；成分股数目越多则越能精确地追踪指数，但包含的股数越多，交易成本和管理成本也相对增加，同时所需的资金规模也必须扩大，这增加了指数追踪的难度。

1.5 研究的不足与改进

从以上文献的简要回顾中可以看出，当前的研究中至少还存在有以下几个方面的不足：

1. 研究者忽略了对不同复制方法和其不同复制效果的对比研究。绝大多数的文献仅仅考虑了某一种复制方法（主要是优化复制）的具体应用，而没有把不同的复制方法同时放在一起加以比较。

2. 研究者还忽略了对不同算法的计算情况加以对比。尽管用到的算法很多，不同的算法之间似乎也都有优势，但对不同算法之间在计算结果上的差异问题，当前的研究中并没有加以对比分析。

3. 现有的研究中很少考虑个股以及组合整体的流动性问题，更没有考虑组合仓位调整所带来的市场冲击成本问题。

4. 在我国内陆，对指数化投资和指数衍生品的研究还刚刚起步，绝大多数的文献还停留在对一些基础问题的介绍上，对指数的优化复制问题，尤其是基于我国内陆证券

市场实际的指数的优化复制问题的研究还是空白。

这四个方面正是本研究努力去改进并力图有所突破的问题，具体如下：

1. 本研究将同时考虑三种复制方法：优化复制、抽样复制和完全复制，并以上证 180 指数加以实证研究。尽管前两种方法是本研究的重点，但为了比较，对完全复制问题也一并加以考虑，对不同复制方法下的不同的数学建模、数据处理以及复制效果等问题同时放在一起加以比较。

2. 本研究中除了用到序贯二次规划法 (Sequential Quadratic Programming ,SQP) 外，还将用到遗传算法 (genetic algorithm , GA) 进行优化求解。同时将这两种方法的具体应用及其计算结果加以简要比较。

3. 个股和组合整体的流动性问题以及组合仓位调整所带来的市场冲击成本问题在本研究中将重点加以考虑。本研究中将基于市场微观结构理论的一些研究成果并利用高频成交数据对个股的流动性问题进行综合评价，并在抽样复制中，以个股流动性指标抽样，力图提高指数组合的整体流动性。进而在此基础上，对原模型进行改进，加入基于市场冲击成本考虑的流动性惩罚约束，旨在没有恶化追踪误差的情况下，让模型优化计算的结果能把更多的资金配置在更有流动性的资产上。此外，通过限定个股在组合中的比重以及通过监测个股的仓位变动占其流通股本的比例等指标来分散组合的集中度，减少组合仓位调整所带来的市场冲击成本。

4. 显然本研究一定程度上将弥补国内在指数组合优化方面研究的不足。

2. 模型

2.1 一般模型

所谓的“一般模型”是与下文的“流动性惩罚约束模型”相区分而言的。前者是不考虑流动性惩罚约束，同时也是当今国外研究文献中常见的一种模型，所以称之为“一般模型”。后者则加入了基于市场冲击成本（impact cost）考虑的流动性惩罚约束，因此是对一般模型的改进。加入流动性惩罚约束的目的是在没有大幅度放大追踪误差的前提下使得模型计算的结果能让更多的资金配置于比较有流动性的资产上。流动性惩罚因子（liquidity penalty factor）的大小取值则是基于市场冲击成本的考虑计算出来的，具体下文详述。尽管在抽样复制中已经按照个股的流动性指标进行抽样，目的也是试图提高指数组合中的个股流动性，但不同的是只要按照这种抽样方法被抽样上来的个股其流动性大小在之后的模型计算中并没有区别对待，而加入流动性惩罚约束之后就可以对此加以区分，这样更贴近实际。

不过，无论是“一般模型”还是下文的“流动性惩罚约束模型”都考虑优化复制、抽样复制以及完全复制三种不同的情况。尽管前两种情况是本课题的研究对象，但为了对不同复制方法的复制效果进行比较，本课题对完全复制问题也一并予以计算。

2.1.1 优化复制问题

（一）基本思想

此中的优化复制就是上文所说的优化抽样复制，其是单阶段优化法，即让计算机自动完成抽样，而且抽样和权重优化再配置同时进行。具体讲，就是通过对标的指数中各成份证券的历史数据和相关指标进行分析，并利用最优化算法模型来寻找一个最佳的追踪组合，使得该组合相对标的指数的追踪误差最小。此中让计算机来回答哪些股票应当选入，而且对被选入股票的确切数目不加约束。如果优化求解出来的某只股票的权重值为零，就剔除该股票；不为零，就保留该股票，并按计算出来的权重值配置资产。

（二）模型构建

1. 目标函数

正如前述，所谓的“优化复制”就是通过权重的优化再配置来寻找一个含有“部分”成份证券的最优的追踪组合，并使得该组合相对标的指数的一些考核标准最优。这里的考核标准一般有两个，一是追踪误差，二是超额收益，由于本研究中设定的目标是追踪指数（track the index），而不是战胜指数（beat the index），因此只考虑追踪误差，不考

虑超额收益。在 Roll (1992)、Buckley 和 Korn (1998)、Cornnor 和 Leland (1995) 以及 Larsen 和 Resnick (1998) 等人的文献中把追踪误差 (tracking error, 简称 TE) 定义为追踪组合收益率 R_{pt} 和标的指数收益率 R_{it} 之差的方差, Beasley、Meade 和 Chang (2001) 不同意这种定义, 如果 $R_{pt} - R_{it}$ 为一个常数, 则其方差为零, 则基于这种定义的追踪误差为零, 显然这种定义忽略了这种情况下的 R_{pt} 与 R_{it} 之间的偏差, 因此他们采用了误差平方和的均值平方根 (root mean squared error) 的定义。本研究中对追踪误差的定义与 Beasley 等人一致, 即目标函数如下:

$$\text{Min}_{q_i} TE = \frac{1}{T} \sqrt{\sum_{t=1}^T \left(\ln\left(\frac{I_t}{I_{t-1}}\right) - \sum_{i=1}^N \left[\frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \ln\left(\frac{p_{it}}{p_{i,t-1}}\right) \right] \right)^2}$$

令 $R_{it} = \ln(I_t / I_{t-1})$ 和 $R_{it} = \ln(p_{it} / p_{i,t-1})$, 经代换得:

$$\text{Min}_{q_i} TE = \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N \left[\frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} R_{it} \right] \right)^2}$$

其中: T 是基于历史时期即样本内时间区间 $0, 1, 2, \dots, T$ 来考察股票与指数的价值变化, $T+1$ 代表一个决策时点 (decision point), 在该时点把当前组合调整为一个新的追踪组合; p_{it} 代表 t 时刻 ($t = 0, \dots, T$) 个股 i ($i = 1, 2, \dots, N$) 的价格, $\ln(p_{it} / p_{i,t-1})$ 即个股对数收益率 R_{it} ; \bar{p}_i 代表追踪组合调整 (或构建) 时刻的个股当日收盘价; I_t 代表 t 时刻指数的点位, $\ln(I_t / I_{t-1})$ 即指数对数收益率 R_{it} ; q_i 为决策变量 (即待求变量), 代表 $T+1$ 时点个股 i 的持有数量。

2. 需要考虑的约束条件

(1) 组合内个股集中度: 即组合内单只股票的持有数量不得超过一定的比例, 也不能小于一定的比例:

$$\alpha_i \leq \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \leq \beta_i$$

其中: α_i 为追踪组合中个股 i 的最小持有比例, 实践中 α_i 代表个股 i 的“最小持有

水平”(minimum holding level), 本模型中取值为 0; β_i 为追踪组合中个股 i 的最大持有比例, 则有 $0 \leq \alpha_i \leq \beta_i \leq 1$, β_i 限制了追踪组合中个股 i 的风险敞口 (risk exposure), 本模型中取值为 0.1;

(2) 资金规模: 即持有的追踪组合市值加上仓位调整成本等于初始资金规模:

$$\sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| = V$$

其中: \tilde{q}_i 为组合调整之前对应个股的持有量, 创设问题中 \tilde{q}_i 为零; γ 为交易中涉及到的佣金、印花税等费率的综合比例, 本模型中取值为 0.003; $\gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i|$ 为仓位调整所带来的交易成本; V 为初始资金规模, 其值等于每次调整前一日收盘后的追踪组合市值与因指数调整而被调出的个股悉数卖出而带来的交易成本的差额。本研究中一律假定最初 (即 2002 年 7 月 1 日) 的资金规模为 10 亿元人民币。

(3) 交易成本: 即仓位调整所带来的交易成本不能超过初始资金规模的一定比例。因指数调整而被调出的个股需悉数抛出, 因此而带来的交易成本是个给定值, 无法优化。能够加以考虑的是因指数调整而被调入的个股以及保留在指数内的个股的仓位变动所带来的交易成本, 其不能超过初始资金规模的一定比例, 此中为 c , 本模型中取值为 0.0005, 即仓位调整所带来的交易成本最多不超过初始资金规模的 0.0005:

$$\gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| \leq cV$$

(4) 市场冲击成本: 即个股仓位调整引起市场价格波动所带来的成本。创设中由于仓位变动大, 市场冲击成本高; 调整中仓位变动小, 市场冲击成本一般不大。此中对冲击成本的考虑主要有两个方面: 一是创设时通过限制单只股票买入的比例, 即根据买入数量占该股流通股本的比例来决定 α_i 和 β_i 的取值; 二是允许分阶段建仓, 来分散一次性大量建仓所可能带来的市场冲击成本。实际上这样的考虑是对可能的市场冲击成本的间接控制, 没有对市场冲击成本加以量化研究, 这就是流动性惩罚约束模型中需要进一步改进的地方。不过, 仍然需要指出的是, 对市场冲击成本的精确度量是不可能的, 所有的量化研究只是一种近似, 因为影响市场冲击成本的因素与影响股价波动的因素一样的多而取法确定。

(5) 最小交易手数: 由于国内证券市场允许的最小交易手数为 1 手, 因此在根据计算结果进行资产配置时需要对个股的买卖数量按照四舍五入法加以取整。当然也可以

在模型中加入整数约束，只是取整所带来的现金误差很小（在百万分之三左右），几乎可以忽略，因此本研究中仍是进行事后取整，没有考虑整数约束问题。

综上可得优化复制的数学模型为：

$$\begin{aligned}
 \min_{q_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N \left[\frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} R_{it} \right] \right)^2} \\
 s.t. & \\
 1) \quad &\alpha_i \leq \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \leq \beta_i \quad \dots\dots\dots (1) \\
 2) \quad &\sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| = V \\
 3) \quad &\gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| \leq cV
 \end{aligned}$$

追踪组合的初次建仓即创设问题中，由于 $\tilde{q}_i = 0$ ，因此通过定义权重变量

$w_i = \bar{p}_i q_i / \sum_{i=1}^N \bar{p}_i q_i$ 而把上述模型变换为：

$$\text{目标函数：} \min_{w_i} TE = \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N w_i R_{it} \right)^2}$$

约束条件：

- 1) $\alpha_i \leq w_i \leq \beta_i$;
- 2) 但要增加约束 $\sum_{i=1}^N w_i = 1$;

3) 由于决策变量为 w_i ，所以不必考虑规模的约束。而且由于 $\tilde{q}_i = 0$ ，模型（1）中

的约束 2) 变为 $\sum_{i=1}^N \bar{p}_i q_i = \frac{V}{1+\gamma}$ ，该约束在事后的资产配置即求解具体的个股手数时加上，具体体现在如下公式：

$$\sum_{i=1}^N \bar{p}_i q_i = \frac{V}{1+\gamma}$$

$$\text{由 } w_i = \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \text{ 得:}$$

$$q_i = \frac{w_i \sum_{i=1}^N \bar{p}_i q_i}{\bar{p}_i} = \frac{w_i \frac{V}{1+\gamma}}{\bar{p}_i}$$

4) 由于创设时, 交易成本是规模的一定比例的一个常数, 所以对交易成本无需加以约束, 模型(1)中的约束3)取消, 交易成本事后计算为 $\text{cost} = \gamma \sum_{i=1}^N \bar{p}_i q_i = \frac{\gamma V}{1+\gamma}$;

综上可得创设时的优化抽样复制模型:

$$\begin{aligned} \text{Min}_{w_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N w_i R_{it} \right)^2} \\ \text{s.t.} & \dots\dots\dots (2) \\ 1) & \alpha_i \leq w_i \leq \beta_i \\ 2) & \sum_{i=1}^N w_i = 1 \end{aligned}$$

2.1.2 抽样复制问题

(一) 基本思想

此中的抽样复制实际上就是上文所说的分层抽样复制, 其是两阶段优化法, 即第一阶段是抽样, 一般做法是先将市场上所有证券按某一特征分成几个或几十个互不相交的行业类别, 以不同行业类别在目标指数中所占的权重来决定在该行业类别内的投资量, 然后分别从每个行业类别中按照与行业整体相关性、总市值及市盈率水平等有关指标的综合评价结果, 选出最能代表该行业类别的样本证券; 本课题的做法与此大体一致, 只是顺序有所不同, 先按流动性和调整流通市值(按照180指数编制原理中规定的分级靠档方法对成分股的流通市值进行调整)两个指标对180指数的180只成分股进行综合排名, 再根据排名结果和行业分布状况进行抽样, 在综合考虑行业分布的情况下剔除综合排名靠后的部分成分股, 即选入剩下的成分股。第二阶段的做法一样, 样本选出之后则是权重的优化再配置, 通过对最优化算法模型的应用, 求出追踪组合内各样本证券的最优权重, 使得组合的表现与标的指数相一致, 即追踪误差最小, 同时保证较小的调整频率和追踪成本。

(二) 抽样方法

样本空间为 180 指数的 180 只成分股，即指数成分股之外的个股不加考虑。抽样所依据的指标是流动性和调整流通市值两个指标，每个指标的具体计算方法如下：

1. 流动性指标

对个股流动性的定义以及计算方法基于上证联合研究计划第二期课题“上海证券市场日内流动性与波动性研究”(上海交大-国泰君安课题组，课题负责人：杨朝军、李迅雷)的研究成果。该研究以现代证券市场微观结构理论为基础，根据西方学术界对市场流动性的研究成果和中国股市的运行特点，建立了股票市场流动性的指标体系，同时对这些指标的有效性(包括相关性和平稳性)进行了检验。该研究以原上证 30 指数样本股为研究对象，对这些股票从 2000 年 9 月 6 日至 2000 年 12 月 29 日期间的每分钟交易数据进行了实证分析，结果表明这些流动性衡量指标的定义、计算及综合评判方法比较的科学和可行，因此本课题直接基于以上研究成果对 180 指数成分股的流动性加以衡量。

(1) 定义指标

1) 市场宽度 (Breadth or Width)：即交易价格偏离市场有效价格的程度，它是投资者支付的流动性升水。市场宽度的观测值是做市商的买卖差额，只有当这个差额为零时，才可认为这个市场完全达到流动性。

由于我国国内的证券市场是指令驱动市场 (order-driven market)，没有做市商的买卖报价，因此对市宽度的衡量需要基于国内证券市场的实际情况进行变通。杨朝军等人的研究报告中对此界定如下：

$$W_{\alpha,\Delta} = 2 * (H_{\tau,\Delta} - L_{\tau,\Delta}) / (H_{\tau,\Delta} + L_{\tau,\Delta})$$

即 Δ 时间间隔的市场宽度定义，其中 H、L 分别为达到 换手率的最高价与最低价， τ,Δ 指在 Δ 时间间隔内首次达到 换手率的时间。杨朝军等人的研究报告中用到的换手率为 0.01% 或 0.005%。

2) 市场深度 (Depth)：即在不影响当前价格下的成交量。从市场深度来看，流动性好意味着在当前价格下投资者可以大量买入或抛出股票。市场深度的观测值是某一时刻做市商在委托簿(Order Book)报出的委托数量。杨朝军等人的研究报告中对此做如下界定：

Δ 时间间隔的深度 $D_{c,\Delta}$ ，是指自 t 时始到股票价格首次超越区间 $[P_t - c, P_t + c]$ 时的换手率， 为首次超越的时间，c 为价格波动幅度。杨朝军等人的研究报告中采用容错法，得到合适的波动临界值为价格 P 的 0.12%。

3) 市场弹性 (Resiliency): 即由交易引起的价格波动消失的速度, 弹性通过时间来表示, 它表示价格偏离均价后首次返回所需要的时间。杨朝军等人的研究报告中对此做如下界定:

Δ 时间间隔的弹性 $R_{c,\Delta}$ 是指自 t 起在 Δ 时间间隔内, 股票价格偏离 P_t 的幅度为 c 并首次返回 P_t 的时间。

(2) 综合评价

综合评价是通过描述被评价事物的多个指标来进行的, 如果将指标看成变量, 则在几何上将形成一个高维空间, 而每个被评价的事物由反映它的多个指标值在该空间中决定一个点。因此, 从几何角度来看, 综合评价的对象就是高维空间中的一些点, 综合评价问题就变为对这些点作出总体评价或排序。一个直观而自然的想法就是在空间中确定出参考点, 比如最优样本点、最劣样本点等, 然后计算各样本点到参考点的距离, 距离最优样本点越近越好, 距离最劣样本点越远越好, 这就是距离综合评价方法的基本思想。本课题中距离的计算选用欧氏距离。

具体评价步骤如下:

- 1) 指标同向化。将逆向指标正向化, 一般取倒数, 若数据过小, 则可同乘 100;
- 2) 无量纲化。可以进行标准化处理, 可用比重法, 即: 用 X 中每一元素除以其所在列所有元素平方和的算术平方根。得到矩阵 Y^T
- 3) 构造加权数据矩阵。设以确定各指标的权重为 w_1, w_2, \dots, w_p , 以它们为主对角线元素构造对角矩阵 W , 则加权数据矩阵为 $Y^T W = Y$ 。
- 4) 确定参考样本。一般以最优样本和最劣样本为参考样本。
- 5) 计算距离, 并按照距离大小进行排序。

2. 调整流通市值指标

调整流通市值是按照 180 指数编制原理中规定的分级靠档方法对成分股的流通市值进行调整的。180 指数的分级靠档比例如下表 2.1 所示:

表 2.1 分级靠档比例

流通比例 (%)	≤ 10	(10, 20]	(20, 30]	(30, 40]	(40, 50]	(50, 60]	(60, 70]	(70, 80]	> 80
加权比例 (%)	流通比例	20	30	40	50	60	70	80	100

计算时先计算调整流通股本, 再乘以相应的价格, 得出调整流通市值, 并按其进行排序。

3. 综合评价

把以上两个指标的排名结果加和，再根据加和的值进行综合排名，即得个股的流动性综合评价价值。

4. 抽样

根据各成分股的行业属性计算 180 成分股数目在各行业内的分配比例，再根据这个比例计算将要剔除的股票数目在各行业内的大致分配。之后在综合评价结果的基础上，综合考虑行业的分散性手动挑出应当剔除的股票。

(三) 模型构建

此中的优化模型与优化复制中的模型在思路是一致的，所不同的是由于抽样后的 150 只股票要求其最终持有量最少为 5000 股（即 50 手，这是一个经验值，可以根据实际情况加以调整。取 50 手的目的是保证每一只股票最少的代表性，不至于因为计算出来的手数少得可以忽略而丧失了最低代表性），所以必须在原模型（1）中加入最少持有量的约束：

$$q_i \geq 5000$$

即原模型（1）变为：

$$\begin{aligned} \text{Min}_{q_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \frac{\sum_{i=1}^N \bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i} R_{it} \right)^2} \\ \text{s.t.} \\ 1) \quad \alpha_i &\leq \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \leq \beta_i \quad \dots\dots\dots (3) \\ 2) \quad q_i &\geq 5000 \\ 3) \quad \sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| &= V \\ 4) \quad \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| &\leq cV \end{aligned}$$

创设问题的模型变换也与优化复制中的模型一致，每股最低不得低于 50 手的约束可以通过约定的 α_i 值来实现， α_i 值的求解过程如下：

$$q_i \geq 5000$$

$$q_i = \frac{w_i \sum_{i=1}^N \bar{p}_i q_i}{\bar{p}_i} \geq 5000$$

由于 $\tilde{q}_i = 0$ ，由原约束3) 可得 $\sum_{i=1}^N \bar{p}_i q_i = \frac{V}{1+\gamma}$ ，代入得：

$$q_i = \frac{w_i \frac{V}{1+\gamma}}{\bar{p}_i} = \frac{w_i}{\frac{\bar{p}_i(1+\gamma)}{V}} \geq 5000$$

得：

$$w_i \geq \frac{(1+\gamma)5000}{V} \bar{p}_i$$

所以：

$$\alpha_i = \frac{(1+\gamma)5000}{V} \bar{p}_i$$

综上可得，创设时的模型如下：

$$\begin{aligned} \text{Min}_{w_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{It} - \sum_{i=1}^N w_i R_{it} \right)^2} \\ \text{s.t.} & \\ 1) & \frac{(1+\gamma)5000}{V} \bar{p}_i \leq w_i \leq \beta_i \quad \dots\dots\dots (4) \\ 2) & \sum_{i=1}^N w_i = 1 \end{aligned}$$

由于只是选择部分成分股进行复制，因此此中的 c 取值为 0.0004，即仓位调整所带来的交易成本最多不超过初始资金规模的 0.0004，其余各参数取值以及各变量的含义与优化复制中的模型（1）和（2）一致。

2.1.3 完全复制问题

（一）基本思想

正如上文所述，完全复制就是购买标的指数中的所有成份证券，并且按照每种成份证券在标的指数中的权重确定购买的比例以构建追踪组合从而达到复制指数的目的。尽管实际情况要复杂的多，原因是指数是一个“纸面上的组合”（paper portfolio），每种成份证券在标的指数中的权重时时刻刻在发生变化，以某一时刻的相对权重值来确定组合的结构显然不能保证组合的走势与指数完全一致，因此实务中即便是完全复制也要根据追踪误差的偏离状况对组合进行动态调整。本课题中由于要进行比较，所以对完全复制

的情况也要一并计算，即在一定的资金规模（10 亿元）下配置资产，并考察其相对于标的指数的表现。

（二）变量求解

由于受资金规模的限制，完全复制问题实际上就是求解如下非线性方程组，得出个股的持有数量 q_i ，然后再进行资产配置，并计算一定时间段内的追踪效果。具体如下：

$$\begin{aligned} 1) \quad & \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} = S_w \\ 2) \quad & \sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| = V \end{aligned}$$

创设时，由于 $\tilde{q}_i = 0$ ，原方程组 2) 变为 $\sum_{i=1}^N \bar{p}_i q_i = \frac{V}{1+\gamma}$ ；

创设时，交易成本是规模的一定比例，因此 $\text{cost} = \gamma \sum_{i=1}^N \bar{p}_i q_i = \frac{\gamma V}{1+\gamma}$ ；

模型中各变量的含义及参数取值与优化复制中的模型（1）一致。求解结果经取整后得到的个股持有手数见附录中的附表 4。

2.1.4 样本外绩效考察

优化复制或抽样复制中优化计算的结果是得到最优的持有数量（创设问题时得到的是最优的持有权重），而完全复制中通过求解非线性方程组来得到个股的持有数量，不过无论是哪种方法，都要根据计算的结果再进行资产配置，并进行样本外（out of sample）绩效考察，样本外绩效如何是指数组合优化成功与否的关键。具体讲，样本外绩效考察就是考察该追踪组合在未来若干个交易日相对于标的指数的表现，即计算日均追踪误差

和累积追踪误差，其计算公式分别为 $TE = \frac{1}{T} \sqrt{\sum_{t=1}^T (R_{it} - R_{pt})^2}$ 和 $TTE = \sqrt{\sum_{t=1}^T (R_{it} - R_{pt})^2}$ ；另外

还要考察给定时间段内追踪组合收益率与指数收益率之间的相关系数、追踪组合总的调整成本组合分散化程度等指标。而对样本内的追踪效果则主要计算三个指标，即 TE、TTE 和追踪组合收益率与指数收益率之间的相关系数。

2.2 流动性惩罚约束模型

正如前述，流动性惩罚约束模型是对一般模型的一种改进，加入流动性惩罚约束的

目的是试图对个股的流动性在优化计算中区分对待以贴近实际，目的是在没有恶化追踪误差的前提下使得模型计算的结果能让更多的资金配置于比较有流动性的资产上。本模型仅包括优化复制和抽样复制两种情况，完全复制与此无关。此外，样本外绩效考察同上文所述，在此不再重复。

2.2.1 基于市场冲击成本考虑的流动性惩罚约束

个股流动性的不同直接影响到仓位调整时市场冲击成本的大小。一般讲，个股流动性大小与其仓位调整所导致的市场冲击成本的大小成反向变动关系，市场流动性越高的个股，市场冲击成本也越小，反之相反。可见市场冲击成本可以看做是对个股流动性的一个“惩罚”，流动性越高，这种惩罚就越小，反之也相反。基于这种认识，我们在上述模型（1）至（4）中加入这个惩罚，同时通过限定这个惩罚的最大值就可以实现提高指数组合总体流动性的目的。这个惩罚如下：

$$\sum_{i=1}^N \Delta_i \bar{p}_i |q_i - \tilde{q}_i| \leq dV$$

Δ_i 就是个股的流动性惩罚因子， $\Delta_i \bar{p}_i |q_i - \tilde{q}_i|$ 是个股 i 仓位调整所引起的惩罚值，总的惩罚值不能超过初始资金规模的一定比例，此中为 d 。 Δ_i 的取值则由下式给定：

$$\Delta_i = \frac{1 \times (\Delta A / A) \times e}{L_{180} - L_1} \times (L_i - L_1)$$

其中 $(\Delta A / A)$ 为市场成交金额的变化比例， e 为市场冲击成本关于成交金额的平均弹性，因此 $(\Delta A / A) \times e$ 为市场冲击成本相对其自身的平均变化比例。把市场冲击成本占成交金额变化额的比例记为 1（或为任意正小数，比如 η ，但因为 d 的取值与 η 有关，在这个约束的不等式两边可以同时约去 η ，因此这里可以记为 1），则 $1 \times (\Delta A / A) \times e$ 为市场冲击成本占成交金额变化额的平均比例，这个平均比例再根据个股不同的流动性综合评价进行线性分摊就得到个股不同的市场冲击成本占其成交金额变化额的比例，这就是 Δ_i 的经济含义。这里隐含假设认为标的指数的成分股中最不具流动性的个股的市场冲击成本占其成交金额变化额的比例就是整个股票市场的平均比例 $1 \times (\Delta A / A) \times e$ ，而且这个比例随个股流动性变化而线性变化。剩下的就是取值问题，根据申银万国公司研究人员（2004）对 2003 年 9 月至 12 月沪深股票市场流动性特征的统计测算认为，市场冲击成

本关于成交金额的平均弹性为 1.222 ,因此本研究中 e 的值据此为 1.222 ,而 d 的取值为 Δ_i 平均值扣除不等式两边资金比例影响后的 0.9 ,即把组合总体流动性提高 10%。当然 d 也可以是 $[0,1]$ 之间的其它数 ,数值越小 ,代表组合总体流动性提高的比例越多 ,但这样会大幅度放大组合的追踪误差 ,因此需要在流动性提高比例与追踪误差的恶化之间取得一个平衡。

可见 $\Delta_i \bar{p}_i |q_i - \tilde{q}_i|$ 实际上就是个股 i 仓位变动所带来的市场冲击成本的一种估计 ,因此说这个约束的提出是基于市场冲击成本考虑的流动性惩罚约束。当然这只是一个尽可能准确的估计 ,不过实际的市场冲击成本值是无法准确衡量的 ,因为影响市场冲击成本的因素如同影响股价波动的因素一样多而无法精确计算 ,因此对市场冲击成本也只能是一种估计。

实际上也可以不基于市场冲击成本考虑来对 Δ_i 和 d 进行取值 , Δ_i 只是代表一个流动性惩罚因子 ,具体惩罚多少可以主观而定 ,只要确保 d 的取值为 Δ_i 平均值扣除不等式两边资金比例影响后 $[0,1]$ 之间的的一定比例 ,加入这个约束仍然能够使得模型优化计算的结果是能让更多的资金配置在比较有流动性的资产上。

2.2.2 改进后的模型

(一) 优化复制问题

此中的优化模型是在优化复制的一般模型中加入流动性惩罚约束 ,即在模型 (1) 中加入第四个约束 ,其余不变 ,即得如下模型 (5):

$$\begin{aligned} \min_{q_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N \left[\frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} R_{it} \right] \right)^2} \\ s.t. & \\ 1) & \alpha_i \leq \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \leq \beta_i \dots\dots\dots (5) \\ 2) & \sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| = V \\ 3) & \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| \leq cV \\ 4) & \sum_{i=1}^N \Delta_i \bar{p}_i |q_i - \tilde{q}_i| \leq dV, \Delta_i = \frac{1 \times (\Delta A / A) \times e}{L_{180} - L_1} (L_i - L_1) \end{aligned}$$

同样的，创设问题的模型变换如下：

由于 $\tilde{q}_i = 0$ ，上模型(5)中的约束4)变为 $\sum_{i=1}^N \Delta_i \bar{p}_i q_i \leq dV$ ，不等式两边同除以 $\sum_{i=1}^N \bar{p}_i q_i$

得：

$$\text{左边} = \frac{\sum_{i=1}^N \Delta_i \bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} = \frac{\Delta_1 \bar{p}_1 q_1 + \Delta_2 \bar{p}_2 q_2 + \cdots + \Delta_N \bar{p}_N q_N}{\sum_{i=1}^N \bar{p}_i q_i} = \Delta_1 w_1 + \Delta_2 w_2 + \cdots + \Delta_N w_N = \sum_{i=1}^N \Delta_i w_i$$

$$\text{右边} = dV / \sum_{i=1}^N \bar{p}_i q_i = dV / (\frac{V}{1+\gamma}) = d(1+\gamma)$$

即上模型(5)中的约束4)变为：

$$\sum_{i=1}^N \Delta_i w_i \leq d(1+\gamma)$$

综上则得创设问题中的模型为：

$$\text{Min}_{w_i} TE = \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N w_i R_{it} \right)^2}$$

s.t.

$$1) \alpha_i \leq w_i \leq \beta_i \quad \dots\dots\dots (6)$$

$$2) \sum_{i=1}^N w_i = 1$$

$$3) \sum_{i=1}^N \Delta_i w_i \leq d(1+\gamma), \quad \Delta_i = \frac{1 \times (\Delta A / A) \times e}{L_{180} - L_1} (L_i - L_1)$$

(二) 抽样复制问题

同样的抽样复制中的模型可以做如下变换，即在模型(3)中加入流动性惩罚约束，得：

$$\begin{aligned}
 Min_{q_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \frac{\sum_{i=1}^N \bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} R_{it} \right)^2} \\
 s.t. & \\
 1) & \alpha_i \leq \frac{\bar{p}_i q_i}{\sum_{i=1}^N \bar{p}_i q_i} \leq \beta_i \\
 2) & q_i \geq 5000 \dots\dots\dots (7) \\
 3) & \sum_{i=1}^N \bar{p}_i q_i + \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| = V \\
 4) & \gamma \sum_{i=1}^N \bar{p}_i |q_i - \tilde{q}_i| \leq cV \\
 5) & \sum_{i=1}^N \Delta_i \bar{p}_i |q_i - \tilde{q}_i| \leq dV, \Delta_i = \frac{1 \times (\Delta A / A) \times e}{L_{180} - L_1} (L_i - L_1)
 \end{aligned}$$

创设问题中的模型变换为：

$$\begin{aligned}
 Min_{w_i} TE &= \frac{1}{T} \sqrt{\sum_{t=1}^T \left(R_{it} - \sum_{i=1}^N w_i R_{it} \right)^2} \\
 s.t. & \\
 1) & \frac{(1+\gamma)5000}{V} \bar{p}_i \leq w_i \leq \beta_i \dots\dots\dots (8) \\
 2) & \sum_{i=1}^N w_i = 1 \\
 3) & \sum_{i=1}^N \Delta_i w_i \leq d(1+\gamma), \Delta_i = \frac{1 \times (\Delta A / A) \times e}{L_{180} - L_1} (L_i - L_1)
 \end{aligned}$$

3. 数据及计算

3.1 数据来源

本研究中用到的数据有两类，一是日交易数据，并进行复权调整，用于优化计算；二是高频数据，即 1 分钟成交数据，用于评价个股的流动性。这两类数据均由上海证券交易所信息中心提供。此外，分层抽样中用到的个股的行业归属以及所有上市公司的行业分类则来自于上海证券交易所网站（<http://www.sse.com.cn>）。

由于上证 180 指数对成分股的现金分红不做调整，即让价格自然回落，对指数不做修正，因此本课题在对日交易数据进行复权处理时也对个股的现金分红不做调整，即计算未考虑现金分红的复权日交易数据。在此基础上，再计算指数和个股的日收益率数据。

3.2 参数设定

此中的参数设定是在谢文良、李进生、谢素娟等人（2002）研究结论的基础上结合上证 180 指数的编制方法以及本课题的研究设计加以综合考虑的，具体如下：

1. 样本内（in-sample）时间窗口：原则上是基于前 6 个月的历史数据进行优化计算，如果发生某些成分股的上市日期晚于模型构建之前 6 个月这个时点的，则以该成分股上市日为时间起点。如果上市日期晚于模型构建之前 6 个月这个时点的成分股数目不止一只的，则以最迟上市的日期为模型构建的时间起点。具体见下表 3.1，谢文良、李进生、谢素娟等人（2002）的研究结论认为，在保证市场结构状况一致的情况下用尽可能长的时间窗来进行优化计算。由于上证 180 指数每半年调整一次，因此本课题中最合适的时间窗应当是 6 个月。

2. 样本外（out-of-sample）时间窗口：本研究中同时考虑模型之外 5 个交易日、10 个交易日、30 个交易日、60 个交易日以及截至下次指数调整日的指数组合追踪效果。无论是短期的指数套利决策还是长期的指数化投资决策，都可以参考。由于数据截至日为 2004 年 3 月 15 日，因此在指数第三次调整之后的追踪组合中由于数据长度限制只考虑了模型之外 5 个交易日、10 个交易日、30 个交易日以及截至下次指数调整日的指数组合追踪效果，具体见下表 3.1。

3. 样本股数目：完全复制中以上证 180 指数的所有 180 只成分股的成交数据进行计算；优化复制中也以上证 180 指数的所有 180 只成分股为样本空间，但对最终的指数组合的抽样数目不加限制；抽样复制中则事先从 180 只成分股中抽样 150 只，再进行优化计算，对最终的指数组合的抽样数目严格限定在 150 只。

表 3.1 模型时间窗

决策时点 (decision point)	样本内 (in-sample)	样本外 (out-of-sample)	
		总区间	考察区间
创设问题 (2002/07/01)	2002/01/11-2002/06/30	2002/07/01-2002/12/31	5、10、30、60 日以及 截至下次指数调整日
第一次调整 (2003/01/02)	2002/07/01-2002/12/31	2003/01/02-2003/07/01	5、10、30、60 日以及 截至下次指数调整日
第二次调整 (2003/07/01)	2003/01/07-2003/06/30	2003/07/01-2003/12/31	5、10、30、60 日以及 截至下次指数调整日
第三次调整 (2004/01/02)	2003/11/18-2003/12/31	2004/01/02-2004/03/15	5、10、30 日以及截至 下次指数调整日

