

基于遗传神经网络的指数跟踪优化方法

刘 磊

(中国人民银行 郑州中心支行, 郑州 450002)

摘 要 针对使用完全复制法进行指数跟踪的缺点和仅以跟踪误差作为指数跟踪目标的不足, 以跟踪误差最小化和超额收益最大化两者的权衡作为指数跟踪的目标函数, 综合考虑实际中的交易成本、现金、卖空限制等约束, 建立指数跟踪优化模型, 并采用二进制和实数值混合编码的遗传 BP 网络对指数跟踪管理中的资金进行优化配置. 该算法能同时优化网络结构和权值矢量, 并结合遗传算子和 Solis&Wets 算子生成后代使遗传搜索空间的群体多样性更好, 加快了遗传算法的收敛速度, 提高了连接权系数的优化精度. 跟踪深证 100 指数的实证结果表明: 应用遗传神经网络算法进行指数跟踪的效果明显优于完全复制法, 并且实现了目标指数的动态跟踪.

关键词 遗传神经网络; 指数跟踪; 跟踪误差; 超额收益率

Index tracking optimization method based on genetic neural network

LIU Lei

(Zhengzhou Central Branch, The People's Bank of China, Zhengzhou 450002, China)

Abstract For the disadvantages of full replication approach and only adopting tracking error as the object of index tracking, an index tracking optimal model is constructed in which a tradeoff between minimizing tracking error and maximizing excess return is treated as objective function, and practical constraints such as total transaction cost limit, cash flow constraint and short sale restriction are considered. Moreover, an improved genetic neural network using binary and real-valued combination coding method is presented to optimize the assets allocation in index tracking management. The combination of genetic algorithm and Solis&Wets operator, elitist selection make the genetic search space more diverse, accelerates the genetic algorithm's convergence speed and improve network convergence speed. Experimental results for shenzhen 100 index tracking verify that assets allocation and the performance of index tracking obtained by genetic neural network algorithm are superior to those of full replication approach, and realize dynamic tracking of the target index.

Keywords genetic neural network; index tracking; tracking error; excess return

1 引言

指数跟踪方法是对指数基金等指数产品进行消极资产配置的方法, 通过对交易成本等进行约束, 购买部分或者全部目标指数成分证券, 构