3 . 3 数据处理

优化复制和抽样复制中的优化计算是基于 Matlab6.5 优化工具箱 (Optimization Toolbox) 的编程技术进行的 ; 完全复制中则基于 Matlab6.5 编程来求解非线性方程组。优化计算中用到的算法的基本思想是用于解决非线性约束下的非线性优化问题的序贯二次规划法 (Sequential Quadratic Programming , SQP) , 可以选择的参数包括标准算法与大规模算法 (Largescale on/off) Levenberg-Marquardt 搜索法与 Gauss-Newton 搜索法 (Levenberg-Marquardt on/off) 等 ; SQP 的思想以及相关参数的含义参见 Matlab6.5 的帮助文件和有关优化计算的文献资料。限于篇幅 , 数据处理过程以及优化计算编程在此也不赘述。

而个股流动性衡量的数据的处理是在 SAS8.2 环境下进行的。具体讲是利用每次调整 (包括首次发布) 时对应的 180 指数成分股的前 6 个月的每分钟成交数据 , 分别计算该 180 只股票的三个流动性指标值 , 然后按距离评价法进行综合评价 , 并给出排名。宽度指标越小 , 流动性越强 ; 深度指标越大 , 流动性越强 ; 弹性指标越小 , 流动性越强。故将深度指标逆向化 , 深度的倒数越小 , 流动性越强。三个指标所占权重分别取为 0.331、0.529、0.14 (基于杨朝军等人的研究结果) 。取出 Y 中各向量的最小值 , 作为一个最优的样本 , 然后计算各个样本的指标到最优样本的距离。距离越小 , 则表示流动性越强 , 距离越大 , 表示流动性越差。个股流动性综合评价的计算结果限于篇幅略。

流动性综合评价之后再根据行业分布进行抽样的数据处理也是在 SAS8.2 环境下进行的。抽样的结果见附录中的附表 1。

3 . 4 计算结果

以下计算结果包括了两类模型 (即一般模型和流动性惩罚约束模型) 、三种复制方法 (即优化复制、抽样复制和完全复制) 以及四次组合变动 (即创设问题、上证 180 指数第一至第三次调整后的指数组合的相应调整) 的情况。优化复制和抽样复制中都用到

了优化算法，完全复制中则是求解非线性方程组。此外，抽样复制中还要根据流动性指标和调整流通市值指标进行抽样，流动性惩罚约束模型下还要决定流动性惩罚因子的取值，所有这些结果按模型分类如下。

3.4.1 一般模型下的计算结果

3.4.1.1 优化计算的迭代信息

优化计算时需要根据上下界和初始值的不同进行调试，选择最优的计算结果。限于篇幅，此中的迭代情况仅以优化复制中的创设问题（2002/07/01）的有关信息加以简要说明，其余略。方程在经过 20 次迭代后优化成功终止，迭代过程和收敛程度的信息见下表 3.2：

表 3.2 优化求解的相关信息

Diagnostic Information	
Number of variables: 180	
Functions	
Objective:	objfun
Gradient:	finite-differencing
Hessian:	finite-differencing (or Quasi-Newton)
Constraints	
Nonlinear constraints:	do not exist
Number of linear inequality constraints:	0
Number of linear equality constraints:	1
Number of lower bound constraints:	180
Number of upper bound constraints:	180
Algorithm selected	
medium-scale	
Optimization terminated successfully:	
Magnitude of directional derivative in search direction	
less than 2*options.TolFun and maximum constraint violation	
is less than options.TolCon	
Active Constraints: omitted	

目标方程总共经过了 3821 次估计，方程值从第一次迭代时的 0.000326921 一直下降到第 20 次迭代时的 $5.92795e-5$ ，迭代进程见下图 3.1。

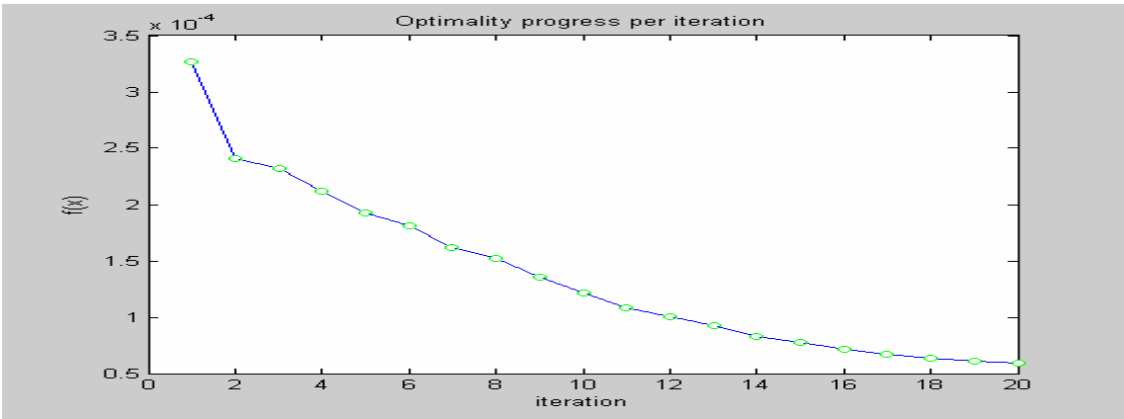


图 3.1 优化求解的迭代进程

3.4.1.2 追踪组合资产配置

假设最初的资金规模为 10 亿元，交易成本（包括佣金、印花税等的综合）为交易市值的 0.003（即模型中的 γ 值），优化复制中 c 的取值为 0.0005，抽样复制中 c 的取值为 0.0004，即仓位调整所带来的交易成本分别不超过初始资金规模的 0.0005 和 0.0004。根据优化计算出来的结果进行资产配置，实际买入的个股手数限于篇幅略，而组合市值、交易成本以及仓位调整比例见下表 3.3：

表 3.3 追踪组合资产配置

	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始 资金规模	1e+9	1e+9	1e+9	7.7154e+8	7.7344e+8	7.7204e+8
追踪组合总 市值	9.9701e+8	9.9701e+8	9.9701e+8	7.7116e+8	7.7312e+8	7.7165e+8
取整 现金差额	-2779.8	-2806	-1847.3	2785.2	3708.3	-669.24
取整 现金误差	-2.7882e-6	-2.8144e-6	-1.8529e-6	3.6117e-6	4.7965e-6	-8.6729e-7
调出仓位 比例	-	-	-	0.02223	0.02651	0.04841
调出成本	-	-	-	51461.51	6.1508e+4	112131.25
调入及仓位 变动成本	-	-	-	3.8574e+5	3.0934e+5	3.97e+5
交易总成本	2.991e+6	2.991e+6	2.991e+6	4.3721e+5	3.7085e+5	5.0913e+5
交易总成本 占初始资金 规模比例	0.003	0.003	0.003	5.6667e-4	4.7949e-4	6.5946e-4
	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始 资金规模	9.038e+8	9.0523e+8	8.7982e+8	9.296e+8	9.0263e+8	8.7791e+8
追踪组合总 市值	9.0336e+8	9.0487e+8	8.7919e+8	9.2915e+8	9.0227e+8	8.7735e+8
取整 现金差额	-12356	-5219.9	1586.3	-5264.2	-2003.0	-949.39
取整 现金误差	-1.3679e-5	-5.7687e-6	1.8043e-6	-5.6657e-6	-2.2199e-6	-1.0821e-6
调出仓位 比例	0.03852	0.06461	0.05188	0.01883	0.06929	0.04328
调出成本	104441.47	1.7549e+5	136943.93	52515.50	1.8767e+5	114007.75
调入及仓位 变动成本	4.5185e+5	3.6209e+5	6.3387e+5	4.5185e+5	3.6101e+5	5.5859e+5
交易总成本	5.5629e+5	5.5092e+5	7.7081e+5	5.1732e+5	5.4869e+5	6.726e+5
交易总成本 占初始资金 规模比例	6.155e-4	5.9386e-4	8.761e-4	5.565e-4	6.0788e-4	7.6614e-4

3.4.1.3 组合分散化

优化复制和抽样复制中的单只股票的仓位调整市值占整个组合市值比例的描述性统计值见下表 3.4，为了比较，把完全复制的情况也一并列上。可见无论是优化复制、

抽样复制还是完全复制，组合内个股的持股集中度都不高，也没有一只股票持有比例突破了模型中设定的 0.1 上限。当然对优化复制和抽样复制来讲，如有需要也可以进一步调整模型中的权重上下界 α_i 和 β_i 的取值，来达到所要求的组合内个股的持股集中度。

表 3.4 组合内个股的持股集中度

占组合市值比例	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.081492	0.082716	0.026125	0.048381	0.061740	0.028198
最小值	0.000000	0.000029	0.001849	0.000001	0.000028	0.001797
平均值	0.006005	0.006667	0.005556	0.005814	0.006667	0.005556
占组合市值比例	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.052670	0.082897	0.042220	0.058705	0.092645	0.047413
最小值	0.000001	0.000031	0.001409	0.000000	0.000027	0.000846
平均值	0.005556	0.006667	0.005556	0.005556	0.006667	0.005556

3.4.1.4 市场冲击成本

优化复制中对冲击成本的考虑主要是通过限制单只股票买入的比例，即买入数量占该股流通股本的比例，同时允许分阶段建仓，这样来分散一次性大量建仓所可能带来的市场冲击成本。而抽样复制中对冲击成本的考虑除了这方面之外，还按照流动性指标进行抽样，保证入选的个股流动性较好，这样市场的冲击成本相应地就小。

尽管没有量化研究，但从下表 3.5 可以看出单只股票的仓位调整数量占流通股本和总股本的比例无论在优化复制中还是在抽样复制中都很小，而且由于允许分阶段建仓，所以这个比例的仓位调整量所带来的市场冲击成本应该不大。

表 3.5 单只股票的仓位调整量占流通股本和总股本的比例

		创设问题 (02/07/01)		第一次调整 (03/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占流通股本比例	最大值	0.022746	0.015092	0.007987	0.010536
	最小值	0.000000	0.000015	0.000001	0.000000
	平均值	0.002368	0.002354	0.000426	0.000439
仓位调整量占总股本比例	最大值	0.012390	0.006056	0.003205	0.002634
	最小值	0.000000	0.000001	0.000000	0.000000
	平均值	0.000832	0.000807	0.000150	0.000149
		第二次调整 (03/07/01)		第三次调整 (04/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占流通股本比例	最大值	0.003810	0.003282	0.005410	0.006248
	最小值	0.000001	0.000000	0.000000	0.000000
	平均值	0.000347	0.000318	0.000447	0.000345
仓位调整量占总股本比例	最大值	0.001647	0.001044	0.001689	0.002138
	最小值	0.000000	0.000000	0.000000	0.000000
	平均值	0.000124	0.000107	0.000146	0.000104

3.4.1.5 追踪误差与相关系数

下表 3.6 给出了样本内以及样本外 5 个交易日、10 个交易日、30 个交易日、60 个交易日和截至下次指数调整日期间的组合追踪误差以及组合与指数日收益率之间的相关系数，其中 TE 为日均追踪误差，TTE 为累积追踪误差， ρ 为组合与指数日收益率之间的相关系数。

表 3.6 追踪误差和日收益率相关系数

	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内 (in sample)						
TE	5.9280e-5* (6.6128e-5)	6.5029e-5* (7.0576e-5)	-	9.0647e-5* (9.4197e-5)	1.0583e-4* (1.0990e-4)	-
TTE	0.0062* (0.0069)	0.0068* (0.0074)	-	0.0113* (0.0118)	0.0126* (0.0137)	-
ρ	0.9995	0.9994	-	0.9935	0.9916	-
样本外 (out of sample)						
5 日						
TE	7.5426e-4	7.3595e-4	1.8559e-4	5.2865e-4	4.4264e-4	7.6079e-5
TTE	0.0038	0.0037	9.2794e-4	0.0026	0.0022	3.8040e-4
ρ	0.9993	0.9992	0.9991	0.9942	0.9926	0.9999
10 日						
TE	4.7730e-4	4.8494e-4	1.1488e-4	4.1989e-4	5.0858e-4	4.9608e-5
TTE	0.0048	0.0048	0.0011	0.0042	0.0051	4.9608e-4
ρ	0.9992	0.9991	0.9995	0.9953	0.9938	0.9999
30 日						
TE	2.7275e-4	3.0127e-4	4.9594e-5	2.2431e-4	2.7750e-4	3.3110e-5
TTE	0.0082	0.0090	0.0015	0.0067	0.0083	9.9331e-4
ρ	0.9988	0.9985	0.9995	0.9951	0.9934	0.9999
60 日						
TE	1.6241e-4	1.7893e-4	3.0098e-5	1.7568e-4	1.9232e-4	2.4895e-5
TTE	0.0097	0.0107	0.0018	0.0105	0.0115	0.0015
ρ	0.9985	0.9982	0.9996	0.9942	0.9927	0.9999
截至下次指数调整日						
TE	1.0828e-4	1.1953e-4	2.1922e-5	1.6300e-4	1.7010e-4	3.3221e-5
TTE	0.0136	0.0151	0.0027	0.0187	0.0196	0.0038
ρ	0.9979	0.9975	0.9997	0.9924	0.9917	0.9996
	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内 (in sample)						
TE	1.5772e-4* (1.5615e-4)	1.8655e-4* (1.9053e-4)	-	3.1666e-4* (3.1368e-4)	3.2000e-4* (3.1765e-4)	-
TTE	0.0173*	0.0216*	-	0.0098*	0.0099*	-

指数组合优化方法、模型与应用 - 基于上证 180 指数的实证研究

	(0.0172)	(0.0205)		(0.0097)	(0.0098)	
ρ	0.9926	0.9888	-	0.9897	0.9889	-
样本外 (out of sample)						
5 日						
<i>TE</i>	7.0220e-4	6.3684e-4	8.7233e-5	8.2211e-4	0.0016	1.0582e-4
<i>TTE</i>	0.0035	0.0032	4.3616e-4	0.0041	0.0080	5.2908e-4
ρ	0.9923	0.9886	0.9999	0.9906	0.9850	0.9999
10 日						
<i>TE</i>	3.9097e-4	3.4436e-4	8.1297e-5	5.6266e-4	8.7856e-4	6.2045e-5
<i>TTE</i>	0.0039	0.0034	8.1297e-4	0.0056	0.0088	6.2045e-4
ρ	0.9924	0.9889	0.9995	0.9912	0.9871	0.9999
30 日						
<i>TE</i>	2.0833e-4	1.8573e-4	4.3242e-5	3.3576e-4	4.8845e-4	5.5972e-5
<i>TTE</i>	0.0063	0.0056	0.0013	0.0101	0.0147	0.0017
ρ	0.9923	0.9891	0.9995	0.9913	0.9851	0.9998
60 日						
<i>TE</i>	1.6439e-4	1.6989e-4	2.5759e-5	-	-	-
<i>TTE</i>	0.0099	0.0102	0.0015	-	-	-
ρ	0.9916	0.9883	0.9997	-	-	-
截至下次指数调整日						
<i>TE</i>	1.2838e-4	1.4733e-4	1.9852e-5	2.6155e-4	3.7785e-4	3.9591e-5
<i>TTE</i>	0.0163	0.0187	0.0025	0.0115	0.0166	0.0017
ρ	0.9913	0.9874	0.9998	0.9920	0.9856	0.9998

*注释：*为优化计算未取整的结果，其余数据都为根据最小交易手数取整后计算的结果。

3.4.1.6 追踪组合的表现

针对每一种复制方法下的每一次仓位变动（包括创设问题以及第一次至第三次调整），本研究都画出了指数点位与组合市值走势图、指数与组合样本内累积收益率图、样本外累积收益率图、指数与组合日收益率图以及日收益率差值图等，试图对追踪组合的追踪效果从图形上予以直观说明。不过限于篇幅，以下部分仅以抽样复制中的第一次调整（03/01/02）的情况为例，其余参见附录中的附图 1。

首先根据优化计算出来的个股持有手数和个股在样本内与样本外时间段中的复权日收盘价格计算追踪组合在整个时间段内的市值表现，并与上证 180 指数在该时间内的点位画出指数点位与组合市值走势图，见下图 3.2。

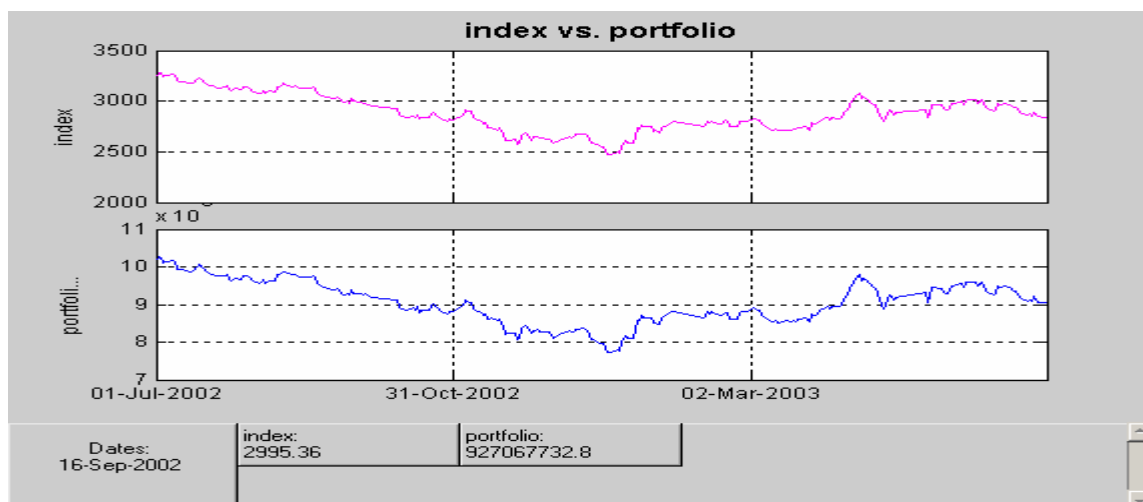


图 3.2 指数点位与组合市值走势 (2002/07/01 ~ 2003/07/01)

指数与组合日收益率图则见下图 3.3。

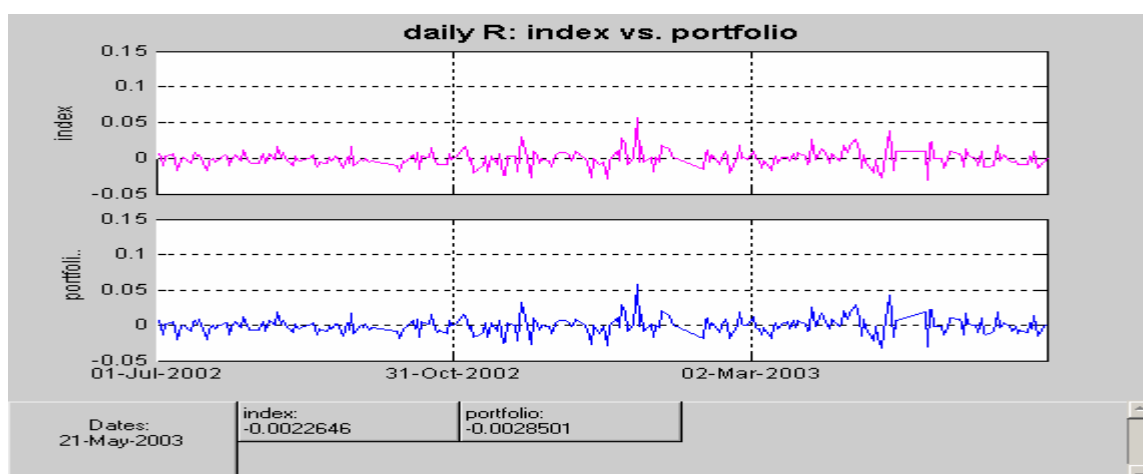


图 3.3 指数与组合日收益率 (2002/07/01 ~ 2003/07/01)

从以上两图中大致可以看出追踪组合与上证 180 指数走势的拟合情况，不过要更准确地反映出组合追踪的精度如何，则需要参考指数与组合在样本内和样本外的累积收益率图以及日收益率差值图，分别见下图 3.4 至图 3.6。

以 2002 年 7 月 1 日作为收益率累积的起点，即当天指数与组合的收益率都为零，之后计算每个交易日相对于 2002 年 7 月 1 日指数与组合各自的对数收益率，即相对 2002 年 7 月 1 日的累积收益率，这样就可以准确地衡量出指数与组合相对涨跌情况。从下图可以看出，样本内指数与组合几乎是同涨同跌，因此追踪的效果相当不错。见表 3.6，事实上，该时间段的实际日均追踪误差仅为 1.0990×10^{-4} ，指数与组合日收益率的相关系数为 0.9916，可见样本内的追踪精度的确很高。

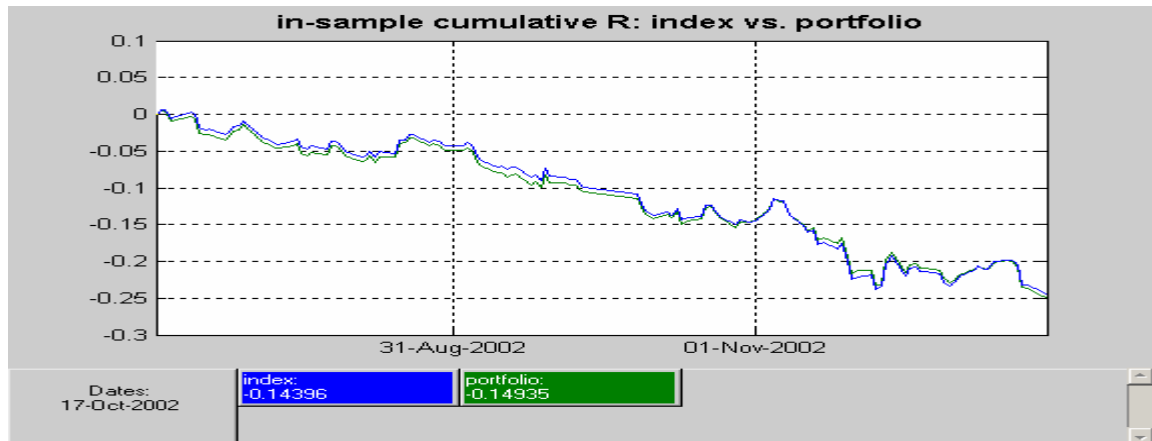


图 3.4 样本内累积收益率 (2002/07/01 ~ 2002/12/31)

样本外的累积收益率情况是衡量优化成功与否的关键。下图刻画出了指数与组合在模型之外的追踪效果，显然上半个时间段效果较好，下半个时间段略差一点，但总体情况还是不错的，见表 3.6，样本外 5 个交易日的日均追踪误差为 $4.4264e-4$ ，累积追踪误差为 0.0022，指数与组合收益率的相关系数为 0.9926，而样本外整个时间段即截至下次指数调整日的日均追踪误差为 $1.7010e-4$ ，累积追踪误差为 0.0196，指数与组合收益率的相关系数为 0.9917，这个结论与下图揭示的结果一致。

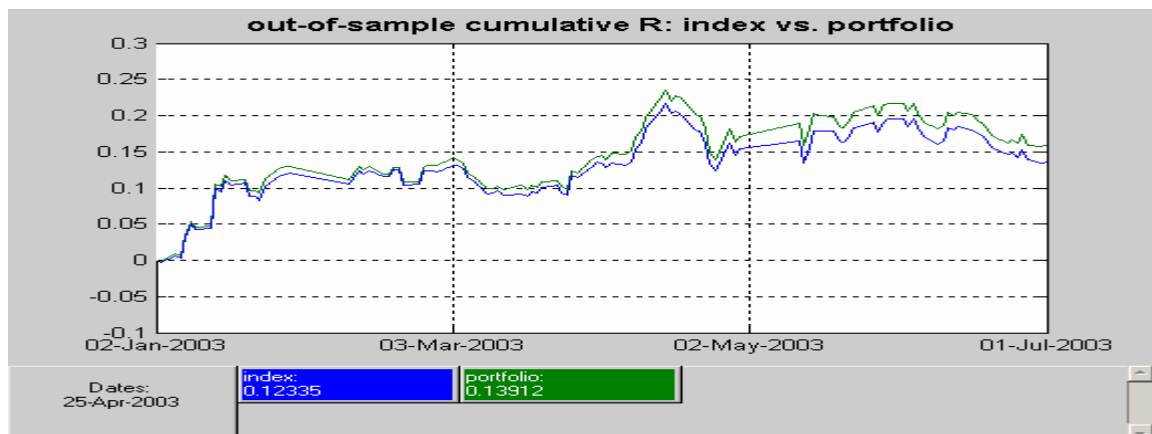


图 3.5 样本外累积收益率 (2003/01/02 ~ 2003/07/01)

下图中差值大于零，意味着组合的日收益率大于指数的日收益率，反之相反。可以看出，指数与组合的日收益率之差基本保持在正负千分之二点五之间，只有个别差值比较大，这说明组合的追踪效果不错。

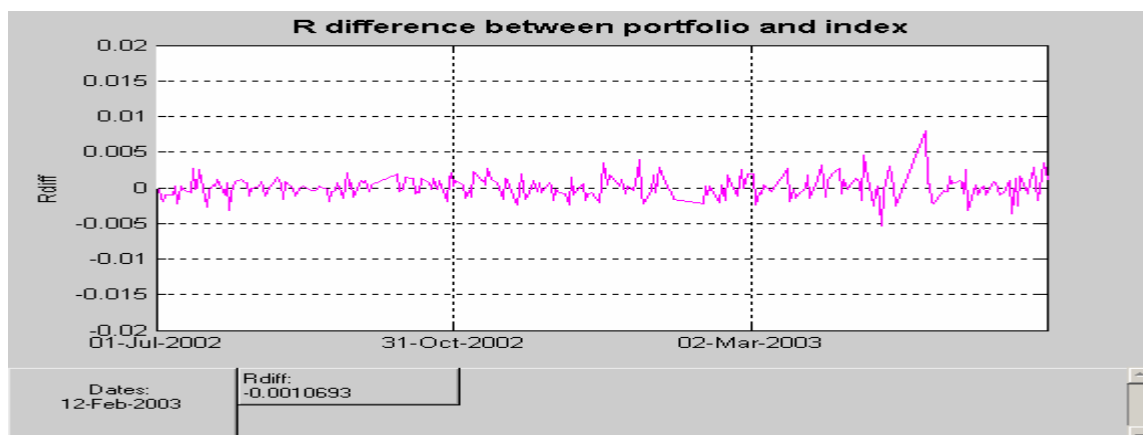


图 3.6 指数与组合的日收益率差值 (2002/07/01 ~ 2003/07/01)

3.4.1.7 追踪组合构成比较

下面给出了追踪组合的构成比较,通过这个比较可以直观看出优化计算之后资金在个股之间的配置变化情况。图 3.7 是组合调整前后的构成比较,图形左边是抽样之后所有 150 只股票的个股持有情况,右边则是取左图前 10 只股票放大比较的情况。可以看出组合调整前后个股的持有量并没有很大的变化。这是由于抽样复制的模型 (3) 中的约束 3) 即仓位调整所引起的交易成本约束作用的结果,这个交易成本最多不超过初值资金规模的 0.0004,因此组合调整时为了保证这个约束成立,优化计算的结果是把资金配置在指数调整中被新调入的成分股上,并减少其它个股的仓位变动,以降低组合仓位调整所引起的交易成本。

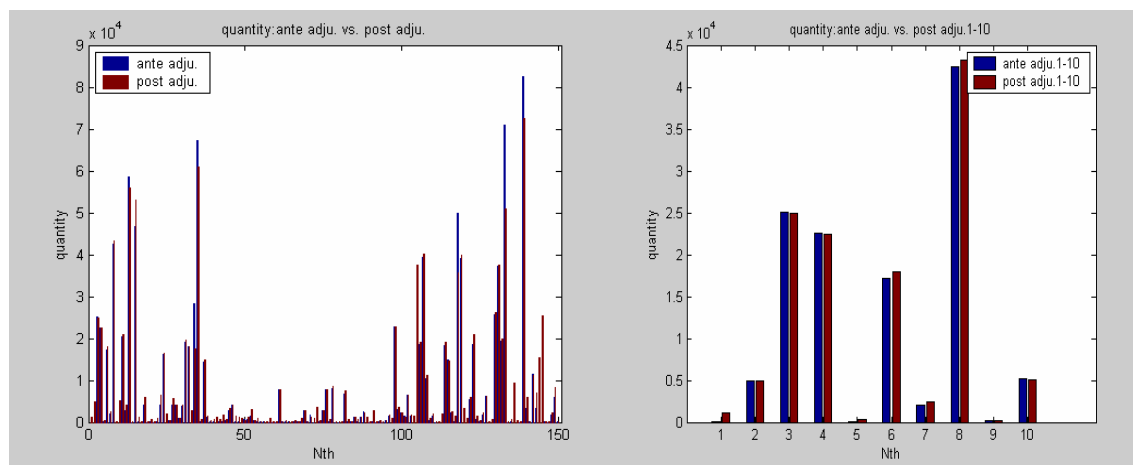


图 3.7 组合调整前后个股持有量比较

图 3.8 是上证 180 指数内对应成分股根据对应权重计算的对应股数与追踪组合内对应成分股的对应股数的比较图,图形右边同样是取左图前 10 只股票放大比较的情况。可以看出根据对应权重计算的上证 180 指数内对应成分股的对应股数与追踪组合内对应成分股的对应股数之间存在很大的区别,这是组合优化的结果,抽样 150 成分股但要使得追踪组合的表现与上证 180 指数的表现一致,这 150 只成分股的对应股数肯定有很大的差别。

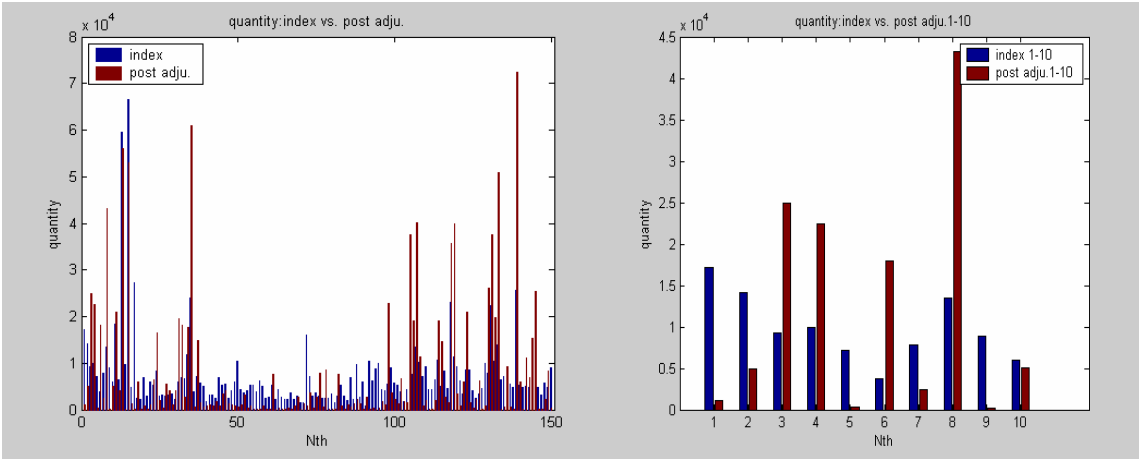


图 3.8 组合与指数对应个股的对应股数比较

3 . 4 . 2 流动性惩罚约束模型下的计算结果

3 . 4 . 2 . 1 优化计算的迭代信息

同样的优化计算时需要根据上下界和初始值的不同进行调试，选择最优的计算结果。限于篇幅，此中的迭代情况仅以抽样复制中的创设问题的有关信息加以简要说明，其余略。方程在经过 18 次迭代后优化成功终止，迭代过程和收敛程度的信息见下表 3.7，本次计算中增加了一个线性不等式约束，这个约束就是模型（8）中的约束 3）即流动性惩罚约束。

表 3.7 优化求解的相关信息

Diagnostic Information	
Number of variables: 150	
Functions	
Objective:	objfun
Gradient:	finite-differencing
Hessian:	finite-differencing (or Quasi-Newton)
Constraints	
Nonlinear constraints:	do not exist
Number of linear inequality constraints:	1
Number of linear equality constraints:	1
Number of lower bound constraints:	150
Number of upper bound constraints:	150
Algorithm selected	
medium-scale	
Optimization terminated successfully:	
Magnitude of directional derivative in search direction	
less than 2*options.TolFun and maximum constraint violation is less than options.TolCon	
is less than options.TolCon	
Active Constraints: omitted	

本次计算中，目标方程总共经过了 2887 次估计，方程值从第一次迭代时的 0.000329195 一直下降到第 18 次迭代时的 0.000116767，迭代进程见下图 3.9。

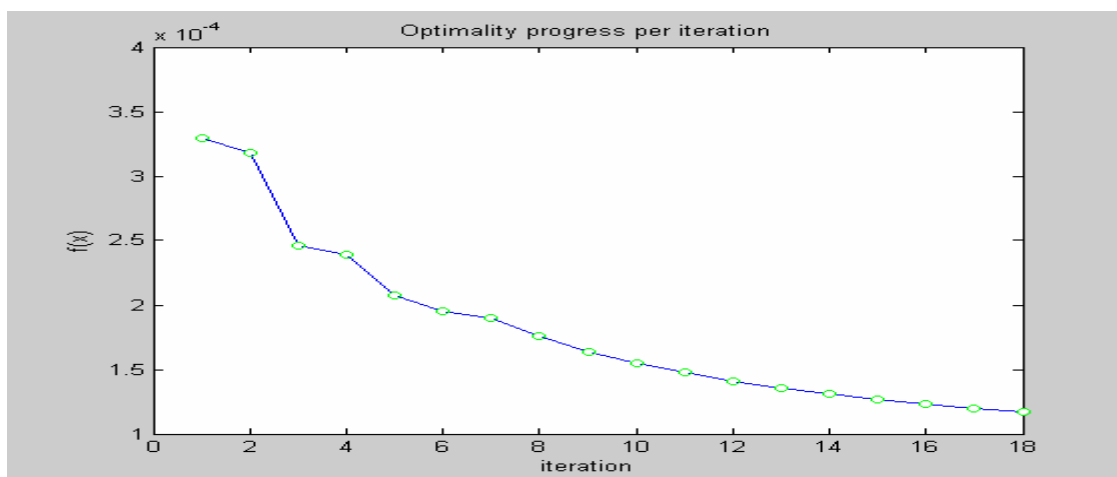


图 3.9 优化求解的迭代进程

3.4.2.2 追踪组合资产配置

同样假设最初的资金规模为 10 亿元，交易成本（包括佣金、印花税等的综合）为交易市值的 0.003（即模型中的 γ 值），优化复制中 c 的取值为 0.0005，抽样复制中 c 的取值为 0.0004，即仓位调整所带来的交易成本分别不超过初始资金规模的 0.0005 和 0.0004。流动性惩罚因子 Δ_i 以及 d 的取值见模型（5）的说明。根据优化计算出来的结果进行资产配置，实际买入的个股手数限于篇幅略，而组合市值、交易成本以及仓位调整比例见下表 3.8：

表 3.8 追踪组合资产配置

	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始 资金规模	1e+9	1e+9	1e+9	7.6917e+8	7.7077e+8	7.7204e+8
追踪组合总 市值	9.9701e+8	9.9701e+8	9.9701e+8	7.6880e+8	7.7046e+8	7.7165e+8
取整 现金差额	-3731.8	-632.43	-1847.3	-4578.6	533.65	-669.24
取整 现金误差	-3.743e-6	-6.3433e-7	-1.8529e-6	-5.9556e-6	6.9265e-7	-8.6729e-7
调出仓位 比例	-	-	-	0.00607	0.02926	0.04841
调出成本	-	-	-	14011.32	67672	112131.25
调入及仓位 变动成本	-	-	-	3.7133e+5	3.0709e+5	3.97e+5
交易总成本	2.991e+6	2.991e+6	2.991e+6	3.8534e+5	3.7476e+5	5.0913e+5
交易总成本 占初始资金 规模比例	0.003	0.003	0.003	5.0098e-4	4.8622e-4	6.5946e-4
	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始 资金规模	8.874e+8	8.955e+8	8.7982e+8	8.8871e+8	8.8507e+8	8.7791e+8
追踪组合总 市值	8.8696e+8	8.9514e+8	8.7919e+8	8.8827e+8	8.8472e+8	8.7735e+8
取整 现金差额	-1466.6	3260.9	1586.3	-571.99	-813.49	-949.39
取整 现金误差	-1.6535e-6	3.6428e-6	1.8043e-6	-6.4394e-7	-9.1949e-7	-1.0821e-6
调出仓位 比例	0.0238	0.05562	0.05188	0.02412	0.07774	0.04328
调出成本	63376.53	1.4945e+5	136943.93	64322	2.0647e+5	114007.75
调入及仓位 变动成本	4.4365e+5	3.5819e+5	6.3387e+5	4.4319e+5	3.5369e+5	5.5859e+5
交易总成本	5.0702e+5	5.0765e+5	7.7081e+5	5.0751e+5	5.6017e+5	6.726e+5
交易总成本 占初始资金 规模比例	5.7135e-4	5.6688e-4	8.761e-4	5.7106e-4	6.3290e-4	7.6614e-4

3.4.2.3 组合分散化

优化复制和抽样复制中的单只股票的仓位调整市值占整个组合市值比例的描述性统计值见下表 3.9, 可见组合内个股的持股集中度不高, 也没有一只股票持有比例突破

了模型中设定的 0.1 上限。当然如有需要也可以进一步调整模型中的权重上下界 α_i 和 β_i 的取值，来达到所要求的组合内个股的持股集中度。

表 3.9 组合内个股的持股集中度

占组合市值比例	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.066696	0.059304	0.026125	0.056124	0.056909	0.028198
最小值	0.000000	0.000022	0.001849	0.000000	0.000023	0.001797
平均值	0.005926	0.006667	0.005556	0.005865	0.006667	0.005556
占组合市值比例	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.063457	0.070368	0.042220	0.062353	0.080161	0.047413
最小值	0.000013	0.000028	0.001409	0.000000	0.000043	0.000846
平均值	0.005556	0.006667	0.005556	0.005556	0.006667	0.005556

3.4.2.4 市场冲击成本及流动性改进

在减少市场冲击成本方面，除了按照一般模型中的通过限制单只股票买入的比例，同时允许分阶段建仓来减少可能的市场冲击成本这种做法之外，此中一个重要的改进就是在模型中加入了基于市场冲击成本考虑的流动性惩罚约束，一个方面是试图对可能的市场冲击成本进行估计，更重要的是让模型优化计算的结果能把更多的资金配置在比较有流动性的资产上，提高组合的总体流动性，进而就间接地减少了可能的市场冲击成本。

下表 3.10 给出了优化复制和抽样复制中单只股票的仓位调整数量占流通股本和总股本的比例，可以看出这个比例很小，而且由于允许分阶段建仓，所以这个比例的仓位调整量所带来的市场冲击成本同样不会大。

表 3.10 单只股票的仓位调整量占流通股本和总股本的比例

		创设问题 (02/07/01)		第一次调整 (03/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占 流通股本比例	最大值	0.013519	0.015785	0.007293	0.005510
	最小值	0.000000	0.000006	0.000000	0.000000
	平均值	0.002387	0.002813	0.000472	0.000454
仓位调整量占 总股本比例	最大值	0.007243	0.007142	0.002926	0.001316
	最小值	0.000000	0.000002	0.000000	0.000000
	平均值	0.000775	0.000876	0.000144	0.000122
		第二次调整 (03/07/01)		第三次调整 (04/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占 流通股本比例	最大值	0.004720	0.002910	0.004153	0.004409
	最小值	0.000000	0.000000	0.000000	0.000000
	平均值	0.000406	0.000371	0.000302	0.000469
仓位调整量占 总股本比例	最大值	0.002678	0.001164	0.000751	0.001453
	最小值	0.000000	0.000000	0.000000	0.000000
	平均值	0.000131	0.000119	0.000085	0.000158

尽管加入流动性惩罚约束之后, 组合的追踪误差有了一定的变化 (比较表 3.6 与下表 3.11), 但组合的总体流动性明显提高了。下图 3.10 是优化复制中的第一次调整 (2003/01/02) 的情况, 图 3.11 是抽样复制中的第一次调整 (2003/01/02) 的情况, 限于篇幅, 仅以以上两例加以说明, 在其它组合变动下都有同样的结论 (两端累积百分比数值大小上会有所差异, 但结论一致), 在此不赘述。图中 case1 代表一般模型下的情况, case2 则代表流动性惩罚约束模型下的情况。先将成分股按照流动性大小排序, 再计算配置于各股上的资金占组合总市值的累积比例, 累积的方向是从流动性最好和最差的两端分别向流动性中间累积, 这就是图中的两端累积百分比。图中左边图形的左半部分红线高于蓝线 (对应于图中右边图形的左半部分两端累积百分比差值大于零), 说明排在流动性较好的一半的资金配置比例是 case2 大于 case1。左边图形的右半部分 (对应于图中右边图形的右半部分) 情况刚好相反, 所以结论也相反。这说明相对于一般模型的结果, 流动性惩罚约束模型情况下资金更多地配置于更有流动性的资产上。资金在不同流动性资产上的配置变化如右边图形中的箭头所示。可见, 流动性惩罚约束模型达到了预期的效果, 即在没有严重恶化组合追踪误差的情况下, 使得模型优化计算的结果能把更多的资金配置在比较有流动性的资产上。

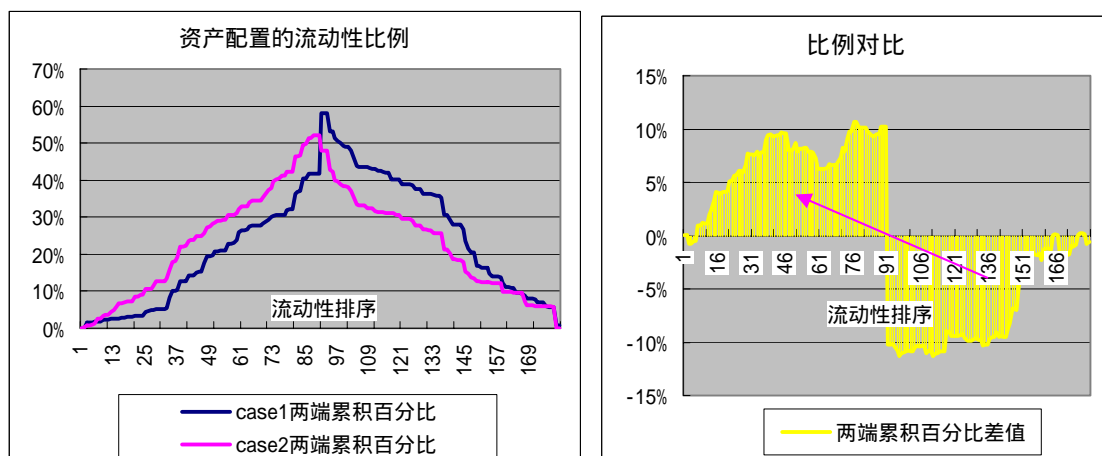


图 3.10 追踪组合总体流动性对比（优化复制中第一次调整）

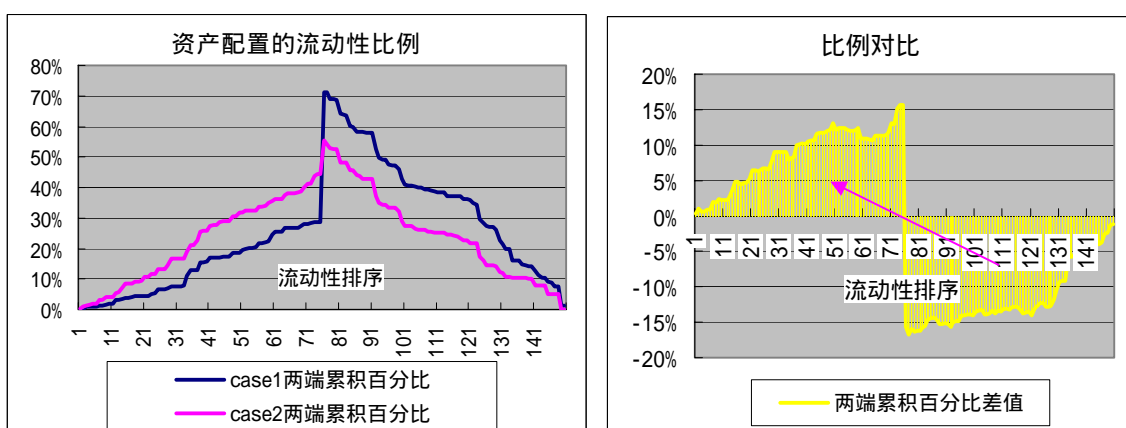


图 3.11 追踪组合总体流动性对比（抽样复制中第一次调整）

3.4.2.5 追踪误差与相关系数

下表 3.11 给出了样本内以及样本外 5 个交易日、10 个交易日、30 个交易日、60 个交易日和截至下次指数调整日期期间的组合追踪误差以及组合与指数日收益率之间的相关系数。完全复制的结果与前述一致。比较表 3.6 与下表 3.11，可以看出加入流动性惩罚约束之后，组合的追踪误差有了明显的放大（主要看创设问题中的追踪误差的比较，之后的调整都基于初始的创设组合进行的，由于根据一般模型和根据流动性惩罚约束模型求解出来的初始的创设组合肯定不同，因此根据之后的调整组合的追踪误差的比较是看不出组合追踪误差放大的迹象），但是由于基数较小，放大后的追踪误差仍然较小。而且随后基于流动性惩罚约束模型求解出来的创设组合之上的组合调整，其追踪误差有些还略小于一般模型中的结果，显然流动性惩罚约束模型改进了之后调整组合的追踪效果。可见，流动性惩罚约束模型总体上略优于一般模型，因为前者比后者提高了组合的总体流动性，又没有严重恶化追踪误差甚至还减少了追踪误差。

表 3.11 追踪误差和日收益率相关系数

	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内 (in sample)						
TE	9.4524e-5* (1.0392e-4)	1.1677e-4* (1.2584e-4)	-	8.9518e-5* (9.5020e-5)	9.8259e-5* (1.0177e-4)	-
TTE	0.0099* (0.0109)	0.0123* (0.0132)	-	0.0112* (0.0119)	0.0123* (0.0127)	-
ρ	0.9987	0.9980	-	0.9941	0.9935	-
样本外 (out of sample)						
5 日						
TE	6.8051e-4	7.8909e-4	1.8559e-4	4.9671e-4	3.6638e-4	7.6079e-5
TTE	0.0034	0.0039	9.2794e-4	0.0025	0.0018	3.8040e-4
ρ	0.9986	0.9979	0.9991	0.9947	0.9944	0.9999
10 日						
TE	4.9446e-4	5.2404e-4	1.1488e-4	4.6481e-4	4.1880e-4	4.9608e-5
TTE	0.0049	0.0052	0.0011	0.0046	0.0042	4.9608e-4
ρ	0.9984	0.9978	0.9995	0.9954	0.9951	0.9999
30 日						
TE	2.7985e-4	2.7570e-4	4.9594e-5	2.0566e-4	2.0470e-4	3.3110e-5
TTE	0.0084	0.0083	0.0015	0.0062	0.0061	9.9331e-4
ρ	0.9979	0.9974	0.9995	0.9956	0.9953	0.9999
60 日						
TE	1.7050e-4	1.6723e-4	3.0098e-5	1.7000e-4	1.6974e-4	2.4895e-5
TTE	0.0102	0.0100	0.0018	0.0102	0.0102	0.0015
ρ	0.9977	0.9971	0.9996	0.9947	0.9944	0.9999
截至下次指数调整日						
TE	1.0574e-4	1.0753e-4	2.1922e-5	1.5402e-4	1.6875e-4	3.3221e-5
TTE	0.0133	0.0135	0.0027	0.0177	0.0194	0.0038
ρ	0.9973	0.9968	0.9997	0.9931	0.9920	0.9996
	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内 (in sample)						
TE	1.2973e-4* (1.2551e-4)	1.5909e-4* (1.5459e-4)	-	2.8853e-4* (2.7065e-4)	3.2595e-4* (3.2590e-4)	-
TTE	0.0143* (0.0138)	0.0175* (0.0170)	-	0.0089* (0.0084)	0.0101* (0.0101)	-
ρ	0.9949	0.9922	-	0.9920	0.9878	-
样本外 (out of sample)						
5 日						
TE	6.5544e-4	5.6702e-4	8.7233e-5	0.0011	0.0015	1.0582e-4
TTE	0.0033	0.0028	4.3616e-4	0.0057	0.0075	5.2908e-4
ρ	0.9946	0.9920	0.9999	0.9909	0.9855	0.9999
10 日						

TE	3.9962e-4	4.1671e-4	8.1297e-5	6.6707e-4	8.9776e-4	6.2045e-5
TTE	0.0040	0.0042	8.1297e-4	0.0067	0.0090	6.2045e-4
ρ	0.9946	0.9920	0.9995	0.9916	0.9866	0.9999
30 日						
TE	2.2212e-4	2.1720e-4	4.3242e-5	3.7317e-4	4.7960e-4	5.5972e-5
TTE	0.0067	0.0065	0.0013	0.0112	0.0144	0.0017
ρ	0.9943	0.9918	0.9995	0.9909	0.9852	0.9998
60 日						
TE	1.6358e-4	1.8026e-4	2.5759e-5	-	-	-
TTE	0.0098	0.0108	0.0015	-	-	-
ρ	0.9935	0.9907	0.9997	-	-	-
截至下次指数调整日						
TE	1.2967e-4	1.4501e-4	1.9852e-5	2.8078e-4	3.7253e-4	3.9591e-5
TTE	0.0165	0.0184	0.0025	0.0124	0.0164	0.0017
ρ	0.9926	0.9897	0.9998	0.9919	0.9858	0.9998

*注释：*为优化计算未取整的结果，其余数据都为根据最小交易手数取整后计算的结果。

3.4.2.6 追踪组合的表现

同样地，针对每一种复制方法下的每一次仓位变动（包括创设问题以及第一次至第三次调整），本研究都画出了指数点位与组合市值走势图、指数与组合样本内累积收益率图、样本外累积收益率图、指数与组合日收益率图以及日收益率差值图等。限于篇幅，以下部分仅以优化复制中的第二次调整（03/07/01）的情况为例，其余参见附录中的附图 1。

根据优化计算的结果画出指数点位与组合市值走势图如下图 3.12。

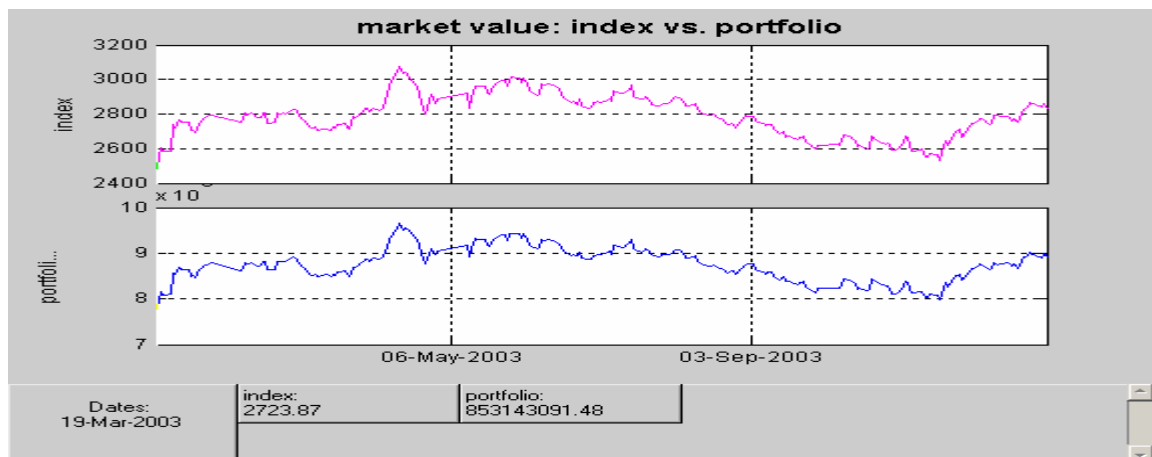


图 3.12 指数点位与组合市值走势（2003/01/07 ~ 2003/12/31）

指数与组合日收益率图则见下图 3.13。

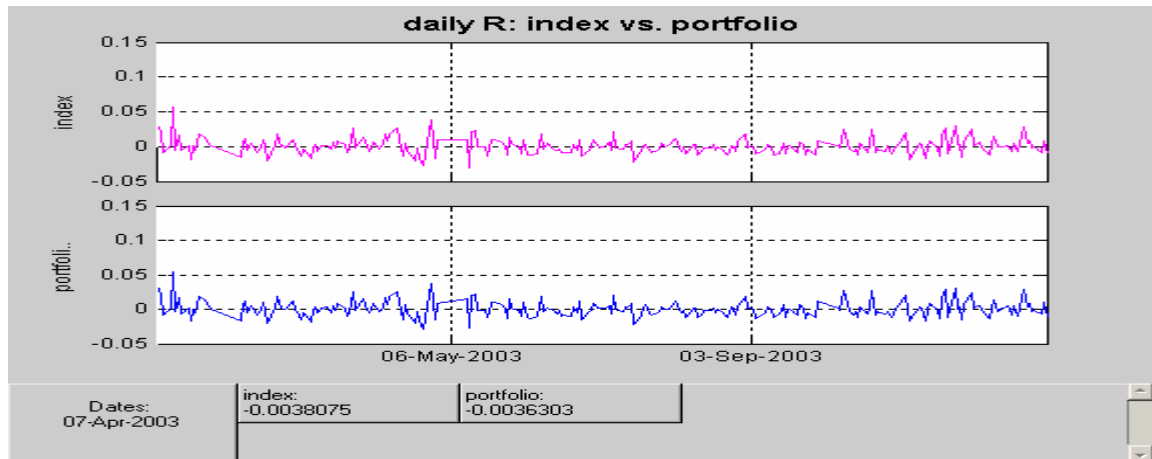


图 3.13 指数与组合日收益率 (2003/01/07 ~ 2003/12/31)

流动性惩罚约束模型下指数与组合在样本内和样本外的累积收益率图以及日收益率差值图，分别见下图 3.14 至图 3.16。

同样以 2003 年 1 月 2 日作为收益率累积的起点来计算每个交易日相对于 2003 年 1 月 2 日指数与组合各自的对数收益率，即相对 2003 年 1 月 2 日的累积收益率。从下图可以看出，样本内的追踪效果也相当不错。见表 3.11，事实上，该时间段的实际日均追踪误差仅为 $1.2551e-4$ ，累积追踪误差为 0.0138，指数与组合日收益率的相关系数为 0.9949，样本内的追踪精度比较好。

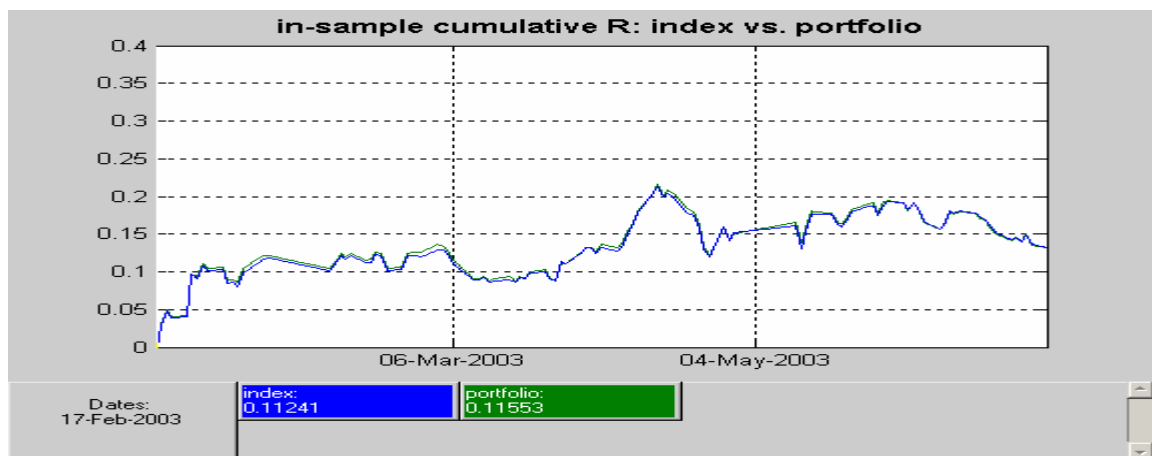


图 3.14 样本内累积收益率 (2003/01/07 ~ 200306/30)

下图刻画出了流动性惩罚约束模型下指数与组合在模型之外的追踪效果，应该说总体效果相当不错，见表 3.11，样本外 5 个交易日的日均追踪误差为 $6.5544e-4$ ，累积追踪误差为 0.0033，指数与组合收益率的相关系数为 0.9946，而样本外整个时间段即截至下次指数调整日的日均追踪误差为 $1.2967e-4$ ，累积追踪误差为 0.0165，指数与组合收益率的相关系数为 0.9926，这个结论与下图揭示的结果一致。

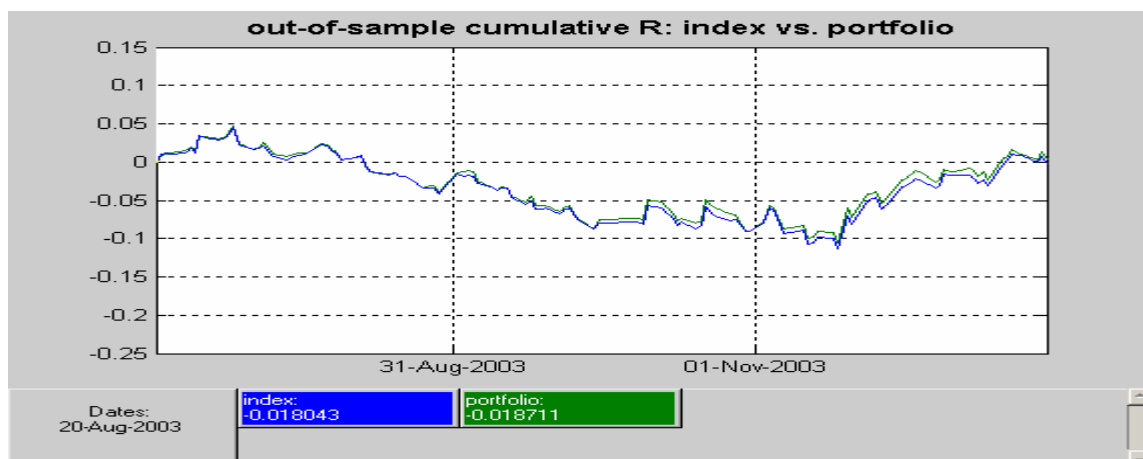


图 3.15 样本外累积收益率 (2003/07/01 ~ 2003/12/31)

从下图可以看出,流动性惩罚约束模型下指数与组合的日收益率之差也基本保持在正负千分之二点五之间,只有个别差值比较大,说明组合的追踪效果不错。

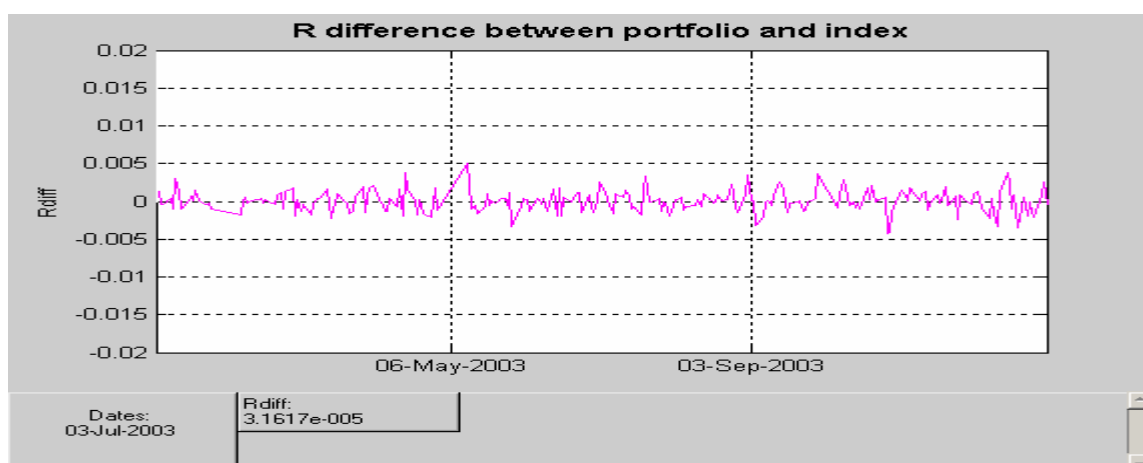


图 3.16 指数与组合的日收益率差值 (2003/01/07 ~ 2003/12/31)

3.4.2.7 追踪组合构成比较

下图 3.17 是流动性惩罚约束模型下组合调整前后的构成比较,图形左边是所有 180 只股票的个股持有情况,右边则是取左图前 10 只股票放大比较的情况。同样可以看出组合调整前后个股的持有量并没有很大的变化。这也是由于优化复制的模型 (5) 中的约束 4) 即仓位调整所引起的交易成本约束作用的结果。

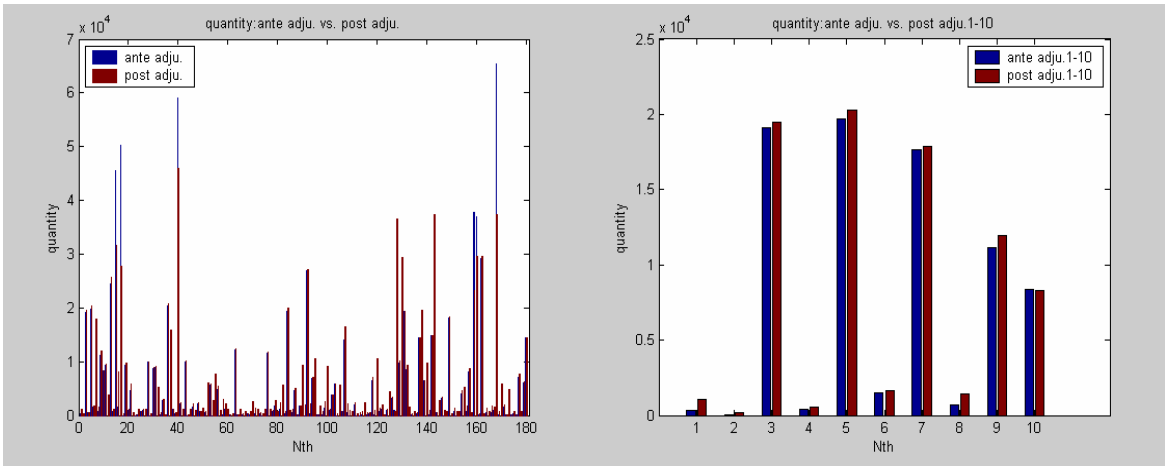


图 3.17 组合调整前后个股持有量比较

图 3.18 是流动性惩罚约束模型下上证 180 指数内对应成分股根据对应权重计算的对应股数与追踪组合内对应成分股的对应股数的比较图，图形右边同样是取左图前 10 只股票放大比较的情况。可以看出根据对应权重计算的上证 180 指数内对应成分股的对应股数与追踪组合内对应成分股的对应股数之间存在很大的区别，这是组合优化的结果。

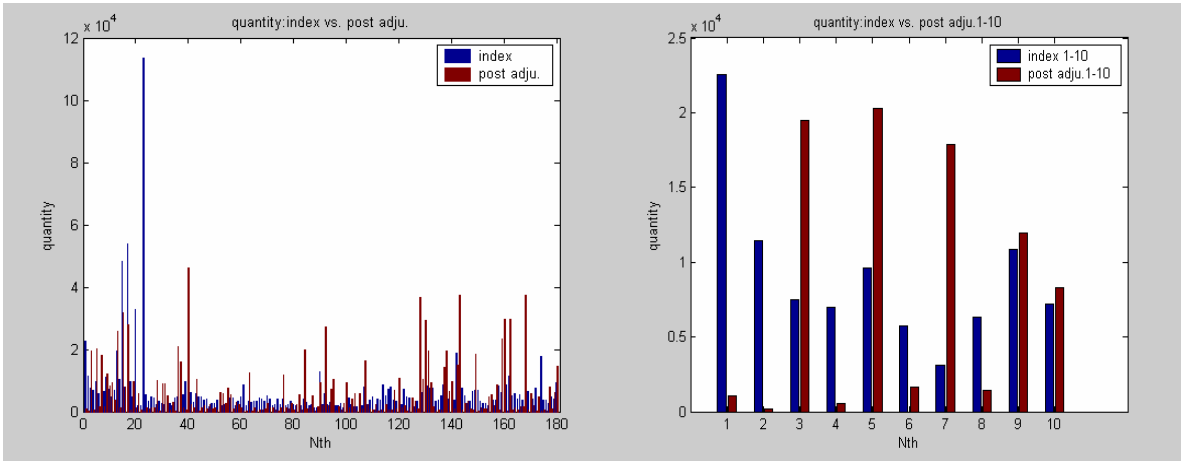


图 3.18 组合与指数对应个股的对应股数比较

4. 算法改进

4.1 遗传算法

遗传算法 (Genetic Algorithm, GA) 最早由 John Holland 于 1975 年提出来, 其模拟达尔文的遗传选择和自然淘汰的生物进化过程, 以一种群体中的所有个体为对象, 并利用随机化技术指导对一个被编码的参数空间进行高效的搜索, 因而遗传算法是一种基于生物自然选择与遗传机理的随机搜索算法。与传统搜索算法不同, 遗传算法从一组随机产生的称为“种群”(population) 的初始解开始搜索过程。种群中的每一个个体是问题的一个解, 称为“染色体”(chromosome)。染色体是一串符号, 比如一个二进制字符串。这些染色体在后续迭代中不断进化, 称为“遗传”。在每一代中用“适应度函数”(fitness function) 来测量染色体的好坏, 生成的下一代染色体称为“后代”(offspring)。后代是由前一代染色体通过“交叉”(crossover) 或“变异”(mutation) 运算形成的。在新的下一代形成过程中, 根据适应度的大小选择部分后代, 淘汰部分后代, 从而保持种群大小是个常数。适应度高的染色体被选中的概率较高。这样经过若干代进化之后, 算法收敛于最好的染色体上, 这个最好的染色体就是待求问题的最优解或次优解。可见, 遗传算法是一种具有“生存 + 检测”能力的搜索算法。这种算法的主要步骤有:

1. **编码**: GA 在进行搜索之前先将解空间的解数据表示成遗传空间的基因型串结构数据, 这些串结构数据的不同组合便构成了不同的点。
2. **初始种群的生成**: 随机产生 N 个初始结构数据, 每个串结构数据称为一个个体, N 个个体构成了一个群体。GA 以这 N 个串结构数据作为初始点开始迭代。
3. **适应度评估检测**: 适应度函数表明个体或解的优劣性。不同的问题, 适应度函数的定义方式也不同。
4. **选择**: 选择的目的是为了从当前群体中选出优良的个体, 使它们有机会作为父代为下一代繁殖子孙。选择过程体现了达尔文的适者生存的原则。
5. **交叉**: 交叉操作是遗传算法中最主要的遗传操作。通过交叉操作可以得到新一代个体, 新个体组合了其父代个体的特征。
6. **变异**: 首先在群体中随机选择一个个体, 对于选中的个体以一定的概率随机地改变串结构数据中某个串的值。同生物界一样, GA 中变异发生的概率很低, 通常取值在 $0.001 \sim 0.01$ 之间。变异为新个体的产生提供了机会。

GA 的计算流程图如下图 4.1 所示, 按照上面的步骤, 从编码和生成初始种群开始, 然后进入一个循环, 通过种群适应度评估, 选择最优的染色体进行交叉或变异以产生新

的染色体，再通过种群适应度评估，再选择最优的染色体进行交叉或变异，这样一直循环下去，直到满足优化终止的条件，找到最优解。

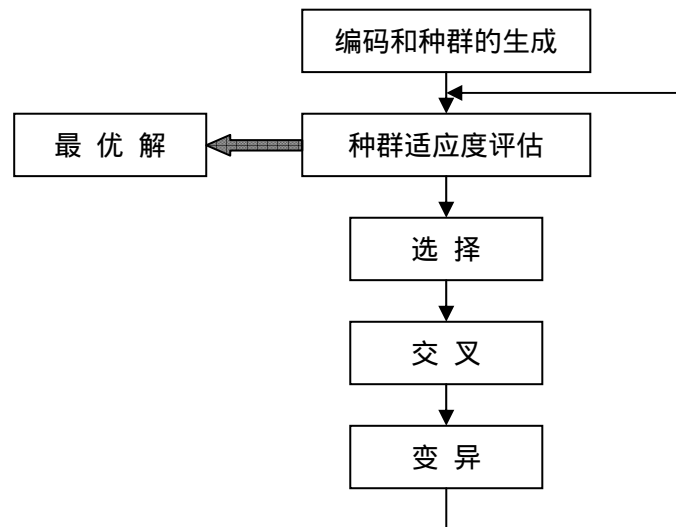


图 4.1 GA 计算流程图

4.2 基于 MATLAB 的遗传算法工具箱

应用 MATLAB 语言编程而成的遗传算法工具箱比较多，比较常用的有 Sheffield 教授于 1994 年编写的 1.2 版本的遗传算法工具箱（Genetic Algorithm Toolbox for MATLAB, v1.2），还有 C.R. Houck、J.A. Joines 以及 M.G. Kay 等人于 1998 年编写的基于 MATLAB5.x 的遗传算法工具箱（The Genetic Algorithm Optimization Toolbox for Matlab 5, GAOT）等。最新版本的 MATLAB7.0（Release 14）已经加载有遗传算法与直接搜索工具箱模块（Genetic Algorithm and Direct Search Toolbox）。

由于 MATLAB7.0 中的遗传算法与直接搜索工具箱模块比较适用于解决无约束优化问题，对非线性约束的非线性规划问题的求解效率很低。本研究在综合比较了各工具箱的优劣之后，选择了 C.R. Houck、J.A. Joines 以及 M.G. Kay 等人的遗传算法工具箱来解决本课题中的优化求解问题。

4.3 流动性惩罚约束模型的 GA 求解

这首先需要回答两个问题：一是为什么要用 GA 求解？二是为什么对流动性惩罚约束模型进行求解？算法的改进主要有两个方面的原因：一是 GA 算法通常情况下比 SQP 算法较有可能找到全局最优解；二是通过不同算法的应用，比较其在优化求解，尤其是本课题的优化求解中的具体应用和优化效果上有何不同，探讨应用 GA 来改进指数组合优化效果的可行性。对流动性约束模型进行求解的原因在于前文通过对一般模型与流动性惩罚约束模型求解结论的比较发现，流动性惩罚约束模型总体上优于一般模型，因为前者比后者提高了组合的总体流动性，又没有严重恶化追踪误差甚至还减少了追踪误

差。因而限于篇幅，一般模型的 GA 求解就不再探讨。

具体的数据处理和计算是基于 C.R. Houck、J.A. Joines 以及 M.G. Kay 等人的 GAOT 的编程技术进行的；GAOT 提供的编码方式包括浮点编码和二进制编码两种，选择、交叉和变异的方法各有几种，通过选择不同的参数就可以实现对选择、交叉和变异的不同方法的选择。GAOT 的思想以及相关参数，包括种群大小（population size）、浮点精度以及最大进化代数等的含义参见 GA 的帮助文件和有关优化计算的文献资料。限于篇幅，数据处理过程以及 GA 优化计算编程在此也不赘述。

用 GA 求解流动性惩罚约束模型的通常方法是以 GAOT 中的 initializega 函数随机生成初始种群，在此基础上进行遗传进化，直到最大进化代数超过模型设定值，终止优化计算，得到最优解。如下图 4.2 所示，这是求解优化复制创设问题中最优权重的 GA 进化过程，种群大小为 100，以 initializega 函数随机生成初始种群，浮点精度为 $1e-6$ ，最大进化代数为 500，结果为 $3.716e-4$ 。下图中，左起第一小图最优适应度（best fitness）和第二小图平均适应度（average fitness）都不断下降，说明随着进化代数的增加，GA 确实在往最小追踪误差方向上搜索。左起第三小图初始种群下 100 个个体的适应度（fitness under the initial population）之间差异很大，然而经过了 500 代进化，有 93 个个体的适应度降低到了最优适应度水平附近（见左起第四小图最终种群下 100 个个体的适应度），初始 180 只成分股的随机持有数量也从左起第五小图所代表的水平进化到了左起第六小图所代表的水平。这些都说明用这种方法进行 GA 求解以及用 GA 求解流动性惩罚约束模型是可行的。

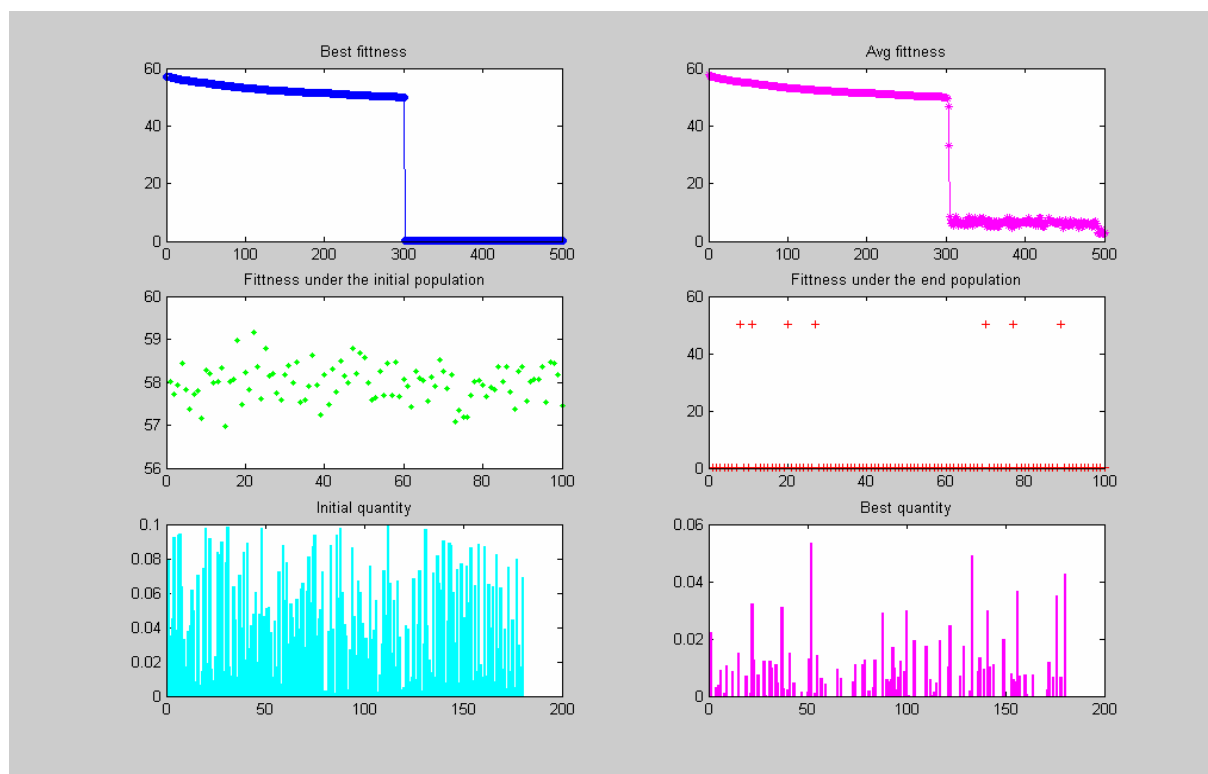


图 4.2 随机生成初始种群的 GA 进化求解过程

然而这种方法的效率很低,从中间过程的跟踪数据中可以看出,经过 301 代进化后,才找到满足原模型约束条件的解(这从上图 4.2 中的左起第一小图最优适应度和第二小图平均适应度的跳跃性下降中可以看得到),再经过近 200 代的进化才达到最终的解,即 $3.716\text{e-}4$ 。如果把最大进化代数增加到 5000,则最终进化结果为 $3.1404\text{e-}4$ 。而在流动性约束模型的 SQP 求解中,优化复制创设问题的样本内追踪误差为 $9.45238\text{e-}5$ 。可见,以 initializega 函数随机生成初始种群来进行 GA 求解需要很大的进化代数才可能达到 SQP 算法下的解,优化计算的效率很低。

很显然,一种改进的方法是以 SQP 优化计算的结果为初始种群来对流动性惩罚约束模型进行 GA 求解,即在 SQP 计算的结果上再用 GA 进行优化,试图进一步改进优化的结果(或者看看有没有进一步优化的可能)。本研究正是按照这种思路进行优化计算的。具体讲,每次在 GA 计算之前,先进行 SQP 算法下的优化求解,再把这个结果作为初始种群进行 GA 求解,把最大进化代数增加到 10000,看看最终的进化结果有没有改进以及改进的幅度如何。计算过程如下图 4.3 所示:

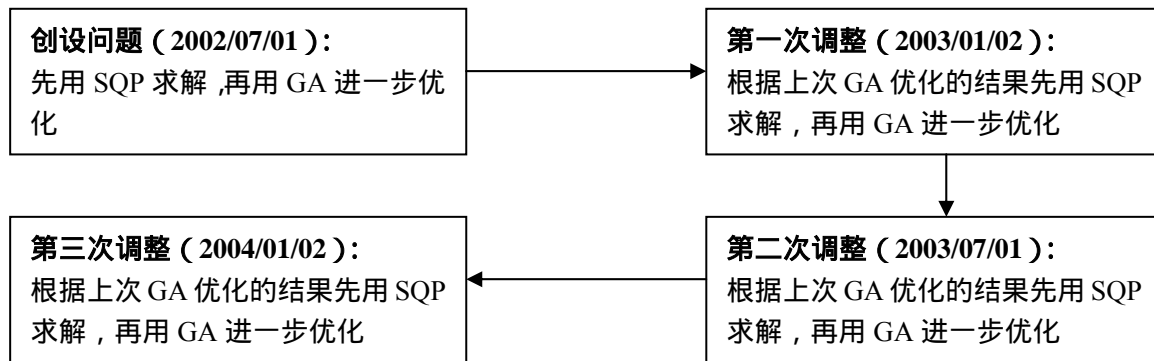


图 4.3 流动性惩罚约束模型的 GA 求解过程

4.4 计算结果

4.4.1 基于 SQP 的 GA 计算

首先在 SQP 算法下进行优化计算,然后再以 SQP 优化计算的结果构建 GA 计算的初始种群,种群大小定为 100,最大进化代数为 10000,并设定好其它参数,然后编写 GA 求解的调用程序,在 MATLAB 环境下进行优化计算。下表 4.1 列出了优化复制和抽样复制下样本内 SQP 计算的最小追踪误差和 GA 改进的情况。为了对比,把单纯用 SQP 算法求解流动性惩罚约束模型的结果也一并列出。可以看出,GA 在 SQP 计算结果的基础上都有所改进(表现在 GA 在 SQP 计算结果的基础上进一步降低了样本内追踪误差),只是改进的幅度非常小。不过把 SQP/GA(即 GA 在 SQP 计算结果的基础上的进一步优化)的计算结果与单纯用 SQP 计算的结果比,可以看出两种方法下共六个调整组合有三个调整组合样本内追踪误差小于单纯用 SQP 计算的结果,但也有另外三个调整组合样本内追踪误差大于单纯用 SQP 计算的结果。可见有 SQP/GA 计算的结果优于 SQP 的

情况，也有劣于 SQP 的情况。而且无论是优于还是劣于，两者之间的差距都不大。

表 4.1 SQP 基础上的 GA 计算结果比较

样本内 追踪误差	创设问题 (02/07/01)		第一次调整 (03/01/02)	
	优化复制	抽样复制	优化复制	抽样复制
SQP	9.4524e-5	1.1677e-4	8.9518e-5	9.8259e-5
SQP/GA	9.4524e-5/9.3964e-5	1.1677e-4/1.1632e-4	9.0200e-5/9.0161e-5	9.555e-5/9.5516e-5
样本内 追踪误差	第二次调整 (03/07/01)		第三次调整 (04/01/02)	
	优化复制	抽样复制	优化复制	抽样复制
SQP	1.2973e-4	1.5909e-4	2.8853e-4	3.2595e-4
SQP/GA	1.2272e-4/1.2268e-4	1.6425e-4/1.6421e-4	2.8083e-4/2.807e-4	3.7643e-4/3.7623e-4

优化复制中的第二次调整问题 (2003/07/01) 的 GA 求解的相关信息见下图 4.4 (限于篇幅, 仅以此为例加以说明, 其余略), 左起第一小图最优适应度和第二小图平均适应度都不断下降, 说明随着进化代数的增加, GA 在 SQP 计算结果的基础上又有所改进。不过从纵坐标看, 改进的幅度非常小。初始 180 只成分股的持有数量也从左起第五小图所代表的水平进化到了左起第六小图所代表的水平, 从图形看, 这两者之间的差异不大。对实际持有的个股手数的比较支持了这一结论, GA 只是在 SQP 的基础上进行了微调。

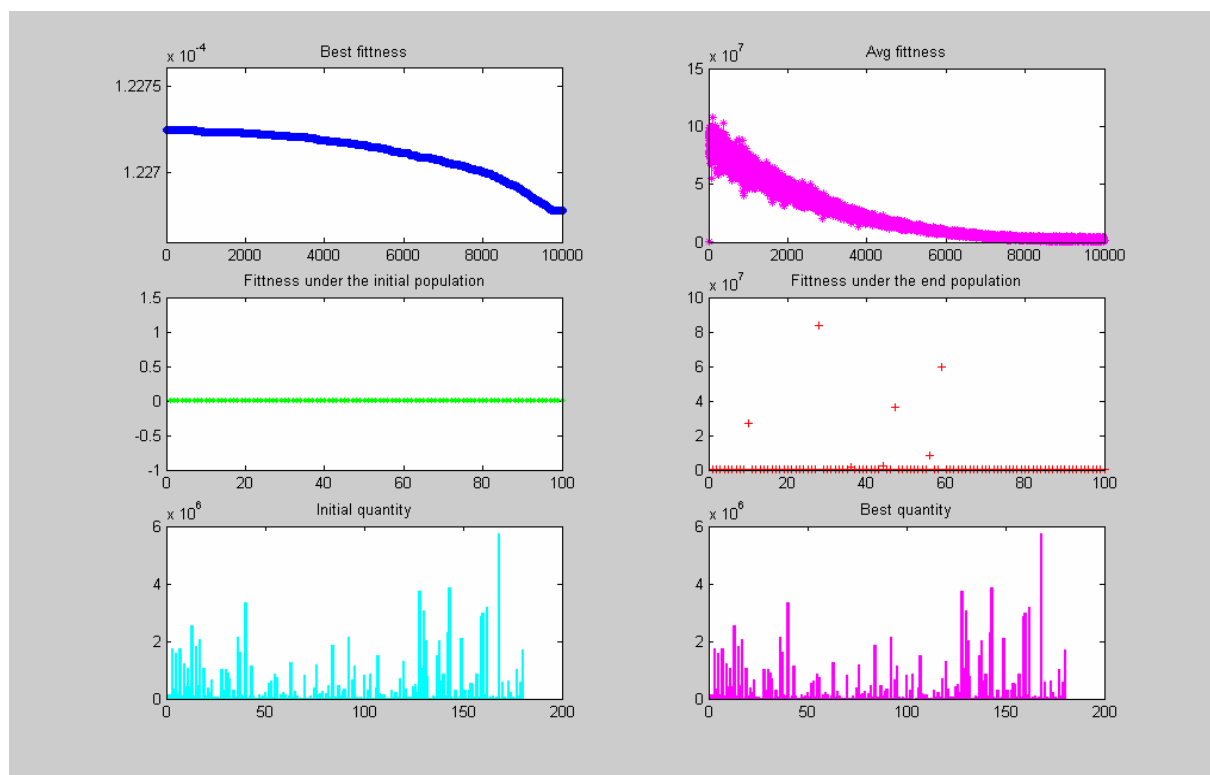


图 4.4 以 SQP 计算结果为初始种群的 GA 进化求解过程

4.4.2 追踪组合资产配置

同样假设最初的资金规模为 10 亿元, 交易成本 (包括佣金、印花税等的综合) 为交易市值的 0.003 (即模型中的 γ 值), 优化复制中 c 的取值为 0.0005, 抽样复制中 c 的

取值为 0.0004，即仓位调整所带来的交易成本分别不超过初始资金规模的 0.0005 和 0.0004。流动性惩罚因子 Δ_i 以及 d 的取值见模型 (5) 的说明。根据 GA 进一步优化的结果进行资产配置，实际买入的个股手数见附录中的附表 2 和附表 3，而组合市值、交易成本以及仓位调整比例见下表 4.2，为了比较，把完全复制的结论也一并列上。

表 4.2 追踪组合资产配置

	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始资金规模	1e+9	1e+9	1e+9	7.6923e+8	7.708e+8	7.7204e+8
追踪组合总市值	9.9701e+8	9.9701e+8	9.9701e+8	7.6885e+8	7.7049e+8	7.7165e+8
取整现金差额	-7482.4	3607.1	-1847.3	-4971.5	3913	-669.24
取整现金误差	2.495e-6	1.5401e-6	-1.8529e-6	-6.4662e-6	5.0786e-6	-8.6729e-7
调出仓位比例	-	-	-	0.00615	0.029222	0.04841
调出成本	-	-	-	14186.76	67583.86	112131.25
调入及仓位变动成本	-	-	-	3.8463e+5	3.0818e+5	3.97e+5
交易总成本	2.991e+6	2.991e+6	2.991e+6	3.9882e+5	3.7577e+5	5.0913e+5
交易总成本占初始资金规模比例	0.003	0.003	0.003	5.1847e-4	4.8750e-4	6.5946e-4
	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
初始资金规模	8.8063e+8	8.8803e+8	8.7982e+8	8.7075e+8	8.9615e+8	8.7791e+8
追踪组合总市值	8.8019e+8	8.8767e+8	8.7919e+8	8.7031e+8	8.9580e+8	8.7735e+8
取整现金差额	1093	5013.8	1586.3	5303.9	-3616.7	-949.39
取整现金误差	1.2418e-6	5.6483e-6	1.8043e-6	6.0942e-6	-4.0374e-6	-1.0821e-6
调出仓位比例	0.02667	0.04647	0.05188	0.02896	0.07526	0.04328
调出成本	70472.77	1.2382e+5	136943.93	75666.06	2.0239e+5	114007.75
调入及仓位变动成本	4.3949e+5	3.5533e+5	6.3387e+5	4.3542e+5	3.5854e+5	5.5859e+5
交易总成本	5.0997e+5	4.7915e+5	7.7081e+5	5.1109e+5	5.6093e+5	6.726e+5
交易总成本占初始资金规模比例	5.7909e-4	5.3956e-4	8.761e-4	5.8695e-4	6.2593e-4	7.6614e-4

4.4.3 组合分散化

优化复制和抽样复制中根据 GA 计算的单只股票的仓位调整市值占整个组合市值比例的描述性统计值见下表 4.3，可见组合内个股的持股集中度不高，也没有一只股票持有比例突破了模型中设定的 0.1 上限。当然如有需要也可以进一步调整模型中的权重上下界 α_i 和 β_i 的取值，来达到所要求的组合内个股的持股集中度。

表 4.3 组合内个股的持股集中度

占组合市值比例	创设问题 (02/07/01)			第一次调整 (03/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.066726	0.059375	0.026125	0.057009	0.051449	0.028198
最小值	0.000000	0.000022	0.001849	0.000002	0.000024	0.001797
平均值	0.005926	0.006667	0.005556	0.005866	0.006667	0.005556
占组合市值比例	第二次调整 (03/07/01)			第三次调整 (04/01/02)		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
最大值	0.050170	0.071751	0.042220	0.063350	0.079780	0.047413
最小值	0.000000	0.000031	0.001409	0.000000	0.000022	0.000846
平均值	0.005556	0.006667	0.005556	0.005556	0.006667	0.005556

4.4.4 市场冲击成本及流动性改进

下表 4.4 给出了优化复制和抽样复制中根据 GA 计算的单只股票的仓位调整数量占流通股本和总股本的比例，可以看出这个比例很小，而且由于允许分阶段建仓，所以这个比例的仓位调整量所带来的市场冲击成本同样不会大。

表 4.4 单只股票的仓位调整量占流通股本和总股本的比例

		创设问题 (02/07/01)		第一次调整 (03/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占流通股本比例	最大值	0.013531	0.015759	0.006814	0.005151
	最小值	0.000000	0.000006	0.000000	0.000000
	平均值	0.002388	0.002811	0.000472	0.000426
仓位调整量占总股本比例	最大值	0.007258	0.007131	0.002734	0.001360
	最小值	0.000000	0.000002	0.000000	0.000000
	平均值	0.000777	0.000875	0.000148	0.000114
		第二次调整 (03/07/01)		第三次调整 (04/01/02)	
		优化复制	抽样复制	优化复制	抽样复制
仓位调整量占流通股本比例	最大值	0.003566	0.004911	0.004434	0.005467
	最小值	0.000000	0.000000	0.000000	0.000000
	平均值	0.000419	0.000264	0.000303	0.000349
仓位调整量占总股本比例	最大值	0.001952	0.002786	0.002219	0.001367
	最小值	0.000000	0.000000	0.000000	0.000000
	平均值	0.000130	0.000090	0.000098	0.000106

同样根据 GA 计算的结果是，组合的总体流动性得到了提高。下图 4.5 是优化复制中的第二次调整（2003/07/01）的情况，图 4.6 是抽样复制中的第二次调整（2003/07/01）的情况，限于篇幅，仅以以上两例加以说明，在其它组合变动下都有同样的结论（两端累积百分比数值大小上会有所差异，但结论一致），在此不赘述。图中 case1 代表一般模型下的情况，GA 则代表流动性惩罚约束模型 GA 求解下的情况。图中左边图形的左半部分红线高于蓝线（对应于图中右边图形的左半部分两端累积百分比差值大于零），说明排在流动性较好的一半的资金配置比例是 GA 大于 case1。左边图形的右半部分（对应于图中右边图形的右半部分）情况刚好相反，所以结论也相反。这说明相对于一般模型的结果，流动性惩罚约束模型情况下资金更多地配置于更有流动性的资产上。资金在不同流动性资产上的配置变化如右边图形中的箭头所示。尽管优化复制中第二次调整下流动性改进的幅度不是很大，但总体上看流动性惩罚约束模型用 GA 求解的结论与 SQP 算法下的结论是一致的。

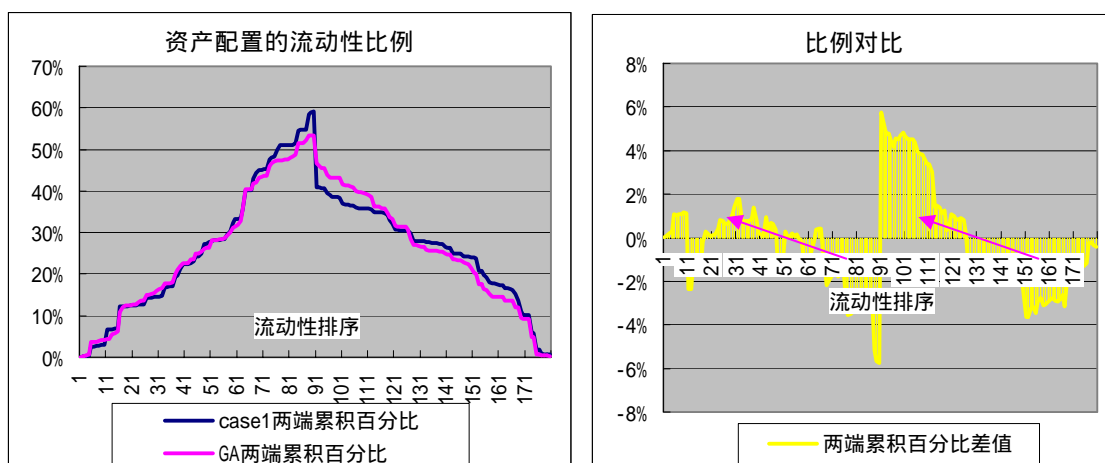


图 4.5 追踪组合总体流动性对比（优化复制中第二次调整）

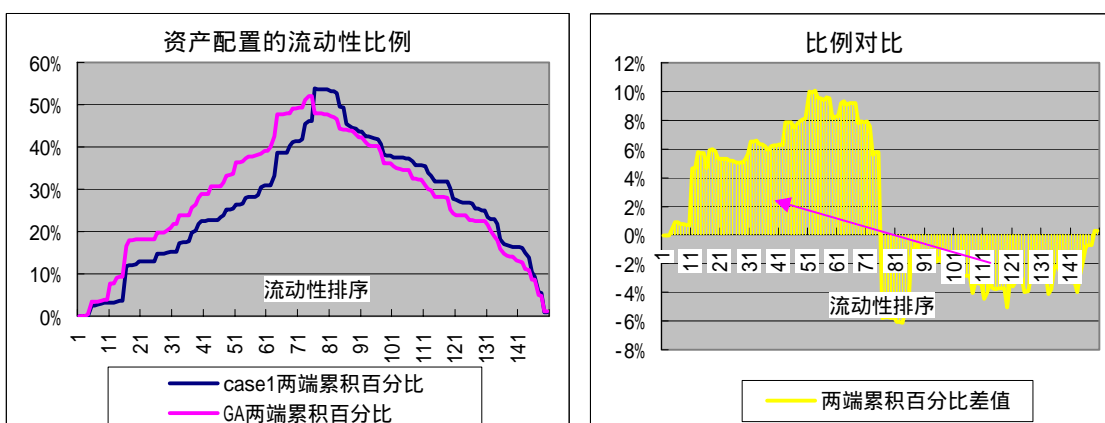


图 4.6 追踪组合总体流动性对比（抽样复制中第二次调整）

4.4.5 追踪误差与相关系数

下表 4.5 给出了根据 GA 优化结果计算的样本内以及样本外 5 个交易日、10 个交易日

日、30 个交易日、60 个交易日和截至下次指数调整日期期间的组合追踪误差以及组合与指数日收益率之间的相关系数。完全复制的结果与前文一致，为了便于比较，在此一并列出。虽然 GA 在 SQP 计算结果的基础上都有小幅度的改进（表现在 GA 在 SQP 计算结果的基础上进一步降低了样本内追踪误差），但仔细比较表 3.11 与下表 4.5，可以看出 SQP/GA（即 GA 在 SQP 计算结果的基础上的进一步优化）与单纯用 SQP 计算的结果差异不大，无论是样本内还是样本外，有 SQP/GA 计算的结果优于 SQP 的情况，也有劣于 SQP 的情况。而且无论是优于还是劣于，两者之间的差距都不大。

表 4.5 追踪误差和日收益率相关系数

	创设问题（02/07/01）			第一次调整（03/01/02）		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内（in sample）						
TE	9.3964e-5* (1.0331e-4)	1.1632e-4* (1.2552e-4)	-	9.0200e-5* (9.6494e-5)	9.5516e-5* (9.9206e-5)	-
TTE	0.0099* (0.0108)	0.0122* (0.0132)	-	0.0113* (0.0121)	0.0119* (0.0124)	-
ρ	0.9987	0.9981	-	0.9943	0.9934	-
样本外（out of sample）						
5 日						
TE	6.8028e-4	7.8931e-4	1.8559e-4	5.3074e-4	3.5195e-4	7.6079e-5
TTE	0.0034	0.0039	9.2794e-4	0.0027	0.0018	3.8040e-4
ρ	0.9986	0.9979	0.9991	0.9948	0.9943	0.9999
10 日						
TE	4.9390e-4	5.2491e-4	1.1488e-4	4.8264e-4	3.9703e-4	4.9608e-5
TTE	0.0049	0.0052	0.0011	0.0048	0.0040	4.9608e-4
ρ	0.9985	0.9978	0.9995	0.9954	0.9951	0.9999
30 日						
TE	2.7966e-4	2.7583e-4	4.9594e-5	2.1367e-4	1.8867e-4	3.3110e-5
TTE	0.0084	0.0083	0.0015	0.0064	0.0057	9.9331e-4
ρ	0.9980	0.9974	0.9995	0.9956	0.9953	0.9999
60 日						
TE	1.7031e-4	1.6727e-4	3.0098e-5	1.6643e-4	1.6219e-4	2.4895e-5
TTE	0.0102	0.0100	0.0018	0.0100	0.0097	0.0015
ρ	0.9977	0.9971	0.9996	0.9949	0.9944	0.9999
截至下次指数调整日						
TE	1.0560e-4	1.0756e-4	2.1922e-5	1.4127e-4	1.5944e-4	3.3221e-5
TTE	0.0133	0.0136	0.0027	0.0162	0.0183	0.0038
ρ	0.9973	0.9968	0.9997	0.9939	0.9923	0.9996
	第二次调整（03/07/01）			第三次调整（04/01/02）		
	优化复制	抽样复制	完全复制	优化复制	抽样复制	完全复制
样本内（in sample）						
TE	1.2268e-4* (1.2274e-4)	1.6421e-4* (1.6111e-4)	-	2.807e-4* (2.6978e-4)	3.7623e-4* (3.6122e-4)	-

指数组合优化方法、模型与应用 - 基于上证 180 指数的实证研究

<i>TTE</i>	0.0135* (0.0141)	0.0181* (0.0177)	-	0.0087* (0.0084)	0.0117* (0.0112)	-
ρ	0.9947	0.9915	-	0.9922	0.9857	-
样本外 (out of sample)						
5 日						
<i>TE</i>	6.2489e-4	6.5244e-4	8.7233e-5	8.4711e-4	0.0016	1.0582e-4
<i>TTE</i>	0.0031	0.0033	4.3616e-4	0.0042	0.0080	5.2908e-4
ρ	0.9945	0.9913	0.9999	0.9922	0.9818	0.9999
10 日						
<i>TE</i>	4.3387e-4	4.5121e-4	8.1297e-5	6.9059e-4	8.7952e-4	6.2045e-5
<i>TTE</i>	0.0043	0.0045	8.1297e-4	0.0069	0.0088	6.2045e-4
ρ	0.9946	0.9912	0.9995	0.9912	0.9848	0.9999
30 日						
<i>TE</i>	2.2297e-4	2.1102e-4	4.3242e-5	3.9075e-4	4.2662e-4	5.5972e-5
<i>TTE</i>	0.0067	0.0063	0.0013	0.0117	0.0128	0.0017
ρ	0.9941	0.9913	0.9995	0.9902	0.9861	0.9998
60 日						
<i>TE</i>	1.7057e-4	1.6521e-4	2.5759e-5	-	-	-
<i>TTE</i>	0.0102	0.0099	0.0015	-	-	-
ρ	0.9932	0.9906	0.9997	-	-	-
截至下次指数调整日						
<i>TE</i>	1.4011e-4	1.2687e-4	1.9852e-5	2.9933e-4	3.4364e-4	3.9591e-5
<i>TTE</i>	0.0178	0.0161	0.0025	0.0132	0.0151	0.0017
ρ	0.9918	0.9905	0.9998	0.9911	0.9865	0.9998

*注释：*为优化计算未取整的结果，其余数据都为根据最小交易手数取整后计算的结果。

4.4.6 追踪组合的表现

同样地，针对每一种复制方法下的每一次仓位变动（包括创设问题以及第一次至第三次调整），本研究都画出了 GA 计算结果下的指数点位与组合市值走势图、指数与组合样本内累积收益率图、样本外累积收益率图、指数与组合日收益率图以及日收益率差值图等。限于篇幅，以下部分仅以抽样复制中的第三次调整（04/01/02）的情况为例，其余参见附录中的附图 1。

根据优化计算的结果画出指数点位与组合市值走势图 4.7。

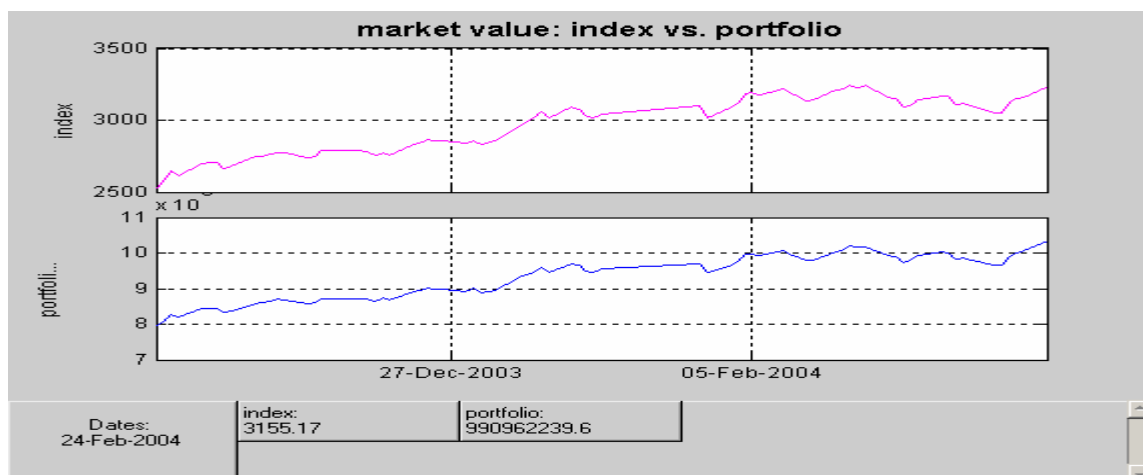


图 4.7 指数点位与组合市值走势 (2003/11/18 ~ 2004/03/15)

指数与组合日收益率图则见下图 4.8。

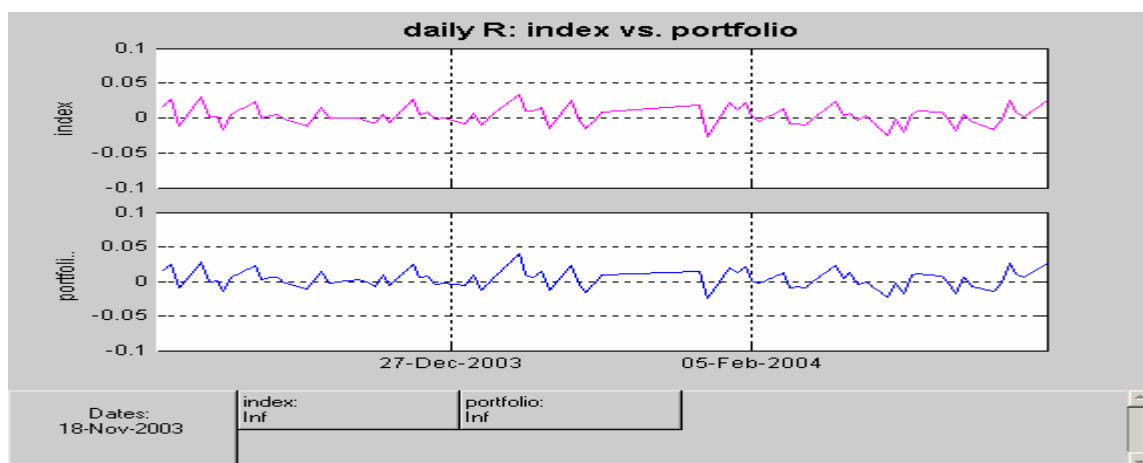


图 4.8 指数与组合日收益率 (2003/11/18 ~ 2004/03/15)

流动性惩罚约束模型下指数与组合在样本内和样本外的累积收益率图以及日收益率差值图,分别见下图 4.9 至图 4.11。同样以 2003 年 11 月 18 日作为收益率累积的起点来计算每个交易日相对于 2003 年 11 月 18 日指数与组合各自的对数收益率,即相对 2003 年 11 月 18 日的累积收益率。从上图可以看出,样本内的追踪效果也相当不错。见表 4.5,事实上,该时间段的实际日均追踪误差仅为 $3.6122e-4$,指数与组合日收益率的相关系数为 0.9857,样本内的追踪精度比同期一般模型下的结果略好。

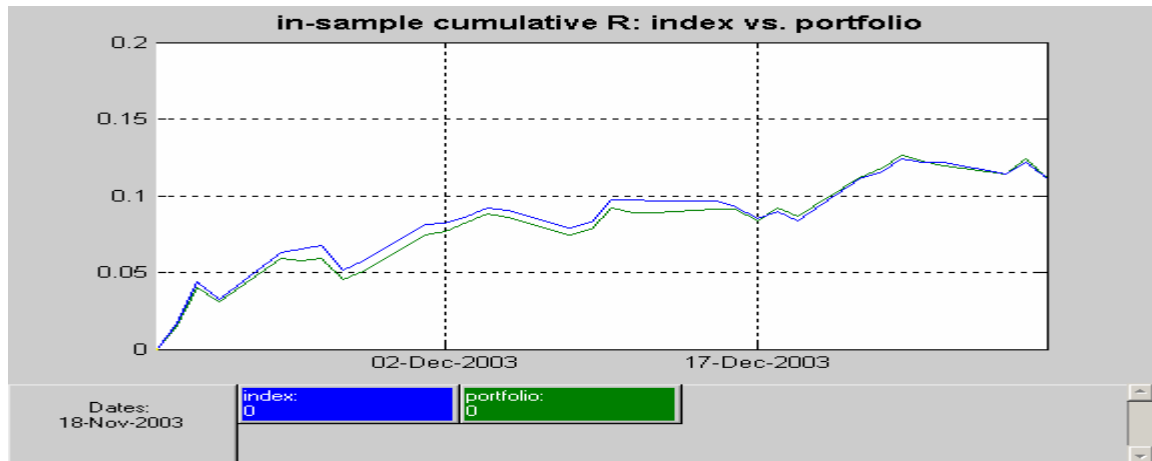


图 4.9 样本内累积收益率 (2003/11/18 ~ 2003/12/31)

下图刻画出了流动性惩罚约束模型 GA 求解下指数与组合在模型之外的追踪效果，尾部有所偏差，但总体效果还不错，见表 4.5，样本外 5 个交易日的日均追踪误差为 0.0016，累积追踪误差为 0.0080，指数与组合收益率的相关系数为 0.9818，而样本外整个时间段即截至下次指数调整日的日均追踪误差为 3.4364×10^{-4} ，累积追踪误差为 0.0151，指数与组合收益率的相关系数为 0.9865，这个结论与上图揭示的结果一致。

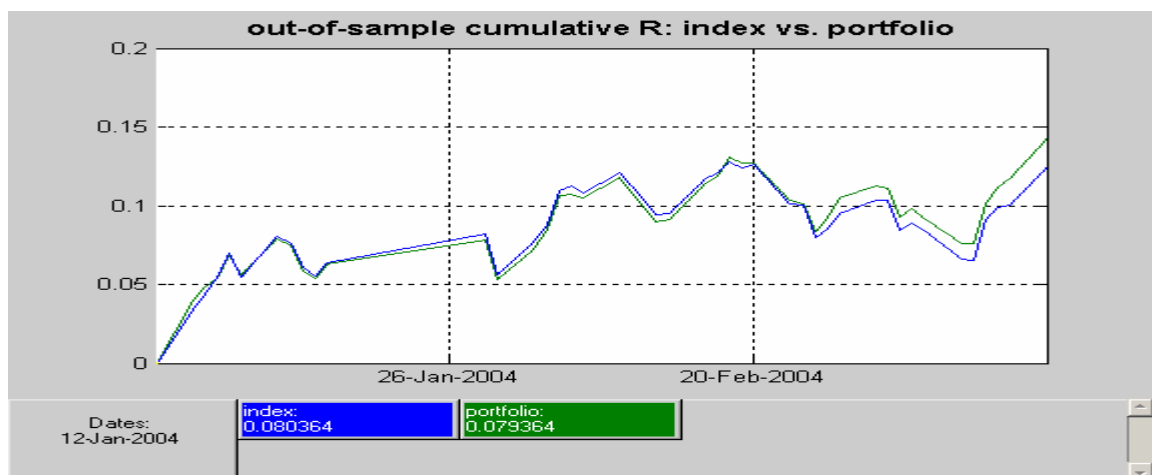


图 4.10 样本外累积收益率 (2004/01/02 ~ 2004/03/15)

从下图指数与组合日收益率差值中可以看出，流动性惩罚约束模型下指数与组合的日收益率之差也基本保持在正负千分之二点五之间，只有个别日的差值比较大，说明组合的追踪效果不错。

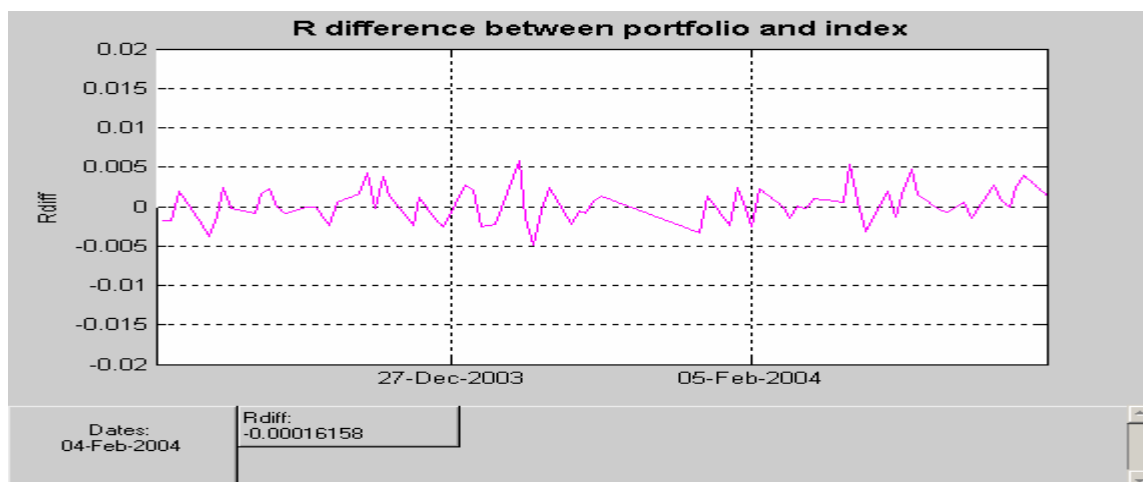


图 4.11 指数与组合的日收益率差值 (2003/11/18 ~ 2004/03/15)

4.4.7 追踪组合构成比较

下图 4.12 是流动性惩罚约束模型 GA 求解下组合调整前后的构成比较,图形左边是抽样之后所有 150 只股票的个股持有情况,右边则是取左图前 10 只股票放大比较的情况。同样可以看出组合调整前后个股的持有量并没有很大的变化。这也是由于抽样复制的模型 (7) 中的约束 5) 即仓位调整所引起的交易成本约束作用的结果。

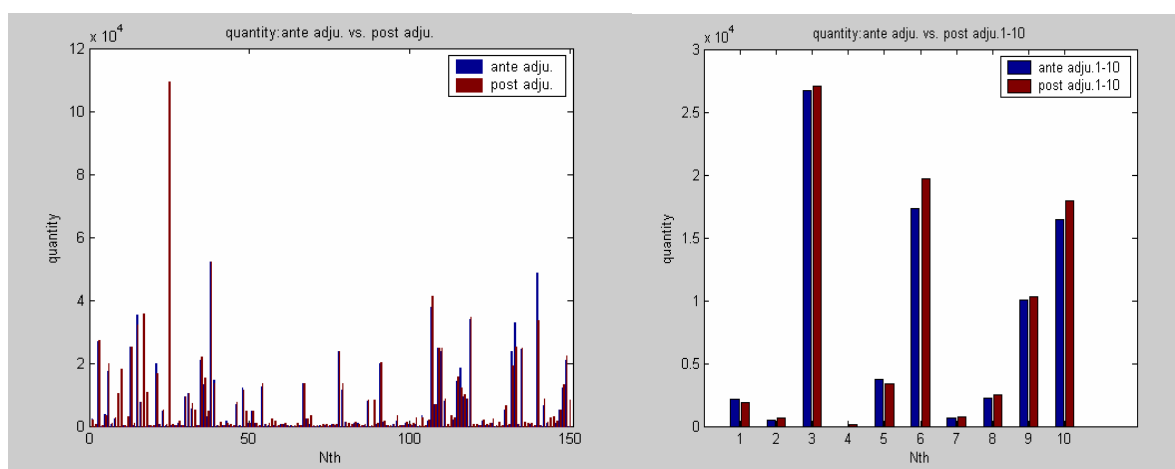


图 4.12 组合调整前后个股持有量比较

图 4.13 是流动性惩罚约束模型 GA 求解下上证 180 指数内对应成分股根据对应权重计算的对应股数与追踪组合内对应成分股的对应股数的比较图,图形右边同样是取左图前 10 只股票放大比较的情况。可以看出根据对应权重计算的上证 180 指数内对应成分股的对应股数与追踪组合内对应成分股的对应股数之间存在很大的区别,这是组合优化的结果。

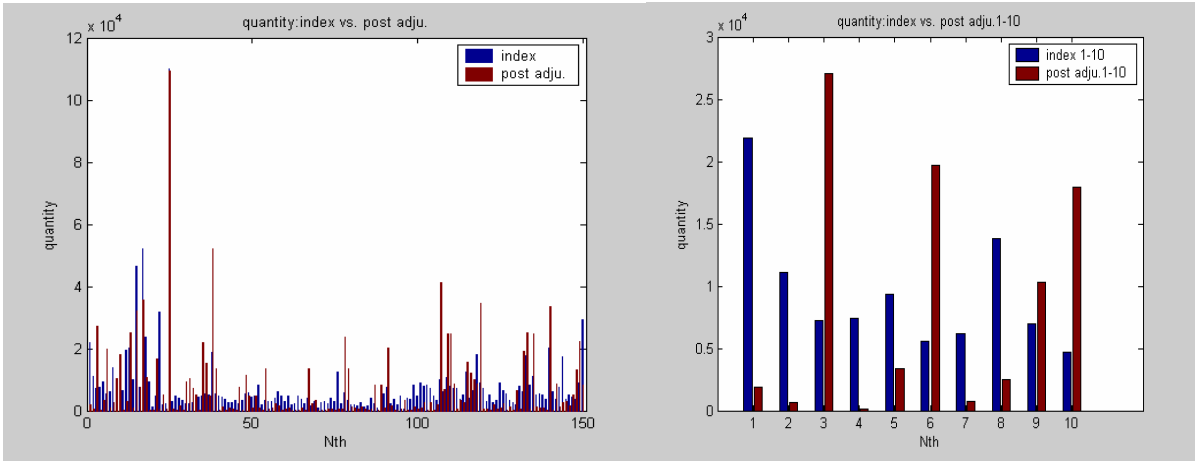


图 4.13 组合与指数对应个股的对应股数比较

5. 实践评估

5.1 评估目的及数据来源

目前我国共有各类准指数基金 9 只，包括天同 180、华安 180、融通 100、易方达 50、基金景福、博时裕富和长盛中信全债等（见下表 5.1）。由于以上几只基金的产品架构与交易模式等诸多方面与国外真正的指数基金比还有很大的差异，而且目前国内对这些基金的投资品种和投资比例还有政策上的限制，因此这些基金都只是一种“准指数基金”。

表 5.1 国内准指数基金

名 称	类 型	设立日期	交易代码	标的指数
基金普丰	封闭式	1999/07/08	184693	深圳 A 股综合指数
基金兴和	封闭式	1999/07/08	500018	上证 A 股综合指数
基金景福	封闭式	1999/12/24	184701	上证 A 股综合指数
天同 180	开放式	2003/02/10	519180	上证 180 指数
华安 180	开放式	2002/10/15	040002	上证 180 指数
博时裕富	开放式	2003/07/10	050002	新华富时中国 A200 复合指数
融通 100	开放式	2003/08/26	161003	深证 100 指数
易方达 50	开放式	2004/03/22	110003	上证 50 指数
长盛全债	开放式	2003/07/31	500080	中信全债指数

从表 5.1 可以看出，只有天同 180 和华安 180 两只准指数基金追踪上证 180 指数，这与本课题所追踪的标的指数一致。不过由于这两只基金只是一种“准指数基金”，除了要投资于国债之外，还有一定比例的资金积极参与一级市场的新股申购，而且华安 180 实际上还是一种指数增强型基金，其投资目标与本课题设定的追踪指数（track the index）的目标不同，因此严格意义上讲，这两只基金与本研究中所要构造的指数组合之间并没有准确的可比性。不过有一点应该肯定，本课题优化的结果从追踪误差上看应该比天同 180 和华安 180 好，如果这个目标没有达到，则本课题研究结果的实用价值就值得质疑。为此，本研究试图根据公开、可获得的数据对天同 180 和华安 180 这两只准指数基金的投资绩效进行简要评估，进而与本课题的研究结果进行比较，目的是为本课题研究结果的实用价值寻找一定的依据。

评估主要依据基金净值（包括日净值和累计净值）以及基金分红的数据，这些数据都是从天同基金管理公司网站（<http://www.ttasset.com>）和华安基金管理公司网站（<http://www.huaan.com.cn>）上获得的。

5.2 天同 180 指数基金

天同 180 指数基金 (以下继续简称“天同 180”) 2003 年 2 月 10 日开始发行, 2003 年 4 月 25 日才有净值数据公布, 因此对天同 180 的绩效评估一律从 2003 年 4 月 25 日开始计算, 截至时间为 2004 年 3 月 15 日。

(1) 净值及收益率表现

首先看看天同 180 的净值表现, 如下图 5.1 所示。这个净值根据公司公布的基金分红数据做了复权调整。

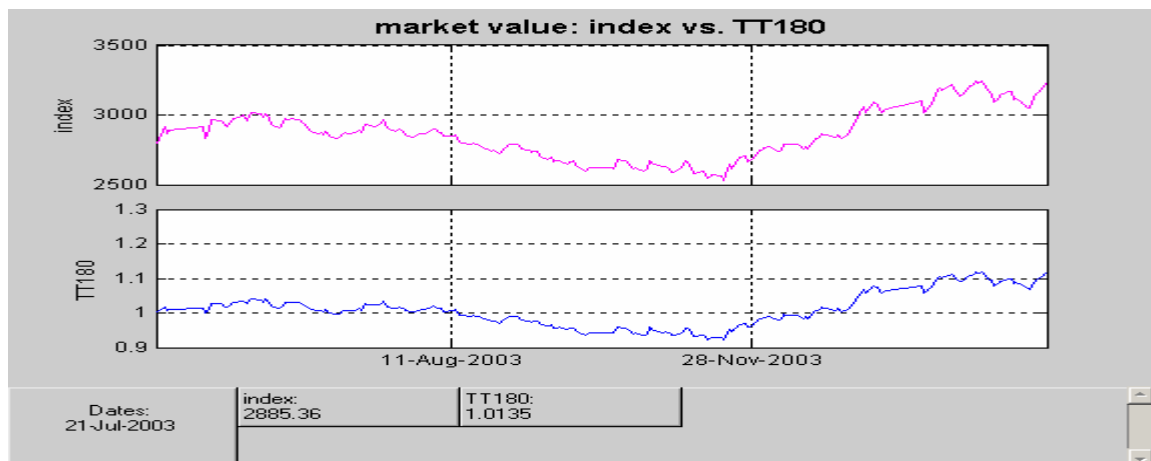


图 5.1 指数与天同 180 的净值表现 (2003/04/25 ~ 2004/03/15)

图 5.2 是天同 180 相对上证 180 指数的日收益率表现, 从图中可以看出似乎天同 180 的日收益率波动小于上证 180 指数的日收益率波动, 一个可能的原因是天同 180 指中持有一定比例的国债。

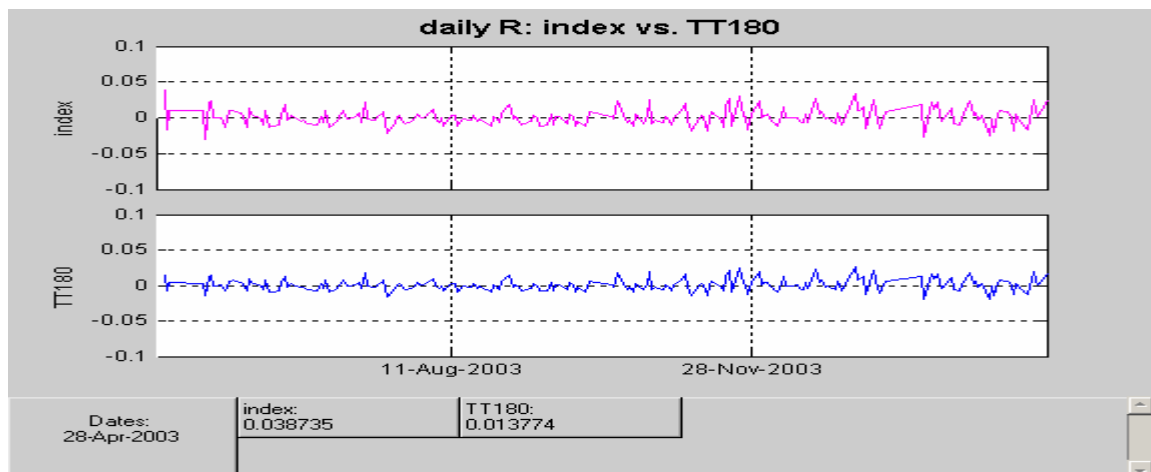


图 5.2 指数与天同 180 的日收益率表现 (2003/04/25 ~ 2004/03/15)

(2) 收益率差值评估

从图 5.3 中可以看出, 天同 180 与指数的日收益率差值还是比较大。

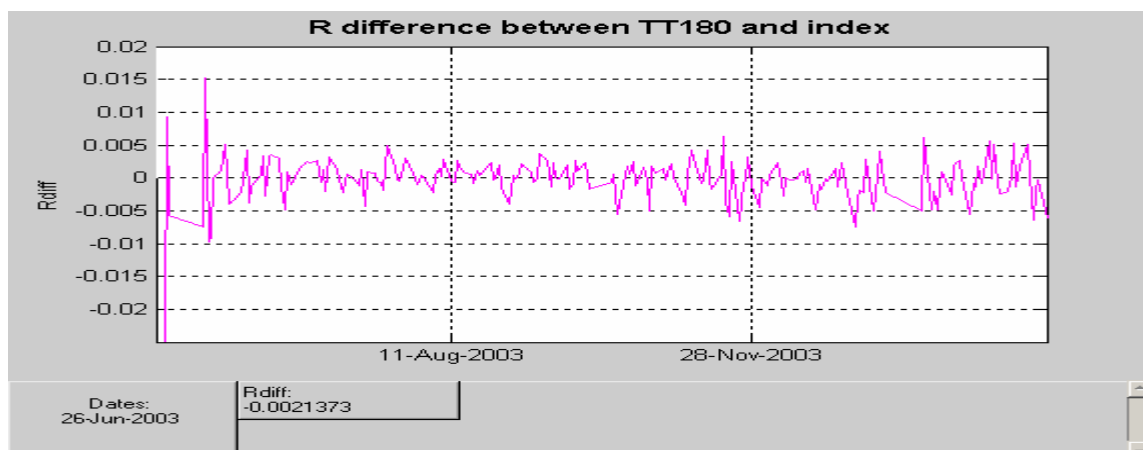


图 5.3 天同 180 与指数的日收益率差值 (2003/04/25 ~ 2004/03/15)

截取其中的一个时间段,即 2003 年 7 月 1 日至 2003 年 12 月 31 日的日收益率差值与优化复制中(限于篇幅,此中仅以一般模型下的 SQP 求解的结果为例)同期样本外的日收益率差值进行对比,如下图 5.4,左边小图为天同 180 与指数的在 2003 年 7 月 1 日至 2003 年 12 月 31 日期间的日收益率差值,右边小图为同期追踪组合与指数的日收益率差值。显然,前者的波动幅度大于后者。如果以其它时间段进行类似的对比,结论还是如此。可见,从收益率差值直观看,天同 180 的实际追踪效果不如本研究中的优化复制的效果。

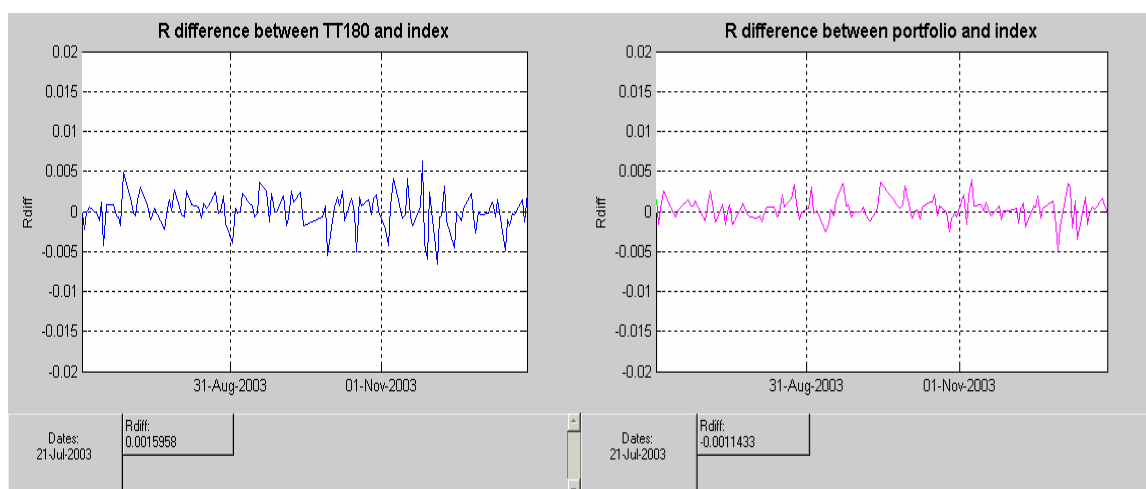


图 5.4 日收益率差值比较 (2003/07/01 ~ 2003/12/31)

(3) 累积收益率评估

图 5.5 是天同 180 与指数的累积收益率。

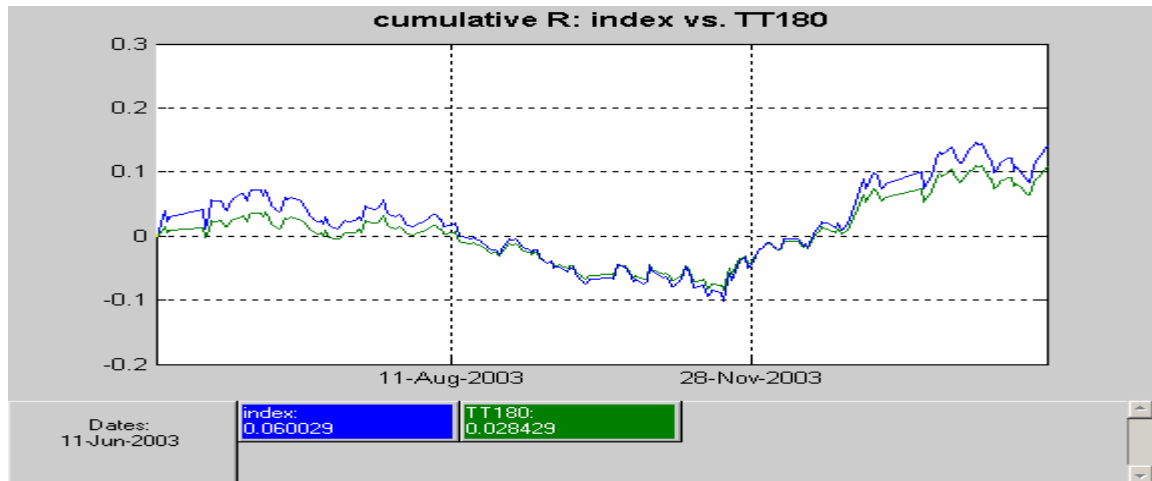


图 5.5 天同 180 与指数的累积收益率 (2003/04/25 ~ 2004/03/15)

截取其中的一个时间段,即 2003 年 4 月 25 日至 2003 年 7 月 1 日的累积收益率与抽样复制中(限于篇幅,此中仅以一般模型下的 SQP 求解的结果为例)同期样本外的累积收益率进行对比,如下图 5.4,左边小图为天同 180 与指数的在 2003 年 4 月 25 日至 2003 年 7 月 1 日的累积收益率,右边小图为同期追踪组合与指数的累积收益率。显然,前者拟合的效果远不如后者。如果以其它时间段进行类似的对比,结论还是如此。可见,从累积收益率直观看,天同 180 的实际追踪效果不如本研究中的抽样复制的效果。

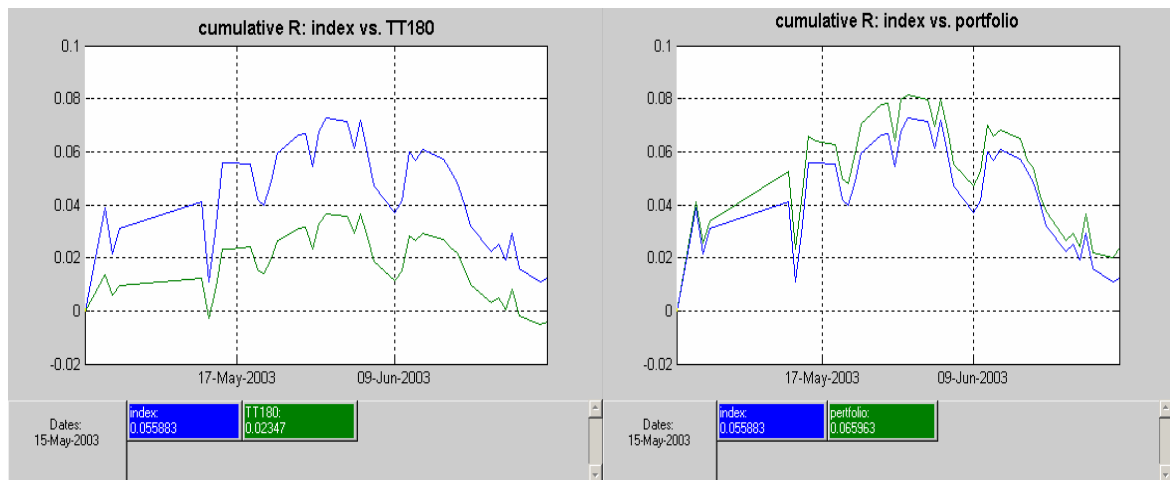


图 5.4 累积收益率比较 (2003/04/25 ~ 2003/07/01)

(4) 追踪误差和相关系数评估

表 5.2 给出了追踪误差和相关系数评估的结果,其中 TE 为日均追踪误差,TTE 为累积追踪误差, ρ 为相关系数,TT180 代表天同 180 指数基金,HA180 代表华安 180 指数增强型基金。优化复制、抽样复制两者都是一般模型下的 SQP 优化求解的结果,限于篇幅,仅以此为例对此加以说明。为了比较上的方便,把华安 180 指数增强型基金的结果一并列上。从表中可以看出,无论是哪个时间段,优化复制、抽样复制两者的追踪误差都远小于天同 180 的追踪误差,多数情况下优化复制、抽样复制两者的追踪误差仅是天同 180 实际追踪误差的一半。

表 5.2 追踪误差和相关系数

	03/01/02~03/07/01			03/07/01~03/12/31			04/01/02~04/03/15		
	TE	TTE	ρ	TE	TTE	ρ	TE	TTE	ρ
TT180	9.4249e-4*	0.0377*	0.9658*	1.8825e-4	0.0239	0.9985	5.1782e-4	0.0228	0.9990
HA180	4.8292e-4	0.0555	0.9546	2.2730e-4	0.0289	0.9893	5.6457e-4	0.0248	0.9891
优化复制	1.6300e-4 (3386e4)*	0.0187 (0.0134)*	0.9924 (0.9872)*	1.2838e-4	0.0163	0.9913	2.6155e-4	0.0115	0.9920
抽样复制	1.7010e-4	0.0196	0.9917	1.4733e-4	0.0187	0.9874	3.7785e-4	0.0166	0.9856

注释：*计算起止时间为 03/04/25~03/07/01

5.3 华安 180 指数增强型基金

华安 180 指数增强型基金（以下继续简称“华安 180”）2002 年 10 月 15 日开始发行，2002 年 12 月 19 日才有净值数据公布，因此对华安 180 的绩效评估一律从 2002 年 12 月 19 日开始计算，截至时间为 2004 年 3 月 15 日。

1. 净值及收益率表现

首先看看华安 180 的净值表现，如下图 5.5 所示。这个净值根据公司公布的基金分红数据做了复权调整。

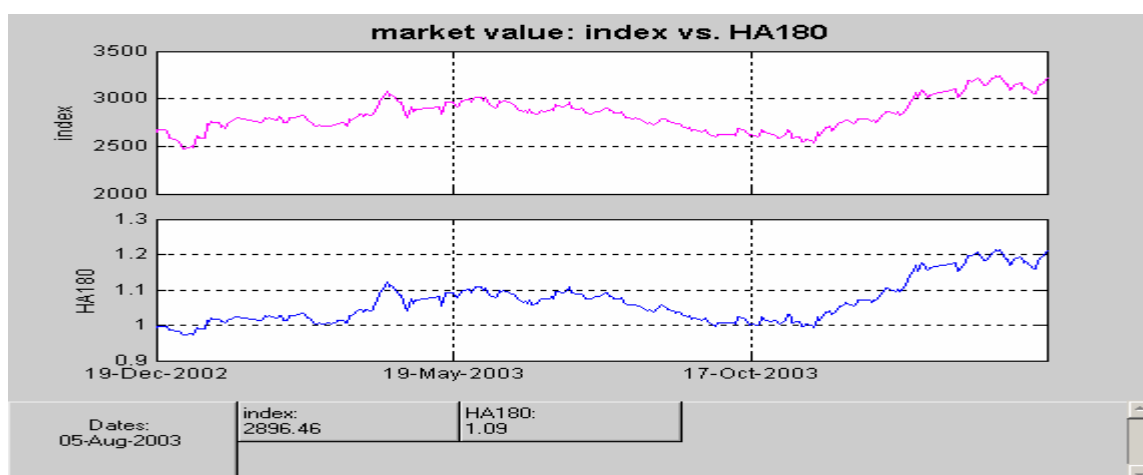


图 5.5 指数与华安 180 的净值表现（2002/12/19 ~ 2004/03/15）

图 5.6 是华安 180 相对上证 180 指数的日收益率表现，从图中可以看出似乎华安 180 的日收益率波动小于上证 180 指数的日收益率波动，同样可能的原因是华安 180 中持有一定比例的国债。

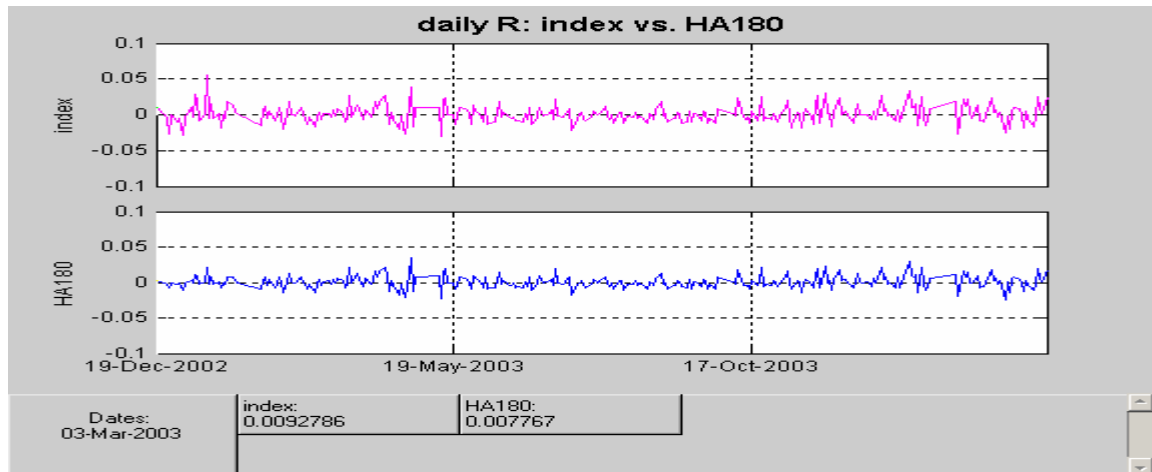


图 5.6 指数与华安 180 的日收益率表现 (2002/12/19 ~ 2004/03/15)

2. 收益率差值评估

从图 5.7 中可以看出，华安 180 与指数的日收益率差值相当大。

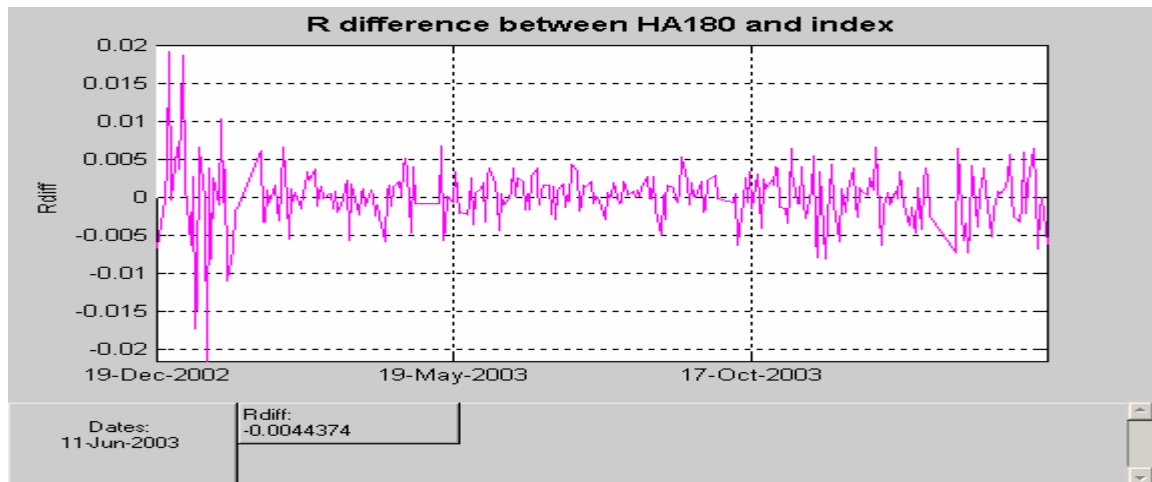


图 5.7 华安 180 与指数的日收益率差值 (2002/12/19 ~ 2004/03/15)

截取其中的一个时间段，即 2003 年 7 月 1 日至 2003 年 12 月 31 日的日收益率差值与抽样复制中（限于篇幅，此中仅以流动性惩罚约束模型下的 SQP 求解的结果为例）同期样本外的日收益率差值进行对比，如下图 5.4，左边小图为华安 180 与指数的在 2003 年 7 月 1 日至 2003 年 12 月 31 日期间的日收益率差值，右边小图为同期追踪组合与指数的日收益率差值。显然，前者的波动幅度大于后者。如果以其它时间段进行类似的对比，结论还是如此。可见，从收益率差值直观看，华安 180 的实际追踪效果不如本研究中的抽样复制的效果。

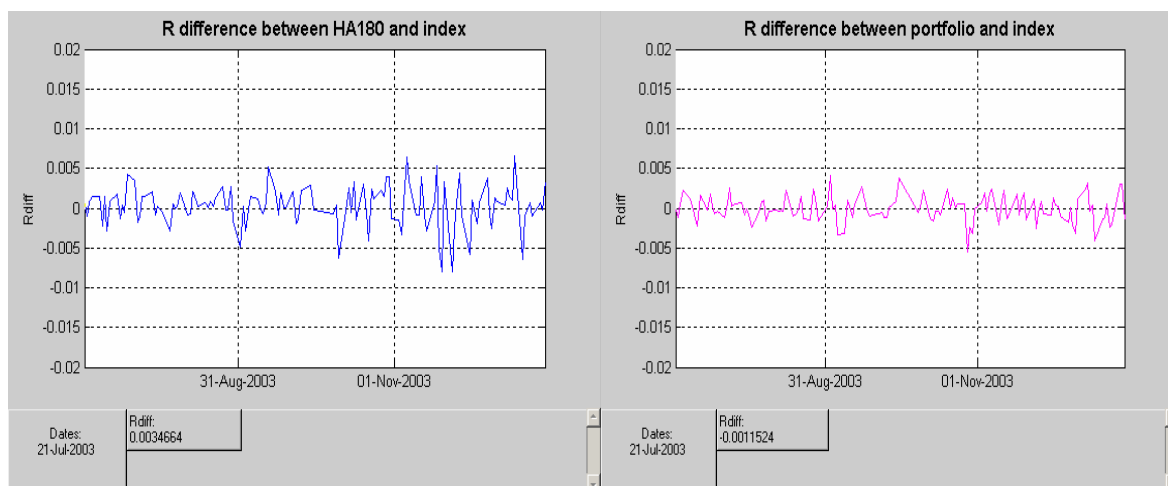


图 5.4 日收益率差值比较 (2003/07/01 ~ 2003/12/31)

3. 累积收益率评估

图 5.5 是华安 180 与指数的累积收益率。

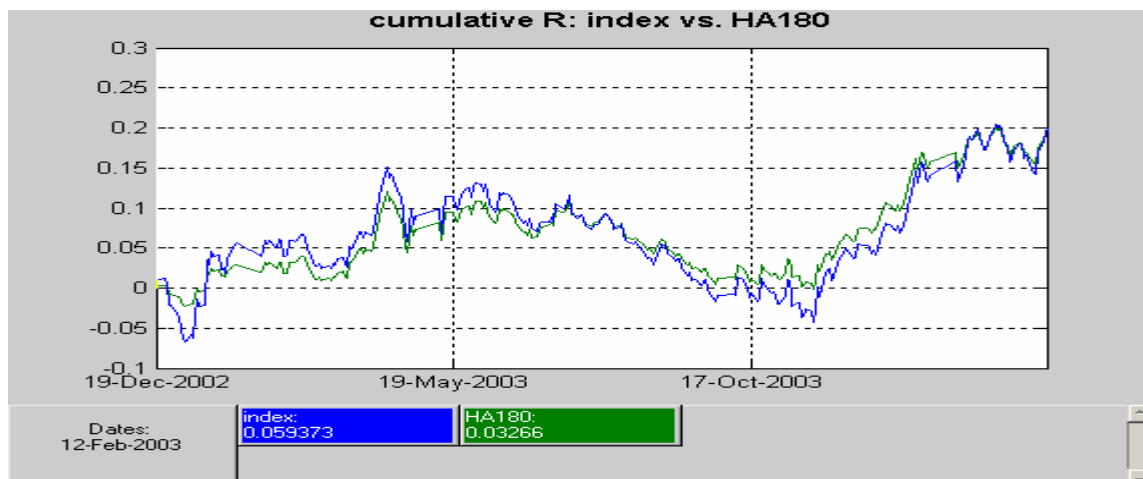


图 5.5 华安 180 与指数的累积收益率 (2002/12/19 ~ 2004/03/15)

截取其中的一个时间段,即 2003 年 1 月 2 日至 2003 年 7 月 1 日的累积收益率与优化复制中(限于篇幅,此中仅以流动性惩罚约束模型下的 GA 求解的结果为例)同期样本外的累积收益率进行对比,如下图 5.4,左边小图为华安 180 与指数的在 2003 年 1 月 2 日至 2003 年 7 月 1 日的累积收益率,右边小图为同期追踪组合与指数的累积收益率。显然,前者拟合的效果远不如后者。如果以其它时间段进行类似的对比,结论还是如此。可见,从累积收益率直观看,华安 180 的实际追踪效果不如本研究中的优化复制的效果。

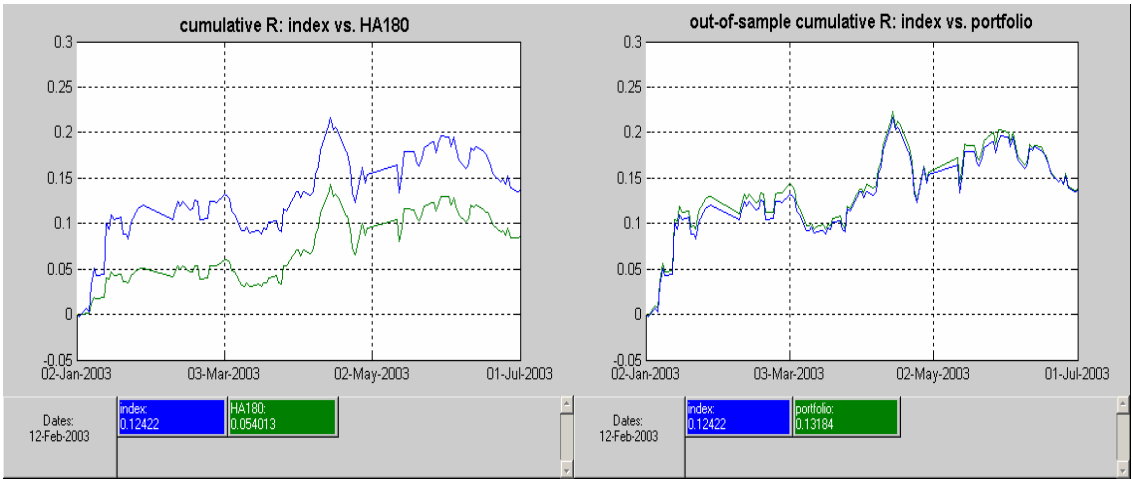


图 5.4 累积收益率比较 (2003/01/02 ~ 2003/07/01)

4 . 追踪误差和相关系数评估

追踪误差和相关系数评估的结果见表 5.2 , 在此就不重复。从表中可以看出 , 无论是哪个时间段 , 优化复制、抽样复制两者的追踪误差都远小于华安 180 的追踪误差 , 多数情况下优化复制、抽样复制两者的追踪误差仅是华安 180 实际追踪误差的一半。

6. 主要结论

本文以上证 180 指数为标的，研究了从 2002 年 7 月 1 日上证 180 指数首次发布之日起，一直到 2004 年 3 月 15 日为止的这个时间段中的指数组合优化问题。文中分别研究了优化复制和抽样复制这两种指数组合优化的方法、模型及运用，而且为了比较，对完全复制问题也做了研究。每一种复制方法下都同时考查了四个组合（即创设问题以及随后的三次调整）的追踪效果（包括追踪误差、追踪成本以及组合流动性等）。针对优化复制和抽样复制问题，文中构造了两类模型（即一般模型和流动性惩罚约束模型），并运用了两种算法（即 SQP 和 GA）在 MATLAB 环境下进行优化求解。抽样复制中先在 SAS8.2 环境下根据个股流动性和调整流通市值两个指标完成抽样，个股流动性的衡量是基于市场微观结构理论的研究成果利用高频数据完成的。此外对天同 180 指数基金和华安 180 指数增强型基金的追踪效果也做了简要评估。主要结论总结如下：

1. 优化复制和抽样复制无论在样本内还是在样本外的追踪效果都相当不错。

这个追踪效果以指数组合收益率相对上证 180 指数收益率的日均追踪误差、累积追踪误差以及相关系数来衡量。一般模型在 SQP 算法下的计算结果见表 3.6，流动性惩罚约束模型在 SQP 算法下的计算结果见表 3.11，流动性惩罚约束模型在 GA 算法下的计算结果见表 4.5，描述性统计值见附表 5 至附表 7，各种情况下的平均值比较见附表 8。从这些表中可知，样本内时间区间为半年且按照最小交易手数取整的情况下，优化复制和抽样复制综合的情况是样本内日均追踪误差在 $6.6128e-5$ 至 $1.9053e-4$ 之间（总平均值即两类模型、两种算法下的综合平均值为 $1.1681e-4$ ），样本内半年累积追踪误差在 0.0069 至 0.0205 之间（总平均值为 0.0132），指数与追踪组合收益率相关系数在 0.9888 至 0.9995 之间（总平均值为 0.9949）；样本外 5 日、10 日、30 日、60 日以及截至下次指数调整日的日均追踪误差在 $1.0560e-4$ 至 $7.8931e-4$ 之间，累积追踪误差在 0.0018 至 0.0196 之间，相关系数在 0.9874 至 0.9999 之间；可见优化复制和抽样复制无论在样本内还是在样本外的追踪效果都相当不错。追踪组合与上证 180 指数之间的追踪效果也可以从累积收益率走势图、收益率差值图等直观反应出来，见附图 3，一般模型在 SQP 算法下的抽样复制中的第一次调整（03/01/02）情况见图 3.2 至图 3.6、流动性惩罚约束模型在 SQP 算法下的优化复制中的第二次调整（03/07/01）情况见图 3.12 至图 3.16 以及流动性惩罚约束模型在 GA 算法下的抽样复制中的第三次调整（04/01/02）情况见图 4.7 至图 4.11。从这些图中可以看出，指数组合对上证 180 指数的追踪效果都很好。

2. 不过相对而言，完全复制的复制精度最高，其次是优化复制，最后是抽样复制。而且样本股数目与追踪误差之间存在着一种此涨彼伏、相互抵消（trade off）的关系。

衡量复制精度的尺度是追踪误差（相关系数也是一个重要的参考指标）。见附表 8，

对优化复制和抽样复制来讲，无论是哪一类模型还是哪一种算法下的计算结果，样本外追踪误差都远大于完全复制的结果。优化复制与抽样复制之间，则是优化复制小于抽样复制，尽管如果就时间长度为 5 至 60 日的样本外追踪误差看，有优化复制小于抽样复制的情况，也有优化复制大于抽样复制的情况，但优化复制小于抽样复制的次数多。而且就样本内追踪误差和时间长度为截至下次指数调整日的样本外追踪误差看，优化复制无一例外都小于抽样复制。可见从复制精度看，完全复制最高，其次是优化复制，最后是抽样复制。抽样复制由于剔除了 30 只“问题证券”从而提高了追踪组合的质量，但却以损害复制精度为代价，其相对于优化复制和完全复制来讲，追踪误差加大。这也从侧面证明了样本股数目与追踪误差之间存在着一种此涨彼伏、相互抵消（trade off）的关系。

3. 但从交易成本看，抽样复制最小，其次是优化复制，最后是完全复制。追踪误差与交易成本之间也存在着相互抵消的关系。

附表 9 计算了包括交易成本的绝对金额和交易成本占资金规模的相对比例这两个指标的平均值。结论非常明显，即从交易成本看，抽样复制最小，其次是优化复制，最多的是完全复制（尽管流动性惩罚约束模型在 SQP 算法下的计算结果是抽样复制的交易成本略大于优化复制，但这只是一种情况，另外两种情况下都是抽样复制小于优化复制）。同样 10 亿元规模的最初资金，整个追踪下来，各自的交易总成本如下：一般模型在 SQP 算法下的计算结果是抽样复制的交易总成本为 4.4615e+6 元，优化复制为 4.5018e+6 元，抽样复制小于优化复制；流动性惩罚约束模型在 SQP 算法下的计算结果是抽样复制的交易总成本为 4.4336e+6 元，优化复制为 4.3909e+6 元，抽样复制也是小于优化复制；流动性惩罚约束模型在 GA 算法下的计算结果是抽样复制的交易总成本为 4.4069e+6 元，优化复制为 4.4109e+6 元，抽样复制还是小于优化复制；而完全复制为 4.9435e+6 元，完全复制比抽样复制和优化复制都要高出 50 万元左右。可见完全复制的高精度是以较大的交易成本为代价的。而抽样复制剔除了 30 只“问题证券”，损害了复制精度但却减少了交易成本。可见追踪误差与交易成本之间也存在着相互抵消的关系。

4. 而从仓位变动对市场的冲击成本看，无论选择哪种复制方法，冲击成本都较小且可控。

由于影响市场冲击成本的因素如同影响股价波动的因素一样多而无法精确计算，因此本研究中没有对因仓位变动而可能带来的市场冲击成本予以度量。不过从个股仓位调整占该股流通股比例的描述性统计中可以大致了解一点。一般模型在 SQP 算法下的计算结果见表 3.3，流动性惩罚约束模型在 SQP 算法下的计算结果见表 3.8，流动性惩罚约束模型在 GA 算法下的计算结果见表 4.2。可见创设时，个股仓位调整占该股流通股的比例在千分之二点五左右，调整时则在万分之四左右，应该说这个比例很小，不足以因之而引起股价很大的波动，因此市场冲击成本应该很小。而且由于允许分阶段建仓，抽样复制中还按照个股的流动性来进行抽样，流动性惩罚约束模型下让更多的资金配置

于比较有流动性的资产上，因此组合建仓时间的延长和组合内个股流动性的提高都会进一步减少可能的市场冲击成本。

5. 其它条件不变的情况下，样本内的时间窗越长，追踪效果越好，而且样本内追踪效果越好，样本外的表现也越好。

第三次调整中由于上证 180 指数纳入 2003 年 11 月 18 日才上市的长江电力(600900) 为成分股，因此样本内的时间统一从 2003 年 11 月 18 日起算至 2004 年 1 月 2 日，只有 32 个交易日。无论是从表 3.6、表 3.11 还是表 4.5，无一例外的是，第三次调整的样本内还是样本外（5 日、10 日、30 日、60 日还是截至下次指数调整日）的追踪误差都大于创设问题以及随后第一和第二调整的追踪误差。两类模型、两种算法下计算的结果都是一致的。可见样本内时间短，样本内的追踪误差就大，并进而影响到样本外的追踪效果。因此，其它条件不变的情况下，样本内的时间窗越长，追踪效果越好，而且样本内追踪效果越好，样本外的表现也越好，这一点与谢文良、李进生、谢素娟（2002）等人的研究结论一致。

6. 资产配置时由于最小交易手数的限制而进行的个股买卖手数取整对现金误差的影响可以忽略不计，对组合样本内追踪误差的影响不能确定，可能加大，也可能减少组合样本内的追踪误差，而且因时间段不同而有所差异。

从现金差额以及这个差额占资金规模的比例看，资产配置时由于最小交易手数的限制而进行的个股买卖手数取整所带来的影响几乎可以忽略不计。一般模型在 SQP 算法下的计算结果见表 3.5，流动性惩罚约束模型在 SQP 算法下的计算结果见表 3.10，流动性惩罚约束模型在 GA 算法下的计算结果见表 4.4，平均现金差额在两三千元左右，占总资金规模的百万分之三左右。而这种取整可能会加大，但也可能会减小组合在样本内的追踪误差。从表 3.5、表 3.10 以及表 4.4 可以看出，在创设问题以及随后的第一次调整中是加大了组合样本内的追踪误差；而在第二次以及第三次调整中则减少了组合样本内的追踪误差。不过无论是加大还是减少，幅度都很小，而且如果把加大和减少相互抵消的话，则可以说资产配置时由于最小交易手数的限制而进行的个股买卖手数取整对组合样本内追踪误差的影响也可以忽略不计。

7. 从追踪组合中个股的集中度看，抽样复制较高，优化复制其次，而完全复制最分散。不过无论哪种方法，都没有突破模型中设定的比例上限。抽样复制中由于考虑了行业的分散性和代表性，因此尽管个股集中度略高一点，也不影响组合的分散化。

追踪组合中的个股集中度指的是个股的资金配置量占组合资金总量的比例。个股集中度的高低反映了组合分散化的程度，而组合分散化减少了其所承担的非系统性风险。本研究中一般模型在 SQP 算法下的计算结果见表 3.4，流动性惩罚约束模型在 SQP 算法下的计算结果见表 3.9，流动性惩罚约束模型在 GA 算法下的计算结果见表 4.3，相比较而言，抽样复制的个股集中度较高（在 0.051449 至 0.092645 之间，平均值为 0.006667），

优化复制其次 (在 0.048381 至 0.081492 之间, 平均值为 0.005728), 而完全复制最分散 (0.000846 至 0.047413 之间, 平均值为 0.005556)。不过无论哪种方法, 组合内个股的持股集中度都不高, 也没有一只股票持有比例突破了模型中设定的 0.1 上限, 即个股占组合总市值的比例都在可以接受的范围之内。抽样复制中由于考虑了行业的分散性和代表性, 因此尽管个股集中度略高一点, 也不影响组合的分散化。

8. 流动性惩罚约束模型达到了预期的效果, 即在没有严重恶化组合追踪误差的情况下, 使得模型优化计算的结果能把更多的资金配置在比较有流动性的资产上。

本研究的一个重要的改进就是在常用的优化模型中加入了基于市场冲击成本考虑的流动性惩罚约束, 一个方面是试图对可能的市场冲击成本进行估计, 更重要的是让模型优化计算的结果能把更多的资金配置在比较有流动性的资产上, 提高组合的总体流动性, 进而就间接减少可能的市场冲击成本。尽管加入流动性惩罚约束之后, 组合的追踪误差有了一定的变化 (比较表 3.6、表 3.11 与表 4.5), 但组合的总体流动性明显提高了。流动性惩罚约束模型在 SQP 算法下的计算结果见图 3.10 与图 3.11, 流动性惩罚约束模型在 GA 算法下的计算结果见图 4.5 与图 4.6, 资金在不同流动性资产上的配置变化如图中的箭头所示。可见, 流动性惩罚约束模型达到了预期的效果, 即在没有严重恶化组合追踪误差的情况下, 使得模型优化计算的结果能把更多的资金配置在比较有流动性的资产上。

9. 尽管遗传算法 (GA) 在序贯二次规划法 (SQP) 计算结果的基础上都有所改进, 但改进的幅度非常小, 而且并没有比单纯应用 SQP 求解的结果好。可见 SQP 算法已经比较好地找到了模型的最优解, GA 算法能够改进的余地很小。

由于 GA 算法通常情况下比 SQP 算法较有可能找到全局最优解, 而且通过不同算法的应用, 可以比较其在优化求解, 尤其是本课题的优化求解中的具体应用和优化效果上有何不同, 因此本研究在 SQP 算法的基础上, 又通过 GA 算法的应用来探讨改进指数组合优化效果的可行性。表 4.1 列出了优化复制和抽样复制下样本内 SQP 计算的最小追踪误差和 GA 改进的情况, 可以看出, GA 在 SQP 计算结果的基础上都有所改进, 表现在 GA 在 SQP 计算结果的基础上进一步降低了样本内追踪误差, 不过改进的幅度非常小。通过表 3.11 与表 4.5 的比较中可以看出, SQP/GA (即 GA 在 SQP 计算结果的基础上的进一步优化) 与单纯用 SQP 计算的结果差异不大, 无论是样本内还是样本外, 有 SQP/GA 计算的结果优于 SQP 的情况, 也有劣于 SQP 的情况。而且无论是优于还是劣于, 两者之间的差距都不大。可见, SQP 算法已经比较好地找到了模型的最优解, GA 算法能够改进的余地很小。

10. 本研究中优化后的指数组合的追踪效果比天同 180 指数基金、华安 180 指数增强型基金的实际追踪效果要好, 这从侧面支持了本研究具有一定的实际应用价值。

尽管这两只基金只是一种“准指数基金”, 而且华安 180 实际上还是一种指数增强

型基金，其投资目标与本课题设定的追踪指数（track the index）的目标不同，因此严格意义上讲，这两只基金与本研究中所要构造的指数组合之间并没有准确的可比性。不过有一点应该肯定，本课题优化的结果从追踪误差上看应该比天同 180 和华安 180 好，如果这个目标没有达到，则本课题研究结果的实用价值就值得质疑。从评估结果的比较中可以看出，本研究中优化后的指数组合的追踪效果比天同 180 指数基金、华安 180 指数增强型基金的实际追踪效果要好。从表 5.2 中可以看出，无论是哪个时间段，优化复制和抽样复制两者的样本外追踪误差都远小于天同 180 指数基金和华安 180 指数增强型基金的实际追踪误差，多数情况下优化复制、抽样复制两者的追踪误差仅是天同 180 指数基金和华安 180 指数增强型基金实际追踪误差的一半。

主要参考文献

- 1 . Albert S. Neubert, “*Indexing for Maximum Investment Results*”, Glenlake Publishing Company Ltd Press, 1998, pp.1-357
- 2 . C.J. Adcock and N. Meade, “*A simple algorithm to incorporate transaction costs in quadratic optimization*”, European Journal of Operational Research, 1994, pp.85-94
- 3 . David D. Yao, Shuzhong Zhang, and Xun Yu Zhou, “*Tracking a Financial Benchmark Using a Few Assets*”, Working Paper, July 2003, pp.6-30
- 4 . David M. Walsh, Kathleen D. Walsh, John P. Evans, “*Assessing estimation error in a tracking error variance minimization framework*”, Pacific-Basin Finance Journal, 1998, pp.175-192
- 5 . Dirk Eddelbuttel and Marseilles, “*A Hybrid Genetic Algorithm for Passive Management*”, Second Conference Computing in Economics and Finance, Society of Computational Economics , Geneva, Switzerland, 26-28 June 1996, pp.1-21
- 6 . Francesco Corielli and Massimiliano Marcellino, “*Factor Based Index Tracking*”, Working Paper, February 2002, pp.6-41
- 7 . G. Bamberg and N. Wagner, “*Equity index replication with standard and robust regression estimators*”, OR Spektrum, 2000, pp.525-543
- 8 . J.E. Beasley, N. Meade, and T.-J. Chang, “*An evolutionary heuristic for the index tracking problem*”, Working Paper, October 2001, pp.2-34
- 9 . J. Shapcott, “*Index Tracking : Genetic Algorithms for Investment Portfolio Selection*”, Working Paper, September 1992, pp.1-16
- 10 . M.A.H Dempster, “*Dynamic index tracking*”, Citicorp/Salomon Smith Barney, PowerPoint, 7 October, 1999, pp.1-14
- 11 . Markus Rudolf, Hans-Jurgen Wolter, and H. Zimmermann, “*A linear model for tracking error minimization*”, Journal of Banking & Finance, 1999, pp.85-103
- 12 . M. Gilli and E. K llezi, “*Threshold accepting for index tracking*”, Working paper available from the first author at Department of Econometrics, University of Geneva, 1211 Geneva 4, Switzerland, 2001
- 13 . N. Meade and G.R. Salkin, “*Developing and maintaining an equity index fund*”, Journal of the Operational Research Society, 1990, 599-607
- 14 . Paolo Dai Pra, Wolfgang J. Runggaldier, and Marco Tolotti, “*Pathwise optimality for benchmark tracking*”, Working Paper, Oct 31, 2002, pp.1-23
- 15 . R. Roll, “*A mean/variance analysis of tracking error*”, The Journal of Portfolio Management, Summer 1992, pp.13-22
- 16 . Soosung Hwang and Stephen E. Satchell, “*Tracking Error: Ex-Ante versus Ex-Post Measures*”, Working Paper, 2001, pp.1-10
- 17 . 陈春锋、陈伟忠, “*ETFs 的发展与创新—国际经验及对我国的启示*”, 载于《山西财经大学学报》, 2003 年第 3 期, pp.65-69
- 18 . 陈春锋、陈伟忠, “*NASDAQ100 指数跟踪股的产品与运作解析*”, 载于《山西财经大学学报》, 2003 年第 6 期, pp.88-92
- 19 . 陈春锋、陈伟忠, “*全球交易所交易基金的发展与创新*”, 载于《外国经济与管理》, 2003 年第 8 期, pp.33-38
- 20 . 陈春锋、陈伟忠, “*从证券价格指数的演变中看全球金融市场的创新方向*”, 载于

《外国经济与管理》，2003 年第 10 期，pp.22-28

21. 陈春锋、陈伟忠，“指数复制方法与算法模型综述”，拟载于《同济大学学报（自然科学版）》2004 年第 11 期，2004 年 3 月录用
22. 陈春锋、陈伟忠，“积极指数化：一种全新的投资模式”，拟载于《证券市场导报》2004 年第 9 期，2004 年 6 月录用
23. 顾娟、刘文红，“指数选择的评价模型”，广发证券/深圳证券交易所研究报告，2003 年，pp.1-18
24. (台湾)谢文良、李进生、谢素娟，“指数模拟策略与参数选择”，《证券市场发展》，2002 年，pp.81-108
25. 杨朝军、李迅雷，“上海证券市场日内流动性与波动性研究”，上证联合研究计划第二期课题研究报告，上海交大 - 国泰君安课题组 2002 年，pp.1-42
26. 王霞，“海外 ETFs 发展与实践研究”，深圳证券交易所研究报告，2002 年，pp1-81
27. 上海证券交易所研究中心，“上证研究指数专辑”，第 2 卷，复旦大学出版社，2003 年，pp.1-266

附录

附表 1 抽样复制中剔除的个股及行业归属

(从 180 只股票中剔除 30 只即抽样 150 只)

代 码	行业归属	综合排名	代 码	行业归属	综合排名	代 码	行业归属	综合排名	代 码	行业归属	综合排名	代 码	行业归属	综合排名
创设问题 (2002/07/01)														
600003	工业	152	600161	健康护理	164	600632	可选消费	157.5	600096	原材料	172	600722	原材料	173
600058	工业	159	600133	金融	155.5	600723	可选消费	163	600123	原材料	177	600746	原材料	179
600368	工业	171	600701	金融	174	600778	可选消费	153.5	600126	原材料	167	600108	主要消费	157.5
600798	工业	178	600872	金融	150.5	600857	可选消费	180	600205	原材料	169	600195	主要消费	175.5
600896	工业	175.5	600156	可选消费	165	600861	可选消费	168	600223	原材料	160.5	600779	主要消费	160.5
600168	公用事业	170	600630	可选消费	155.5	600288	信息技术	162	600399	原材料	166	600887	主要消费	153.5
第一次调整 (2003/01/02)														
600003	工业	158.5	600668	健康护理	171	600158	可选消费	169.5	600657	信息技术	163.5	600307	原材料	165.5
600263	工业	167	600064	金融	163.5	600630	可选消费	168	600096	原材料	172	600395	原材料	176
600528	工业	160	600872	金融	169.5	600632	可选消费	152	600123	原材料	180	600727	原材料	154
600874	工业	147.5	600122	可选消费	178	600723	可选消费	177	600126	原材料	175	600881	原材料	155
600267	健康护理	151	600151	可选消费	156.5	600757	可选消费	162	600205	原材料	179	600108	主要消费	158.5
600332	健康护理	165.5	600156	可选消费	174	600778	可选消费	173	600282	原材料	156.5	600315	主要消费	161
第二次调整 (2003/07/01)														
600086	工业	157.5	600007	金融	176	600631	可选消费	156	600330	信息技术	150.5	600309	原材料	150.5
600263	工业	168	600052	金融	163	600632	可选消费	180	600103	原材料	152	600319	原材料	153.5
600662	工业	157.5	600071	可选消费	163	600757	可选消费	160.5	600111	原材料	172.5	600108	主要消费	175
600874	工业	169	600151	可选消费	165.5	600778	可选消费	178	600187	原材料	172.5	600127	主要消费	177
600129	健康护理	159	600158	可选消费	170	600838	可选消费	174	600223	原材料	153.5	600275	主要消费	163
600181	健康护理	179	600626	可选消费	167	600854	可选消费	165.5	600299	原材料	160.5	600779	主要消费	171
第三次调整 (2004/01/02)														
600345	电讯服务	153	600741	工业	160	600007	金融	180	600372	可选消费	174	600854	可选消费	165.5
600640	电讯服务	158.5	600820	工业	148	600322	金融	171	600555	可选消费	157	600498	信息技术	145
600776	电讯服务	163	600129	健康护理	170	600638	金融	172	600626	可选消费	167.5	600223	原材料	169
600086	工业	149	600488	健康护理	167.5	600639	金融	174	600631	可选消费	156	600127	主要消费	176
600266	工业	158.5	600535	健康护理	165.5	600823	金融	177.5	600677	可选消费	154.5	600180	主要消费	161.5
600377	工业	161.5	600866	健康护理	174	600158	可选消费	177.5	600754	可选消费	179	600779	主要消费	164

附表 2 优化复制中实际持有的个股手数

(流动性惩罚约束模型 GA 求解下的结果，限于篇幅仅以此为例)

代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数
创设问题 (2002/07/01)															
600000	0	600071	10224	600138	0	600221	132	600377	6130	600637	0	600717	0	600835	0
600001	0	600073	8448	600151	0	600223	0	600380	10509	600638	65	600718	2201	600839	66527
600002	26681	600076	8288	600156	3218	600256	10952	600383	0	600639	3351	600722	0	600849	115
600003	0	600085	4991	600158	4364	600258	0	600386	7003	600640	3333	600723	11478	600851	5799
600005	19625	600086	0	600161	3984	600263	0	600395	0	600641	373	600727	17	600854	503

上证联合研究计划课题研究报告

600006	872	600088	2792	600165	5568	600266	981	600399	0	600642	35271	600741	17894	600857	14
600007	16248	600096	0	600168	0	600267	0	600498	8483	600643	9651	600744	20	600861	9
600008	608	600098	20131	600170	6905	600270	0	600500	0	600649	29003	600746	634	600863	0
600009	11493	600100	14986	600171	4816	600282	3685	600508	3041	600651	18997	600748	1334	600866	4821
600010	4454	600102	815	600177	14	600288	0	600519	6037	600652	7163	600770	4874	600868	145
600011	9198	600103	98	600180	0	600299	0	600528	0	600654	47	600776	5173	600871	0
600016	24339	600104	74992	600181	0	600307	18671	600548	0	600657	709	600778	0	600872	0
600018	211	600108	0	600183	933	600309	19	600555	0	600662	0	600779	7898	600874	0
600019	45946	600111	0	600186	0	600313	0	600569	14330	600663	41	600784	0	600881	239
600028	35309	600115	14157	600188	11384	600315	0	600588	475	600664	14206	600795	0	600887	7562
600033	9810	600121	65	600190	0	600319	0	600600	0	600665	19116	600798	0	600890	25
600037	4537	600122	0	600195	3	600322	4315	600602	37	600668	5830	600805	39588	600894	6067
600052	0	600123	3	600196	0	600330	0	600624	6342	600675	9102	600808	63049	600895	14614
600058	594	600126	0	600198	683	600332	321	600626	0	600677	0	600811	64	600896	0
600060	175	600127	0	600205	59	600345	2082	600630	26	600688	19056	600812	63499		
600062	12	600129	519	600207	0	600362	26574	600631	12631	600690	35531	600820	385		
600064	201	600133	0	600210	0	600368	0	600632	499	600701	13	600821	2070		
600068	62	600135	244	600217	0	600372	0	600635	34	600705	788	600832	21		

第一次调整 (2003/01/02)

600000	859	600062	35	600129	545	600226	159	600380	11348	600626	284	600675	9069	600820	632
600001	135	600064	267	600135	545	600228	429	600383	365	600630	448	600677	195	600832	905
600002	27593	600068	211	600138	962	600256	11473	600386	9015	600631	12883	600688	22746	600835	99
600003	355	600071	10345	600151	332	600258	552	600395	691	600632	3232	600690	36033	600838	842
600005	19986	600073	9673	600156	3249	600263	280	600398	223	600635	285	600705	776	600839	66512
600006	1419	600076	9118	600158	4672	600266	2357	600428	373	600637	43	600717	211	600849	100
600007	16283	600085	6225	600165	6085	600267	903	600456	234	600638	378	600718	2350	600851	6055
600008	1006	600086	40	600170	7658	600270	143	600498	9039	600639	3822	600723	12808	600854	500
600009	11588	600088	3454	600171	4929	600275	1137	600500	109	600640	4093	600727	23	600863	129
600010	4600	600096	535	600177	78	600282	3876	600508	3049	600641	1471	600741	20281	600866	5425
600011	9403	600098	20627	600180	40	600299	816	600519	6192	600642	36149	600744	448	600868	1417
600016	28273	600100	15339	600181	954	600307	18637	600528	719	600643	9887	600748	2557	600871	282
600018	482	600102	897	600183	1216	600309	529	600548	416	600649	29217	600757	1442	600872	381
600019	31382	600103	523	600186	120	600315	126	600555	3445	600651	20039	600770	4958	600874	97
600026	1046	600104	43502	600188	11980	600319	387	600569	15010	600652	7325	600776	5346	600881	397
600028	29063	600108	35	600196	122	600322	4688	600578	1215	600654	369	600778	1293	600887	10124
600033	10208	600111	104	600198	1591	600330	654	600585	975	600657	1066	600779	8532	600890	658
600036	5	600115	14763	600205	78	600332	553	600588	734	600660	236	600795	373	600894	6409
600037	4548	600121	145	600207	89	600345	2825	600598	2831	600662	1162	600805	39390	600895	17006
600038	485	600122	127	600210	265	600350	3090	600600	370	600663	517	600808	47631		
600052	139	600123	97	600217	1	600362	21346	600601	333	600664	14711	600811	452		
600058	654	600126	278	600221	268	600372	57	600602	464	600665	20074	600812	31524		
600060	350	600127	34	600223	72	600377	6651	600624	6574	600668	6138	600816	256		

第二次调整 (2003/07/01)

600000	1392	600052	448	600127	1215	600221	8163	600372	55	600601	2112	600668	6295	600812	31801
--------	------	--------	-----	--------	------	--------	------	--------	----	--------	------	--------	------	--------	-------

指数组合优化方法、模型与应用 - 基于上证 180 指数的实证研究

600001	1497	600058	793	600129	523	600223	163	600377	6776	600602	621	600675	8420	600816	242
600002	17495	600060	399	600135	860	600226	199	600380	11325	600624	6824	600677	1054	600820	625
600003	3327	600062	119	600138	2112	600228	999	600383	496	600626	372	600688	22976	600832	1099
600005	15682	600071	10392	600151	655	600236	232	600398	346	600631	12806	600690	38680	600835	287
600006	1553	600072	2815	600158	5540	600253	3763	600428	681	600632	3403	600705	784	600838	799
600007	17313	600073	10251	600165	6085	600256	11712	600456	52	600635	560	600717	379	600839	57524
600008	994	600076	9151	600166	796	600258	1563	600498	8263	600637	411	600718	3032	600849	1
600009	12311	600085	6976	600170	8566	600263	1716	600500	86	600638	755	600724	151	600851	5823
600010	4042	600086	384	600171	7368	600266	2577	600508	3031	600639	4027	600737	517	600854	421
600011	10582	600088	3046	600177	66	600270	62	600519	8422	600640	4940	600741	20899	600863	137
600012	4230	600091	450	600180	1557	600275	1950	600535	0	600641	2229	600744	468	600866	6088
600016	25555	600098	21545	600181	2115	600282	4751	600548	381	600642	37297	600748	3061	600868	2123
600018	949	600100	16032	600183	1334	600299	442	600555	3741	600643	10608	600754	121	600871	465
600019	18289	600102	1272	600186	94	600307	18600	600569	15117	600649	30504	600757	2942	600874	270
600026	4455	600103	803	600187	836	600309	512	600578	1687	600651	20055	600770	5508	600887	10265
600028	20724	600104	33379	600188	12600	600319	563	600581	1484	600652	7611	600776	4913	600890	1076
600030	319	600108	28	600196	409	600322	4517	600585	1247	600654	400	600778	450	600894	5828
600033	10602	600111	887	600198	2576	600330	917	600588	369	600660	382	600779	8412	600895	16962
600036	1146	600115	11450	600205	1056	600345	2795	600591	116	600662	1106	600795	669		
600037	3877	600120	910	600207	41	600350	3402	600597	91	600663	533	600805	28742		
600038	2324	600121	272	600210	273	600357	91	600598	3047	600664	15022	600808	29966		
600050	6662	600126	352	600217	272	600362	21450	600600	520	600665	20011	600811	1313		
第三次调整 (2004/01/02)															
600000	1414	600037	2265	600135	929	600228	980	600398	326	600600	166	600677	1087	600816	24
600001	1097	600038	2621	600138	2508	600231	1	600418	55	600601	1815	600688	19247	600820	783
600002	18361	600050	82377	600151	782	600236	559	600428	802	600602	362	600690	38443	600823	56
600004	120	600058	781	600158	6035	600253	3794	600456	342	600609	0	600705	896	600832	2214
600005	16509	600060	631	600165	4063	600256	11321	600481	6	600626	290	600710	4	600835	447
600006	1365	600062	239	600166	850	600258	1278	600488	113	600631	12643	600717	1090	600838	791
600007	17015	600072	2730	600170	8668	600266	2576	600498	8284	600635	439	600718	3323	600839	45409
600008	881	600073	10332	600171	7257	600270	40	600500	145	600637	266	600724	1011	600849	70
600009	12812	600076	10857	600177	596	600271	55	600508	3458	600638	708	600737	351	600851	4563
600010	3273	600085	6956	600180	928	600275	1437	600511	442	600639	3576	600741	21836	600854	420
600011	10328	600086	205	600183	1256	600282	6501	600519	8460	600640	3876	600744	625	600863	2
600012	3606	600088	2819	600186	508	600296	0	600535	33	600641	1465	600747	0	600866	5915
600015	230	600091	461	600187	701	600307	9733	600548	426	600642	38129	600748	2224	600868	2071
600016	25585	600098	23316	600188	12981	600309	523	600555	3429	600643	10920	600754	47	600871	455
600018	687	600100	13663	600196	712	600322	4814	600569	10470	600649	30496	600757	2894	600887	10319
600019	26308	600102	1160	600205	693	600345	2743	600578	1315	600651	22484	600770	4739	600890	1058
600026	4175	600104	33351	600207	193	600350	1713	600581	2359	600652	9018	600776	3765	600894	5695
600028	28420	600115	11393	600210	717	600357	107	600583	4	600654	902	600779	7736	600895	16962
600029	16301	600120	901	600215	4	600362	20941	600585	1252	600660	806	600795	799	600900	6663
600030	347	600121	414	600217	77	600372	281	600588	662	600663	599	600805	16640		
600031	3	600126	388	600221	6919	600377	6413	600591	9	600664	14753	600808	18522		

600033	10315	600127	971	600223	264	600380	9385	600597	90	600665	19934	600811	812		
600036	1267	600129	510	600226	344	600383	429	600598	2599	600675	10427	600812	28944		

附表 3 抽样复制中实际持有的个股手数

(流动性惩罚约束模型 GA 求解下的结果, 限于篇幅仅以此为例)

代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数
创设问题 (2002/07/01)															
600000	1593	600064	50	600135	352	600221	50	600372	50	600631	14104	600675	8961	600820	50
600001	50	600068	130	600138	50	600256	11879	600377	9129	600635	50	600677	50	600821	605
600002	26440	600071	16933	600151	50	600258	312	600380	10323	600637	50	600688	7188	600832	50
600005	16347	600073	9096	600158	7129	600263	50	600383	50	600638	50	600690	34254	600835	96
600006	1589	600076	9399	600165	6402	600266	50	600386	7280	600639	6087	600705	50	600839	59197
600007	17326	600085	4099	600170	9948	600267	50	600395	50	600640	1049	600717	50	600849	50
600008	422	600086	50	600171	4427	600270	50	600498	9087	600641	259	600718	1089	600851	5713
600009	855	600088	5000	600177	177	600282	4112	600500	50	600642	36550	600727	285	600854	50
600010	8407	600098	20753	600180	50	600299	50	600508	7055	600643	6324	600741	17170	600863	50
600011	14591	600100	12627	600181	50	600307	17370	600519	7136	600649	23197	600744	50	600866	7844
600016	23753	600102	2196	600183	3919	600309	899	600528	50	600651	23219	600748	1577	600868	50
600018	50	600103	120	600186	50	600313	50	600548	50	600652	7858	600770	5041	600871	1687
600019	42594	600104	73249	600188	12541	600315	50	600555	50	600654	50	600776	5620	600874	50
600028	31888	600111	669	600190	50	600319	50	600569	19275	600657	274	600784	50	600881	2301
600033	14497	600115	4904	600196	512	600322	6857	600588	1415	600662	50	600795	50	600890	94
600037	4771	600121	91	600198	221	600330	50	600600	50	600663	50	600805	38398	600894	7871
600052	50	600122	50	600207	50	600332	50	600602	421	600664	13350	600808	56625	600895	20218
600060	50	600127	50	600210	50	600345	2910	600624	8337	600665	17769	600811	50		
600062	50	600129	825	600217	50	600362	35409	600626	50	600668	4336	600812	55152		
第一次调整 (2003/01/02)															
600000	2109	600052	157	600127	404	600223	548	600383	150	600624	8412	600665	17770	600816	54
600001	81	600058	282	600129	1469	600226	127	600386	8479	600626	309	600675	9091	600820	1364
600002	26860	600060	66	600135	1528	600228	231	600398	115	600631	14646	600677	129	600832	252
600005	14356	600062	751	600138	319	600256	13269	600428	87	600635	309	600688	7487	600835	72
600006	2462	600068	163	600165	6598	600258	2418	600456	174	600637	154	600690	34261	600838	53
600007	17496	600071	17184	600170	10212	600266	54	600498	9138	600638	1036	600705	79	600839	59049
600008	653	600073	9099	600171	4725	600270	794	600500	265	600639	7188	600717	60	600849	148
600009	849	600076	10017	600177	547	600275	51	600508	7875	600640	1154	600718	1437	600851	6436
600010	9426	600085	4220	600180	370	600299	114	600519	7256	600641	1370	600741	17416	600854	245
600011	15857	600086	143	600181	1003	600309	1287	600548	447	600642	37823	600744	270	600863	769
600016	24881	600088	5037	600183	4161	600319	238	600555	252	600643	6696	600748	1632	600866	9483
600018	254	600098	20945	600186	772	600322	6942	600569	19833	600649	24562	600770	5040	600868	54
600019	43013	600100	12918	600188	12781	600330	2150	600578	858	600651	23585	600776	5785	600871	1946
600026	5214	600102	2881	600196	639	600345	3391	600585	184	600652	7945	600779	82	600887	315
600028	32219	600103	1328	600198	448	600350	93	600588	1615	600654	375	600795	388	600890	5480
600033	18189	600104	59610	600207	1856	600362	23561	600598	203	600660	391	600805	38711	600894	12339
600036	105	600111	1202	600210	341	600372	985	600600	80	600662	90	600808	32964	600895	20218

指数组合优化方法、模型与应用 - 基于上证 180 指数的实证研究

600037	4774	600115	13169	600217	437	600377	9143	600601	118	600663	1648	600811	356		
600038	228	600121	82	600221	432	600380	10928	600602	564	600664	13478	600812	39248		
第二次调整 (2003/07/01)															
600000	2133	600037	4631	600126	126	600226	227	600398	872	600601	74	600668	50	600811	675
600001	519	600038	304	600135	1531	600228	190	600428	1040	600602	591	600675	9150	600812	24522
600002	26715	600050	109204	600138	282	600236	50	600456	749	600624	9138	600677	288	600816	345
600003	17269	600058	206	600165	6876	600253	51	600498	9579	600635	3312	600688	8437	600820	2040
600005	3742	600060	131	600166	50	600256	13338	600500	291	600637	223	600690	33897	600832	706
600006	17305	600062	800	600170	12089	600258	2425	600508	7871	600638	1232	600705	135	600835	617
600008	635	600072	136	600171	4801	600266	58	600519	7483	600639	7221	600717	271	600839	48468
600009	2249	600073	9352	600177	747	600270	577	600535	51	600640	1980	600718	1448	600849	140
600010	10019	600076	10281	600180	235	600282	50	600548	603	600641	1532	600724	52	600851	6372
600011	16418	600085	5336	600183	4673	600307	50	600555	311	600642	37708	600737	50	600863	762
600012	58	600088	5002	600186	845	600322	8393	600569	19895	600643	6696	600741	17434	600866	9725
600016	25205	600091	51	600188	12528	600345	3603	600578	1099	600649	24732	600744	758	600868	1555
600018	299	600098	21034	600196	669	600350	658	600581	50	600651	23841	600748	1234	600871	2016
600019	35226	600100	13048	600198	223	600357	50	600585	391	600652	7877	600754	50	600887	811
600026	7510	600102	3132	600205	52	600362	23774	600588	1597	600654	350	600770	5119	600890	5108
600028	35445	600104	52078	600207	2018	600372	1337	600591	51	600660	412	600776	7133	600894	12127
600030	50	600115	14519	600210	341	600377	9451	600597	50	600663	1991	600795	224	600895	20876
600033	19859	600120	52	600217	621	600380	11314	600598	708	600664	14074	600805	23513		
600036	663	600121	992	600221	497	600383	1213	600600	607	600665	18286	600808	32815		
第三次调整 (2004/01/02)															
600000	1866	600031	51	600115	13659	600210	1420	600357	517	600588	3237	600664	15599	600811	516
600001	689	600033	16801	600120	76	600215	120	600362	23631	600591	155	600665	12054	600812	24717
600002	27082	600036	368	600121	1198	600217	578	600380	13337	600597	131	600675	9859	600816	1164
600004	100	600037	5247	600126	216	600221	871	600383	1082	600598	1282	600688	8510	600832	869
600005	3355	600038	665	600135	879	600226	235	600398	895	600600	634	600690	34556	600835	974
600006	19703	600050	109202	600138	421	600228	118	600418	464	600601	936	600705	483	600838	217
600008	783	600058	512	600151	55	600231	50	600428	1064	600602	2480	600710	548	600839	33526
600009	2499	600060	135	600165	7687	600236	872	600456	465	600609	53	600717	262	600849	151
600010	10278	600062	1424	600166	85	600253	232	600481	52	600635	2641	600718	1920	600851	8593
600011	17975	600072	198	600170	11302	600256	13534	600500	463	600637	67	600724	699	600863	1329
600012	73	600073	9426	600171	4812	600258	2192	600508	8296	600641	1912	600737	986	600868	2577
600015	3124	600076	10385	600177	741	600270	3163	600511	51	600642	41145	600744	2382	600871	3020
600016	25012	600085	7086	600183	4848	600271	50	600519	8113	600643	6883	600747	52	600887	1666
600018	751	600088	5132	600186	773	600275	51	600548	852	600649	24890	600748	1347	600890	5137
600019	32075	600091	63	600187	71	600282	81	600569	20283	600651	24736	600757	53	600894	13036
600026	7496	600098	22028	600188	13447	600296	514	600578	1660	600652	8518	600770	6433	600895	22391
600028	35690	600100	15299	600196	408	600307	488	600581	234	600654	481	600795	591	600900	8179
600029	10772	600102	4803	600205	727	600309	50	600583	50	600660	3200	600805	19205		
600030	264	600104	52280	600207	2278	600350	417	600585	555	600663	2800	600808	24916		

附表 4 完全复制中实际持有的个股手数
(非线性方程组求解的结果)

代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数	代 码	手 数
创设问题 (2002/07/01)															
600000	11026	600071	2716	600138	3054	600221	5981	600377	3431	600637	5115	600717	6038	600835	4853
600001	13603	600073	2965	600151	4281	600223	3680	600380	2791	600638	5135	600718	3219	600839	24755
600002	8922	600076	2896	600156	4016	600256	4128	600383	2471	600639	3724	600722	2892	600849	5064
600003	8326	600085	3036	600158	1741	600258	2117	600386	2306	600640	1669	600723	4686	600851	3834
600005	9565	600086	3224	600161	1318	600263	2801	600395	3398	600641	4098	600727	4339	600854	4753
600006	6863	600088	2166	600165	2299	600266	4118	600399	3569	600642	12306	600741	8305	600857	2199
600007	3660	600096	2527	600168	3028	600267	1713	600498	2814	600643	7377	600744	8140	600861	4641
600008	7549	600098	5732	600170	5485	600270	2511	600500	3410	600649	12932	600746	2103	600863	4532
600009	12925	600100	6573	600171	5168	600282	3459	600508	2756	600651	9703	600748	4032	600866	4470
600010	8579	600102	3986	600177	5198	600288	1281	600519	1716	600652	6854	600770	2471	600868	20380
600011	5719	600103	11310	600180	2297	600299	2196	600528	2814	600654	8053	600776	3280	600871	4575
600016	17753	600104	23059	600181	3982	600307	4996	600548	6557	600657	2458	600778	4275	600872	6177
600018	6192	600108	7149	600183	4491	600309	3514	600555	2227	600662	4242	600779	4470	600874	4530
600019	57246	600111	3694	600186	7778	600313	2308	600569	9234	600663	6214	600784	4131	600881	6521
600028	64054	600115	6863	600188	4118	600315	1853	600588	686	600664	8500	600795	5661	600887	2013
600033	4701	600121	5559	600190	5353	600319	2166	600600	5977	600665	4942	600798	3513	600890	5546
600037	1853	600122	3651	600195	3569	600322	2908	600602	9544	600668	3437	600805	7487	600894	4709
600052	6638	600123	3397	600196	3494	600330	1575	600624	4219	600675	7978	600808	21606	600895	8557
600058	2910	600126	4429	600198	4017	600332	2704	600626	3869	600677	4477	600811	9193	600896	3629
600060	5648	600127	4947	600205	3843	600345	1359	600630	4860	600688	22282	600812	13376		
600062	5045	600129	1734	600207	5033	600362	6898	600631	5333	600690	10948	600820	6147		
600064	3936	600133	2522	600210	5000	600368	2022	600632	2900	600701	4448	600821	3346		
600068	8073	600135	3129	600217	3779	600372	2814	600635	6536	600705	8965	600832	4722		
第一次调整 (2003/01/02)															
600000	15559	600062	4744	600129	1624	600226	2325	600380	2623	600626	3638	600675	7504	600820	6360
600001	12797	600064	3702	600135	2942	600228	2475	600383	2320	600630	4571	600677	4211	600832	4440
600002	8393	600068	7594	600138	2871	600256	4657	600386	2167	600631	5016	600688	20963	600835	5021
600003	7832	600071	2554	600151	4026	600258	1989	600395	3195	600632	2727	600690	10298	600838	4313
600005	8998	600073	2787	600156	3777	600263	2633	600398	2297	600635	8227	600705	8433	600839	23288
600006	6453	600076	2723	600158	1636	600266	3872	600428	3097	600637	5291	600717	5679	600849	4762
600007	3442	600085	2852	600165	2161	600267	1609	600456	1289	600638	4829	600718	3026	600851	4325
600008	7100	600086	3031	600170	6191	600270	2360	600498	2644	600639	3502	600723	4408	600854	4470
600009	12160	600088	2036	600171	4859	600275	2051	600500	4810	600640	1568	600727	4080	600863	4263
600010	8070	600096	2375	600177	4889	600282	3253	600508	2590	600641	3854	600741	7813	600866	4204
600011	5378	600098	5390	600180	2158	600299	2064	600519	1770	600642	11575	600744	7657	600868	19173
600016	16700	600100	6180	600181	3745	600307	4699	600528	2645	600643	6939	600748	3792	600871	4304
600018	5823	600102	5954	600183	5491	600309	3303	600548	6167	600649	12166	600757	3047	600872	5811
600019	53858	600103	10640	600186	9512	600315	1741	600555	2093	600651	9127	600770	2321	600874	4260
600026	8737	600104	21693	600188	3872	600319	2035	600569	8686	600652	6447	600776	3084	600881	6134
600028	60263	600108	6725	600196	3286	600322	2734	600578	2466	600654	8332	600778	4021	600887	2945

指数组合优化方法、模型与应用 - 基于上证 180 指数的实证研究

600033	4421	600111	3473	600198	3776	600330	1479	600585	5309	600657	2311	600779	4205	600890	5213
600036	24563	600115	6456	600205	3614	600332	2542	600588	637	600660	3893	600795	9053	600894	4429
600037	2262	600121	5228	600207	4733	600345	1275	600598	9490	600662	3989	600805	7043	600895	8049
600038	2061	600122	3433	600210	4701	600350	14479	600600	5621	600663	5843	600808	20327		
600052	6244	600123	3195	600217	3552	600362	6488	600601	8032	600664	9593	600811	9513		
600058	2736	600126	4165	600221	5625	600372	2645	600602	8977	600665	4648	600812	12583		
600060	5312	600127	4653	600223	4615	600377	3226	600624	3968	600668	3232	600816	5863		
第二次调整 (2003/07/01)															
600000	22376	600052	5527	600127	4119	600221	4979	600372	2341	600601	7108	600668	3932	600812	11139
600001	11328	600058	3148	600129	1436	600223	4085	600377	2854	600602	7946	600675	6643	600816	5189
600002	7429	600060	4702	600135	2602	600226	2058	600380	2320	600624	3512	600677	3727	600820	5630
600003	6933	600062	4197	600138	2541	600228	2189	600383	2053	600626	3220	600688	18558	600832	3928
600005	9558	600071	2261	600151	3563	600236	3857	600398	2033	600631	4440	600690	9116	600835	5333
600006	5712	600072	3137	600158	1447	600253	2399	600428	2741	600632	2413	600705	7465	600838	3817
600007	3047	600073	2466	600165	1912	600256	4122	600456	1140	600635	7282	600717	5026	600839	20617
600008	6284	600076	2410	600166	2317	600258	1760	600498	2340	600637	4682	600718	2678	600849	6324
600009	10785	600085	2522	600170	5480	600263	2330	600500	4258	600638	4274	600724	3469	600851	3828
600010	7144	600086	2681	600171	4301	600266	3427	600508	2291	600639	3099	600737	6574	600854	3957
600011	4759	600088	1801	600177	4327	600270	2087	600519	1565	600640	1526	600741	6916	600863	7548
600012	6661	600091	4486	600180	1910	600275	1815	600535	1625	600641	3411	600744	6778	600866	3721
600016	19219	600098	4771	600181	3314	600282	2878	600548	5459	600642	10246	600748	3356	600868	17652
600018	10311	600100	5470	600183	4860	600299	1826	600555	1852	600643	6141	600754	2554	600871	3809
600019	47680	600102	5270	600186	8420	600307	4159	600569	7689	600649	10769	600757	2696	600874	3771
600026	7734	600103	9419	600187	1867	600309	2923	600578	2182	600651	8079	600770	2054	600887	2606
600028	53351	600104	19203	600188	3427	600319	1801	600581	3742	600652	7419	600776	2730	600890	4613
600030	9454	600108	5953	600196	2907	600322	2420	600585	4699	600654	7376	600778	3559	600894	3920
600033	4696	600111	3074	600198	3342	600330	1307	600588	558	600660	3444	600779	3722	600895	9263
600036	32618	600115	5715	600205	3197	600345	1128	600591	4119	600662	3530	600795	8013		
600037	2000	600120	4813	600207	4189	600350	12818	600597	3718	600663	5171	600805	6234		
600038	1822	600121	4628	600210	4160	600357	2171	600598	8401	600664	8491	600808	17995		
600050	112591	600126	3686	600217	3143	600362	5743	600600	4975	600665	4114	600811	8421		
第三次调整 (2004/01/02)															
600000	20066	600037	1794	600135	2334	600228	1967	600398	1824	600600	4461	600677	3343	600816	4654
600001	10159	600038	2213	600138	2279	600231	4473	600418	1568	600601	8288	600688	16641	600820	5049
600002	6662	600050	100966	600151	3196	600236	3459	600428	2458	600602	7125	600690	8175	600823	2018
600004	6832	600058	2823	600158	1298	600253	2151	600456	1022	600609	7466	600705	6694	600832	4933
600005	8571	600060	4217	600165	2144	600256	3697	600481	1566	600626	2887	600710	2817	600835	4782
600006	5123	600062	3762	600166	3118	600258	1579	600488	1719	600631	3982	600717	4507	600838	3423
600007	2732	600072	3095	600170	4914	600266	3073	600498	2099	600635	6531	600718	2402	600839	18488
600008	5635	600073	2212	600171	5230	600270	2808	600500	3818	600637	4199	600724	3111	600849	5671
600009	12685	600076	2161	600177	3880	600271	820	600508	2055	600638	3833	600737	8255	600851	3432
600010	6406	600085	2263	600180	1712	600275	2281	600511	907	600639	2780	600741	6202	600854	3549
600011	4267	600086	2405	600183	4358	600282	2581	600519	1543	600640	1369	600744	6078	600863	6769
600012	5973	600088	1615	600186	7551	600296	2013	600535	1458	600641	3059	600747	5030	600866	3337

600015	17939	600091	4023	600187	1676	600307	3729	600548	4896	600642	9187	600748	3010	600868	15830
600016	18581	600098	4278	600188	3072	600309	2620	600555	1661	600643	5507	600754	2291	600871	3416
600018	9245	600100	4905	600196	2607	600322	2170	600569	6895	600649	9657	600757	2419	600887	4676
600019	42756	600102	4726	600205	2865	600345	1013	600578	1957	600651	7245	600770	1842	600890	4136
600026	6935	600104	17220	600207	3757	600350	11494	600581	3356	600652	6654	600776	2449	600894	3515
600028	47842	600115	5125	600210	5477	600357	1946	600583	1875	600654	6615	600779	3338	600895	8307
600029	21870	600120	4317	600215	4277	600362	5150	600585	4213	600660	3418	600795	7185	600900	26844
600030	8478	600121	4150	600217	2818	600372	2099	600588	607	600663	4638	600805	5590		
600031	1223	600126	3305	600221	4466	600377	2560	600591	3693	600664	11424	600808	16137		
600033	4211	600127	3694	600223	3663	600380	3124	600597	3335	600665	3690	600811	7551		
600036	29251	600129	1289	600226	1846	600383	1843	600598	7533	600675	5957	600812	9989		

附表 5 优化复制下的追踪误差和相关系数的描述性统计值

(两类模型、两种算法下的结果比较，第三次调整的组合因时间过短而除外)

	一般模型 SQP 求解结果				流动性惩罚约束模型 SQP 求解结果				流动性惩罚约束模型 GA 求解结果			
	平均值	最大值	最小值	标准差	平均值	最大值	最小值	标准差	平均值	最大值	最小值	标准差
样本内 (in sample)												
TE^*	1.0255e-4	1.5772e-4	5.9280e-5	4.1060e-5	1.0459e-4	1.2973e-4	8.9518e-5	1.7893e-5	1.0228e-4	1.2268e-4	9.0200e-5	1.4506e-5
TE	1.0549e-4	1.5615e-4	6.6128e-5	3.7609e-5	1.0815e-4	1.2551e-4	9.5020e-5	1.2802e-5	1.0751e-4	1.2274e-4	9.6494e-5	1.1120e-5
TTE^*	0.0116	0.0173	0.0062	0.0045	0.0118	0.0143	0.0099	0.0018	0.0116	0.0135	0.0099	0.0015
TTE	0.0120	0.0172	0.0069	0.0042	0.0122	0.0138	0.0109	0.0012	0.0123	0.0141	0.0108	0.0014
ρ	0.9952	0.9995	0.9926	0.0031	0.9959	0.9987	0.9941	0.0020	0.9959	0.9987	0.9943	0.0020
样本外 (out of sample)												
5 日												
TE	6.6170e-4	7.5426e-4	5.2865e-4	9.6454e-5	6.1089e-4	6.8051e-4	4.9671e-4	8.1381e-5	6.1197e-4	6.8028e-4	5.3074e-4	6.1729e-5
TTE	0.0033	0.0038	0.0026	0.0005	0.0031	0.0034	0.0025	0.0004	0.0031	0.0034	0.0027	0.0003
ρ	0.9953	0.9993	0.9923	0.0030	0.9960	0.9986	0.9946	0.0019	0.9960	0.9986	0.9945	0.0019
10 日												
TE	4.2939e-4	4.7730e-4	3.9097e-4	3.5878e-5	4.5296e-4	4.9446e-4	3.9962e-4	3.9614e-5	4.7014e-4	4.9390e-4	4.3387e-4	2.6053e-5
TTE	0.0043	0.0048	0.0039	0.0004	0.0045	0.0049	0.0040	0.0004	0.0047	0.0049	0.0043	0.0003
ρ	0.9956	0.9992	0.9924	0.0028	0.9961	0.9984	0.9946	0.0016	0.9962	0.9985	0.9946	0.0017
30 日												
TE	2.3513e-4	2.7275e-4	2.0833e-4	2.7390e-5	2.3588e-4	2.7985e-4	2.0566e-4	3.1812e-5	2.3877e-4	2.7966e-4	2.1367e-4	2.9164e-5
TTE	0.0071	0.0082	0.0063	0.0008	0.0071	0.0084	0.0062	0.0009	0.0072	0.0084	0.0064	0.0009
ρ	0.9954	0.9988	0.9923	0.0027	0.9959	0.9979	0.9943	0.0015	0.9959	0.9980	0.9941	0.0016
60 日												
TE	1.6749e-4	1.7568e-4	1.6241e-4	5.8450e-6	1.6803e-4	1.7050e-4	1.6358e-4	3.1509e-6	1.6910e-4	1.7057e-4	1.6643e-4	1.8933e-6
TTE	0.0100	0.0105	0.0097	0.0003	0.0101	0.0102	0.0098	0.0002	0.0101	0.0102	0.0100	0.0001
ρ	0.9948	0.9985	0.9916	0.0028	0.9953	0.9977	0.9935	0.0018	0.9953	0.9977	0.9932	0.0019
截至下次指数调整日												
TE	1.3322e-4	1.6300e-4	1.0828e-4	2.2600e-5	1.2981e-4	1.5402e-4	1.0574e-4	1.9710e-5	1.2899e-4	1.4127e-4	1.0560e-4	1.6548e-5
TTE	0.0162	0.0187	0.0136	0.0021	0.0158	0.0177	0.0133	0.0019	0.0158	0.0178	0.0133	0.0019
ρ	0.9939	0.9979	0.9913	0.0029	0.9943	0.9973	0.9926	0.0021	0.9943	0.9973	0.9918	0.0023

*注释：*为优化计算未取整的结果，其余数据都为根据最小交易手数取整后计算的结果。

附表 6 抽样复制下的追踪误差和相关系数的描述性统计值

(两类模型、两种算法下的结果比较,第三次调整的组合因时间过短而除外)

	一般模型 SQP 求解结果				流动性惩罚约束模型 SQP 求解结果				流动性惩罚约束模型 GA 求解结果			
	平均值	最大值	最小值	标准差	平均值	最大值	最小值	标准差	平均值	最大值	最小值	标准差
样本内 (in sample)												
TE^*	1.1914e-4	1.8655e-4	6.5029e-5	5.0495e-5	1.2471e-4	1.5909e-4	9.8259e-5	2.5460e-5	1.2535e-4	1.6421e-4	9.5516e-5	2.8762e-5
TE	1.2367e-4	1.9053e-4	7.0576e-5	4.9929e-5	1.2740e-4	1.5459e-4	1.0177e-4	2.1592e-5	1.2861e-4	1.6111e-4	9.9206e-5	2.5367e-5
TTE^*	0.0137	0.0216	0.0068	0.0061	0.0140	0.0175	0.0123	0.0025	0.0141	0.0181	0.0119	0.0029
TTE	0.0139	0.0205	0.0074	0.0053	0.0143	0.0170	0.0127	0.0019	0.0144	0.0177	0.0124	0.0023
ρ	0.9933	0.9994	0.9888	0.0045	0.9946	0.9980	0.9922	0.0025	0.9943	0.9981	0.9915	0.0028
样本外 (out of sample)												
5 日												
TE	6.0514e-4	7.3595e-4	4.4264e-4	1.2182e-4	5.7416e-4	7.8909e-4	3.6638e-4	1.7264e-4	5.9790e-4	7.8931e-4	3.5195e-4	1.8267e-4
TTE	0.0030	0.0037	0.0022	0.0006	0.0028	0.0039	0.0018	0.0009	0.0030	0.0039	0.0018	0.0009
ρ	0.9935	0.9992	0.9886	0.0044	0.9948	0.9979	0.9920	0.0024	0.9945	0.9979	0.9913	0.0027
10 日												
TE	4.4596e-4	5.0858e-4	3.4436e-4	7.2487e-5	4.5318e-4	5.2404e-4	4.1671e-4	5.0110e-5	4.5772e-4	5.2491e-4	3.9703e-4	5.2409e-5
TTE	0.0044	0.0051	0.0034	0.0007	0.0045	0.0052	0.0042	0.0005	0.0046	0.0052	0.0040	0.0005
ρ	0.9939	0.9991	0.9889	0.0042	0.9950	0.9978	0.9920	0.0024	0.9947	0.9978	0.9912	0.0027
30 日												
TE	2.5483e-4	3.0127e-4	1.8573e-4	4.9818e-5	2.3253e-4	2.7570e-4	2.0470e-4	3.0947e-5	2.2517e-4	2.7583e-4	1.8867e-4	3.6964e-5
TTE	0.0076	0.0090	0.0056	0.0015	0.0070	0.0083	0.0061	0.0010	0.0068	0.0083	0.0057	0.0011
ρ	0.9937	0.9985	0.9891	0.0038	0.9948	0.9974	0.9918	0.0023	0.9947	0.9974	0.9913	0.0025
60 日												
TE	1.8038e-4	1.9232e-4	1.6989e-4	9.2142e-6	1.7241e-4	1.8026e-4	1.6723e-4	5.6446e-6	1.6489e-4	1.6727e-4	1.6219e-4	2.0862e-6
TTE	0.0108	0.0115	0.0102	0.0005	0.0103	0.0108	0.0100	0.0003	0.0099	0.0100	0.0097	0.0001
ρ	0.9931	0.9982	0.9883	0.0040	0.9941	0.9971	0.9907	0.0026	0.9940	0.9971	0.9906	0.0027
截至下次指数调整日												
TE	1.4565e-4	1.7010e-4	1.1953e-4	2.0679e-5	1.4043e-4	1.6875e-4	1.0753e-4	2.5202e-5	1.3129e-4	1.5944e-4	1.0756e-4	2.1409e-5
TTE	0.0178	0.0196	0.0151	0.0019	0.0171	0.0194	0.0135	0.0026	0.0160	0.0183	0.0136	0.0019
ρ	0.9922	0.9975	0.9874	0.0041	0.9928	0.9968	0.9897	0.0030	0.9932	0.9968	0.9905	0.0026

*注释: *为优化计算未取整的结果,其余数据都为根据最小交易手数取整后计算的结果。

附表 7 完全复制下的追踪误差和相关系数的描述性统计值

	平均值	最大值	最小值	标准差
样本内 (in sample)				
	-	-	-	-
样本外 (out of sample)				
5 日				
TE	1.1630e-4	1.8559e-4	7.6079e-5	4.9206e-5
TTE	5.8150e-4	9.2794e-4	3.8040e-4	2.4603e-4
ρ	0.9996	0.9999	0.9991	0.0004
10 日				

TE	8.1928e-5	1.1488e-4	4.9608e-5	2.6651e-5
TTE	8.0302e-4	1.1000e-3	4.9608e-4	2.4665e-4
ρ	0.9996	0.9999	0.9995	0.0002
30 日				
TE	4.6418e-5	4.9594e-5	4.3242e-5	3.1760e-6
TTE	1.2644e-3	1.5000e-3	9.9331e-4	2.0838e-4
ρ	0.9996	0.9999	0.9995	0.0002
60 日				
TE	2.7497e-5	3.0098e-5	2.4895e-5	2.6015e-6
TTE	1.6000e-3	1.8000e-3	1.5000e-3	1.4142e-4
ρ	0.9997	0.9999	0.9996	0.0001
截至下次指数调整日				
TE	2.4998e-5	3.3221e-5	1.9852e-5	5.8754e-6
TTE	3.0000e-3	3.8000e-3	2.5000e-3	5.7155e-4
ρ	0.9997	0.9998	0.9996	0.0001

附表 8 追踪误差和相关系数平均值的综合比较

(三种复制方法、两类模型和两种算法下的各自结果比较，第三次调整的组合因时间过短而除外)

	一般模型 SQP 求解 的平均值 (1)			流动性惩罚约束模型 SQP 求解的平均值 (2)			流动性惩罚约束模型 GA 求解的平均值 (3)			(1) (2) 和 (3) 总平均值			完全复制
	优化复制	抽样复制	综 合	优化复制	抽样复制	综 合	优化复制	抽样复制	综 合	优化复制	抽样复制	综 合	
样本内 (in sample)													
TE^*	1.0255e-4	1.1914e-4	1.1084e-4	1.0459e-4	1.2471e-4	1.1465e-4	1.0228e-4	1.2535e-4	1.1382e-4	1.0314e-4	1.2306e-4	1.1310e-4	-
TE	1.0549e-4	1.2367e-4	1.1458e-4	1.0815e-4	1.2740e-4	1.1778e-4	1.0751e-4	1.2861e-4	1.1806e-4	1.0705e-4	1.2656e-4	1.1681e-4	-
TTE^*	0.0116	0.0137	0.0126	0.0118	0.0140	0.0129	0.0116	0.0141	0.0128	0.0117	0.0139	0.0128	-
TTE	0.0120	0.0139	0.0129	0.0122	0.0143	0.0133	0.0123	0.0144	0.0134	0.0122	0.0142	0.0132	-
ρ	0.9952	0.9933	0.9942	0.9959	0.9946	0.9952	0.9959	0.9943	0.9951	0.9957	0.9941	0.9949	-
样本外 (out of sample)													
5 日													
TE	6.6170e-4	6.0514e-4	6.3342e-4	6.1089e-4	5.7416e-4	5.9253e-4	6.1197e-4	5.9790e-4	6.0494e-4	6.2819e-4	5.9240e-4	6.1029e-4	1.1630e-4
TTE	0.0033	0.0030	0.0032	0.0031	0.0028	0.0030	0.0031	0.0030	0.0030	0.0031	0.0030	0.0031	5.8150e-4
ρ	0.9953	0.9935	0.9944	0.9960	0.9948	0.9954	0.9960	0.9945	0.9952	0.9957	0.9942	0.9950	0.9996
10 日													
TE	4.2939e-4	4.4596e-4	4.3767e-4	4.5296e-4	4.5318e-4	4.5307e-4	4.7014e-4	4.5772e-4	4.6393e-4	4.5083e-4	4.5229e-4	4.5156e-4	8.1928e-5
TTE	0.0043	0.0044	0.0044	0.0045	0.0045	0.0045	0.0047	0.0046	0.0046	0.0045	0.0045	0.0045	8.0302e-4
ρ	0.9956	0.9939	0.9948	0.9961	0.9950	0.9956	0.9962	0.9947	0.9954	0.9960	0.9945	0.9953	0.9996
30 日													
TE	2.3513e-4	2.5483e-4	2.4498e-4	2.3588e-4	2.3253e-4	2.3421e-4	2.3877e-4	2.2517e-4	2.3197e-4	2.3659e-4	2.3751e-4	2.3705e-4	4.6418e-5
TTE	0.0071	0.0076	0.0074	0.0071	0.0070	0.0070	0.0072	0.0068	0.0070	0.0071	0.0071	0.0071	1.2644e-3
ρ	0.9954	0.9937	0.9945	0.9959	0.9948	0.9954	0.9959	0.9947	0.9953	0.9957	0.9944	0.9951	0.9996
60 日													
TE	1.6749e-4	1.8038e-4	1.7394e-4	1.6803e-4	1.7241e-4	1.7022e-4	1.6910e-4	1.6489e-4	1.6700e-4	1.6821e-4	1.7256e-4	1.7038e-4	2.7497e-5
TTE	0.0100	0.0108	0.0104	0.0101	0.0103	0.0102	0.0101	0.0099	0.0100	0.0101	0.0103	0.0102	1.6000e-3
ρ	0.9948	0.9931	0.9939	0.9953	0.9941	0.9947	0.9953	0.9940	0.9947	0.9951	0.9937	0.9944	0.9997

截至下次指数调整日													
TE	1.3322e-4	1.4565e-4	1.3944e-4	1.2981e-4	1.4043e-4	1.3512e-4	1.2899e-4	1.3129e-4	1.3014e-4	1.3067e-4	1.3912e-4	1.3490e-4	2.4998e-5
TIE	0.0162	0.0178	0.0170	0.0158	0.0171	0.0165	0.0158	0.0160	0.0159	0.0159	0.0170	0.0165	3.0000e-3
ρ	0.9939	0.9922	0.9930	0.9943	0.9928	0.9936	0.9943	0.9932	0.9938	0.9942	0.9927	0.9935	0.9997

*注释：*为优化计算未取整的结果，其余数据都为根据最小交易手数取整后计算的结果。

附表 9 交易成本比较

(三种复制方法、两类模型和两种算法下的各自结果比较)

	一般模型 SQP 求解的平均值 (1)			流动性惩罚约束模型 SQP 求解的平均值 (2)			流动性惩罚约束模型 GA 求解的平均值 (3)			(1) (2) 和 (3) 总平均值			完全复制
	优化复制	抽样复制	综合	优化复制	抽样复制	综合	优化复制	抽样复制	综合	优化复制	抽样复制	综合	
交易总成本	1.1255e+6	1.1154e+6	1.1204e+6	1.0977e+6	1.1084e+6	1.1031e+6	1.1027e+6	1.1017e+6	1.1022e+6	1.1086e+6	1.1085e+6	1.1086e+6	1.2359e+6
占资金规模比例	1.1847e-3	1.1703e-3	1.1775e-3	1.1608e-3	1.1715e-3	1.1662e-3	1.1711e-3	1.1632e-3	1.1672e-3	1.1722e-3	1.1684e-3	1.1703e-3	1.3254e-3

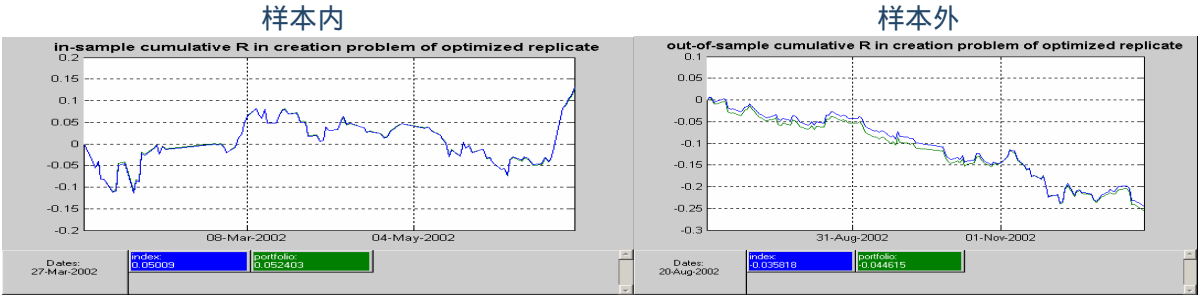
附图 1 追踪组合与上证 180 指数的累积收益率走势图

(三种复制方法、两类模型和两种算法下的各自结果比较)

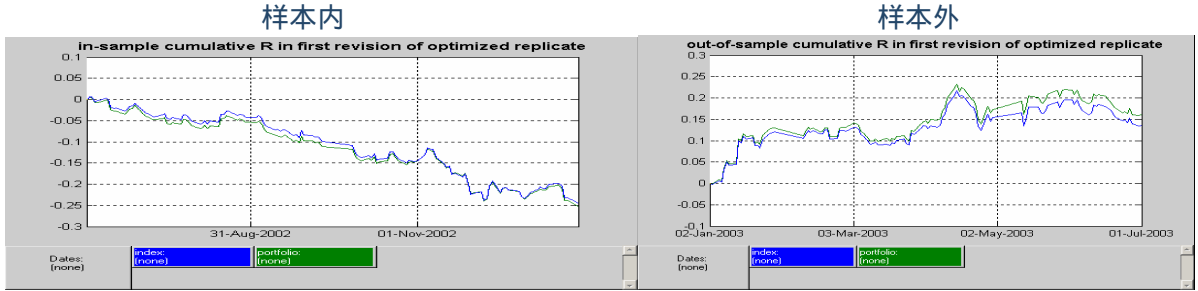
优化复制问题：

一般模型 SQP 算法下的结果：

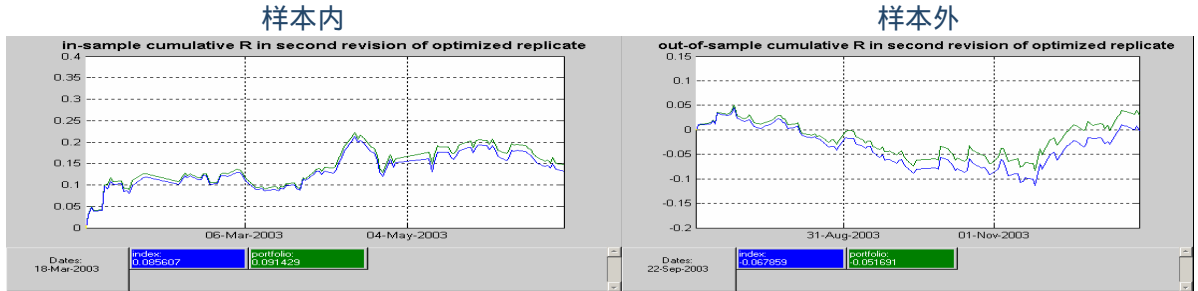
创设问题 (2002/07/01)：



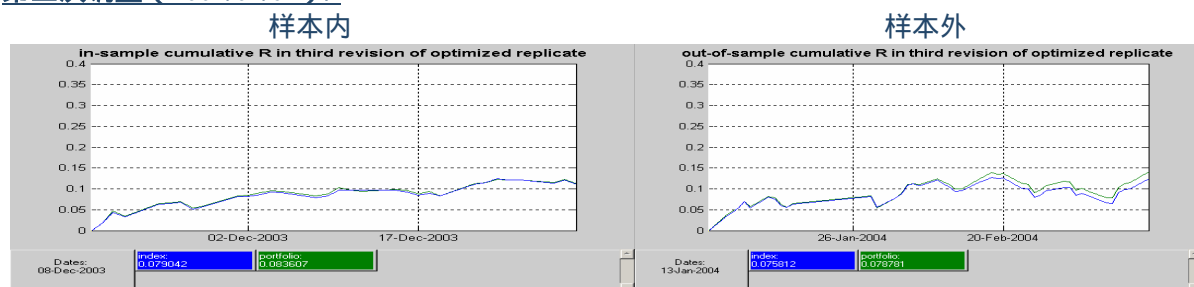
第一次调整 (2003/01/02)：



第二次调整 (2003/07/01)：

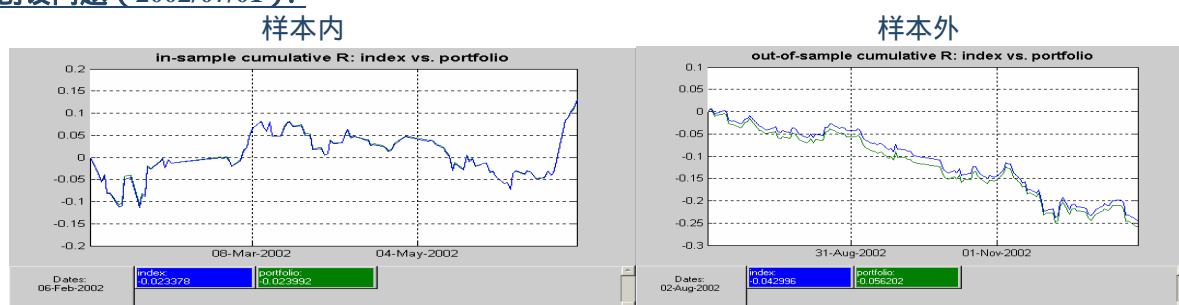


第三次调整 (2004/01/02):

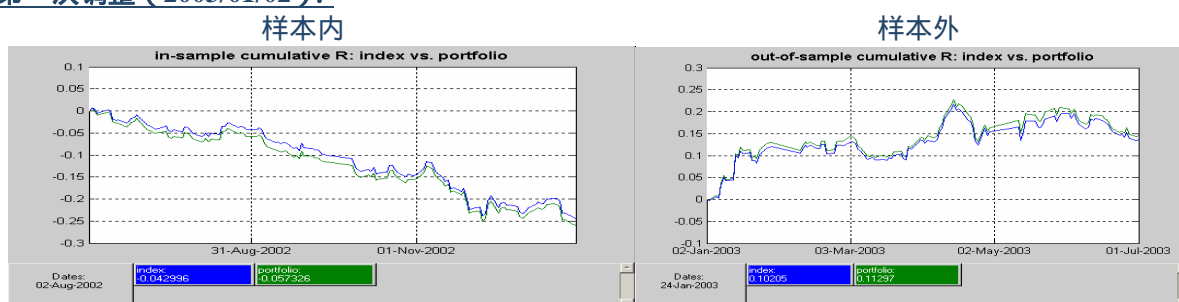


流动性惩罚约束模型SQP 算法下的结果:

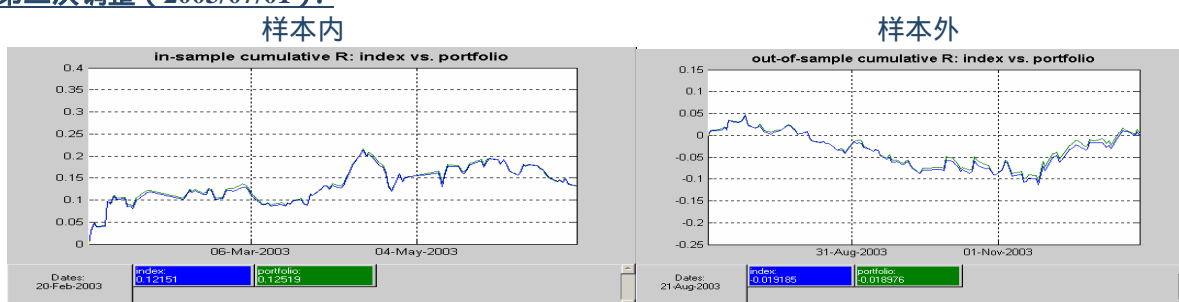
创设问题 (2002/07/01):



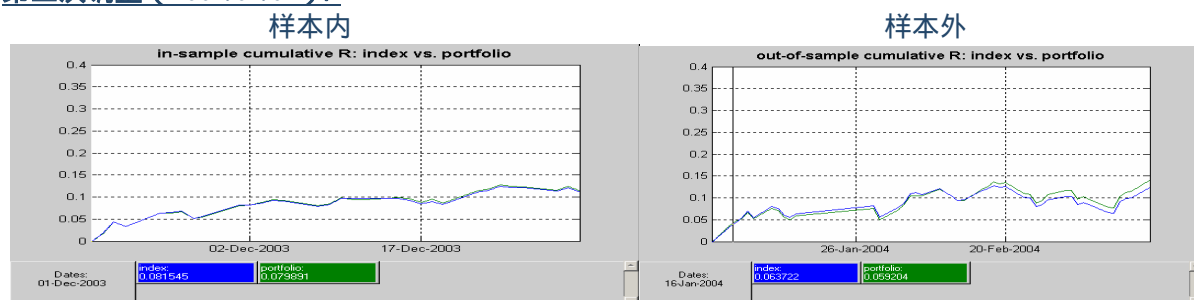
第一次调整 (2003/01/02):



第二次调整 (2003/07/01):

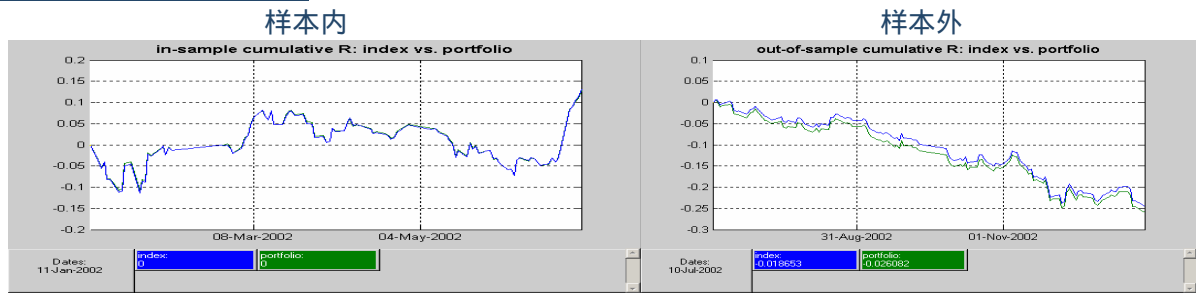


第三次调整 (2004/01/02):

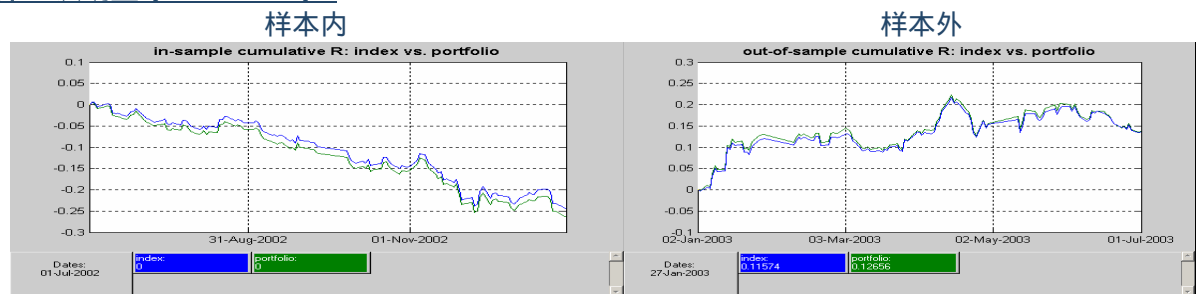


流动性惩罚约束模型 GA 算法下的结果：

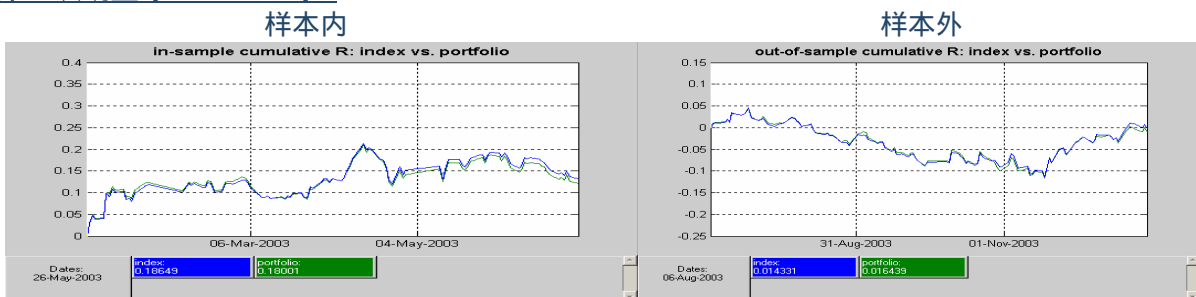
创设问题 (2002/07/01)：



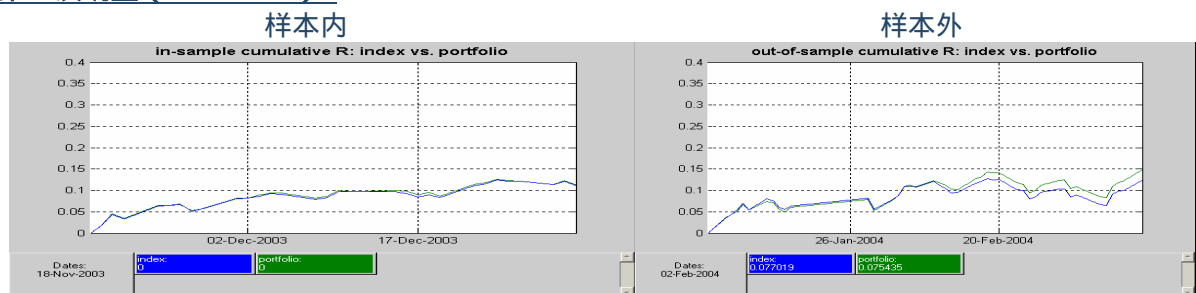
第一次调整 (2003/01/02)：



第二次调整 (2003/07/01)：



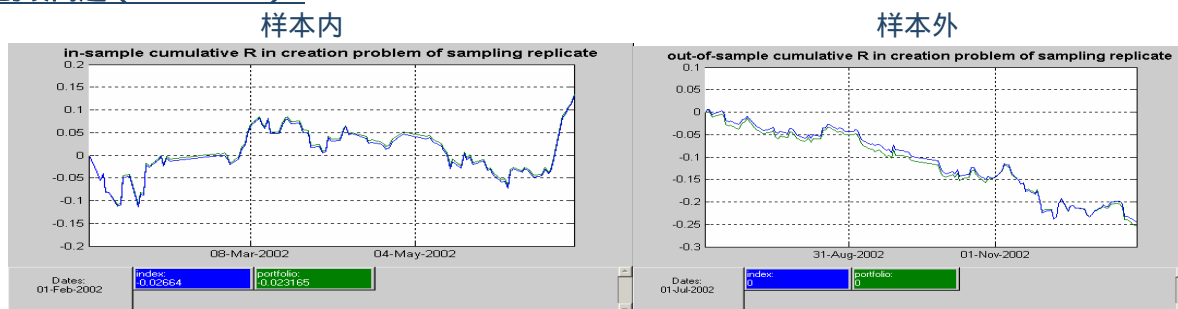
第三次调整 (2004/01/02)：



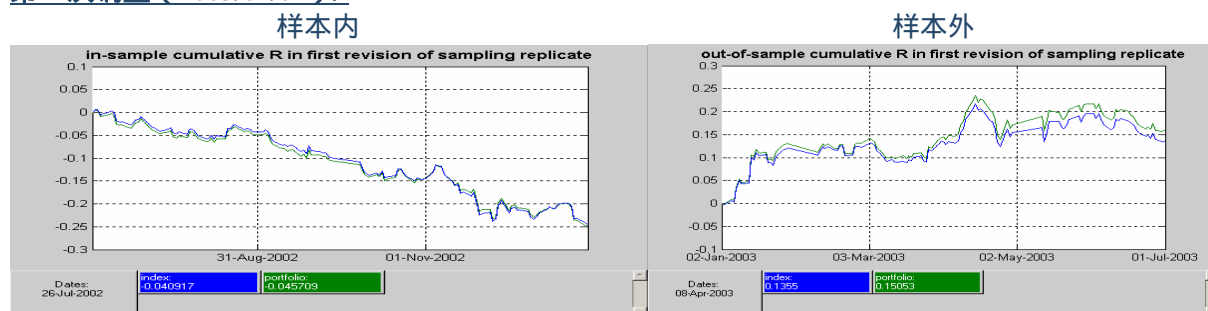
抽样复制问题：

一般模型 SGP 算法下的结果：

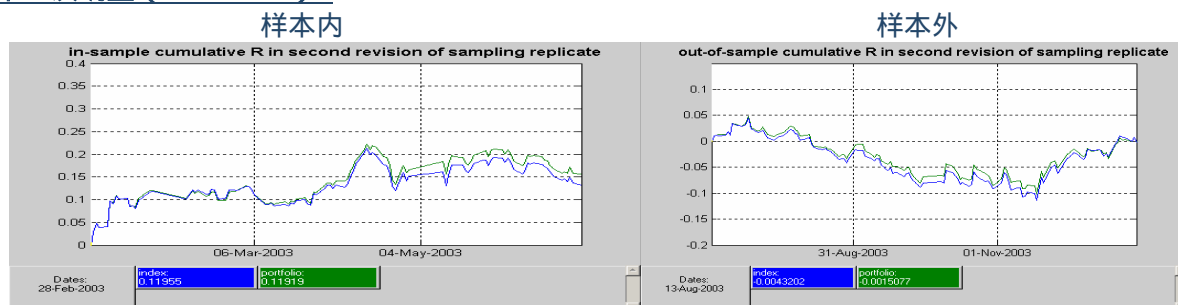
创设问题 (2002/07/01)：



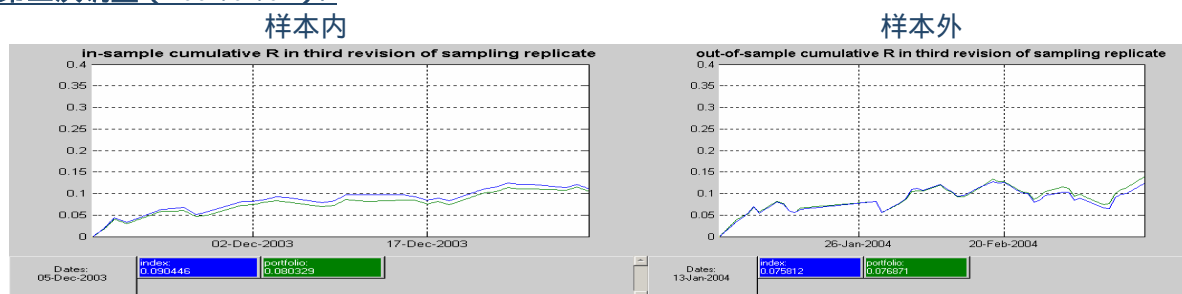
第一次调整 (2003/01/02)：



第二次调整 (2003/07/01)：



第三次调整 (2004/01/02)：

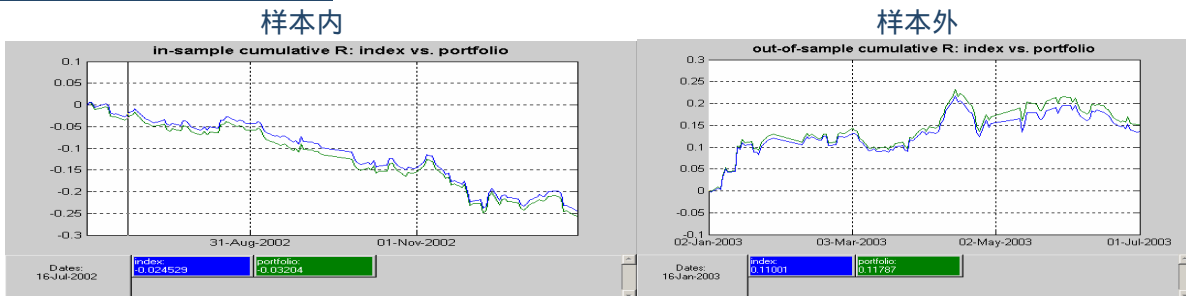


流动性惩罚约束模型 SQP 算法下的结果：

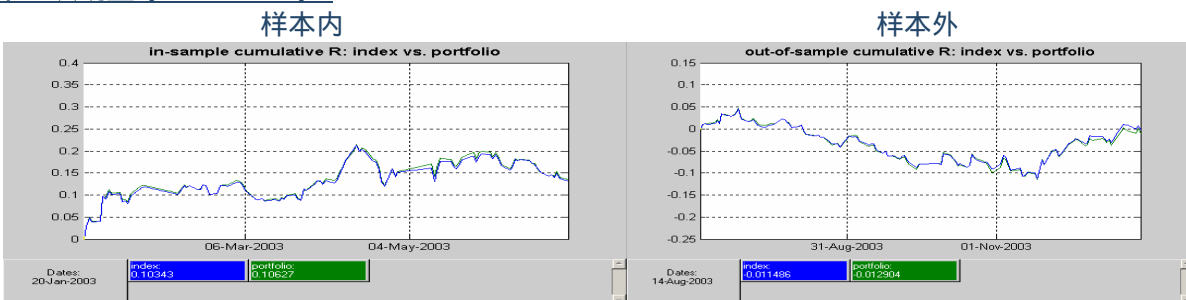
创设问题 (2002/07/01)：



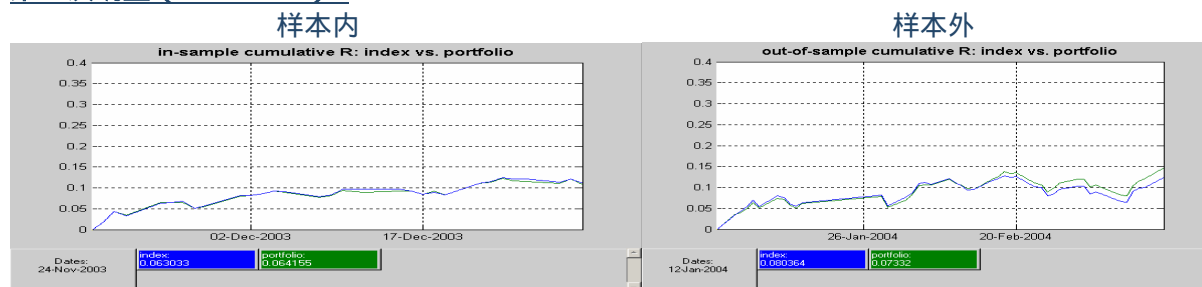
第一次调整 (2003/01/02)：



第二次调整 (2003/07/01)：

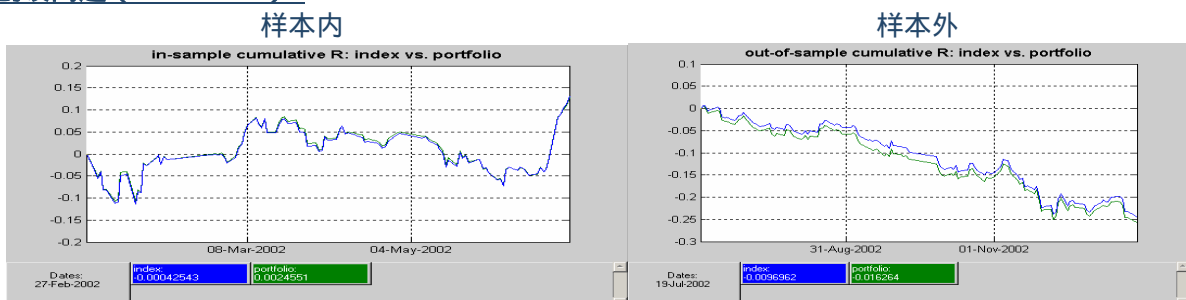


第三次调整 (2004/01/02)：

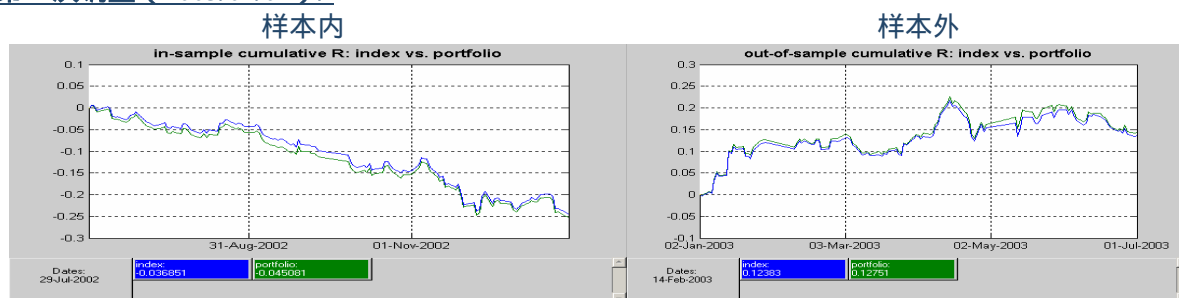


流动性惩罚约束模型 GA 算法下的结果：

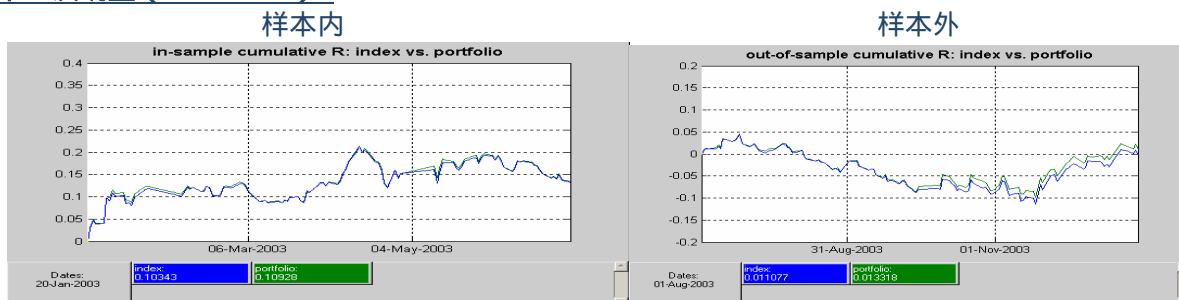
创设问题 (2002/07/01)：



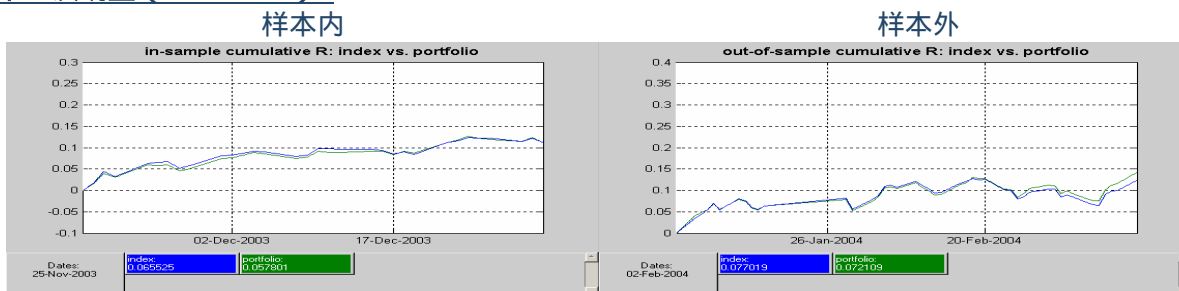
第一次调整 (2003/01/02):



第二次调整 (2003/07/01):

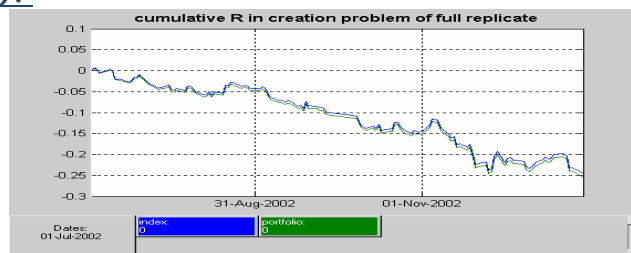


第三次调整 (2004/01/02):

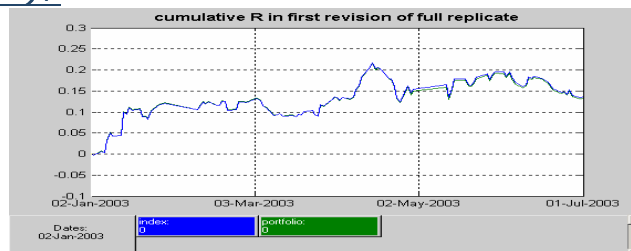


完全复制问题:

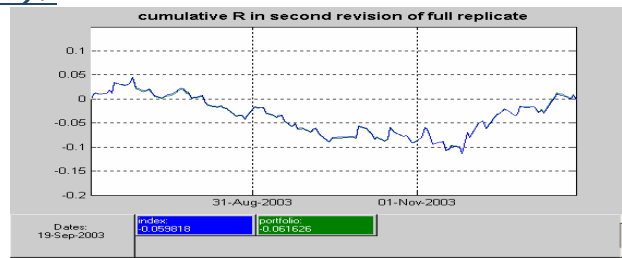
创设问题 (2002/07/01):



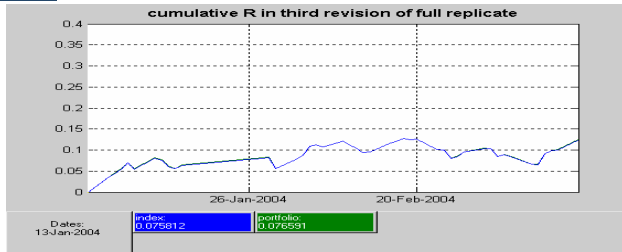
第一次调整 (2003/01/02):



第二次调整 (2003/07/01):



第三次调整 (2004/01/02):



致谢

本课题研究的成功离不开上海证券交易所的大力支持，尤其要感谢副总经理刘啸东博士，其从始至终关注着本研究的进展，课题中一些关键模型的构建和有关数据的处理直接从他细致的指导中受益。同时还要感谢副总经理方星海博士，其为本课题的形成和研究思路的安排提供了很多很有价值的帮助。课题协调人王建春先生和司徒大年先生以及研究中心的卢文莹女士对课题的研究不遗余力地给予了详细而具体的指导和支持。此外，国际发展部陈支左先生对课题的研究提供了宝贵意见，信息中心罗列女士等为本课题提供数据不厌其烦，尽心尽力。谨此，一并深表谢意！

课题组全体成员

2004 年 7 月 2 日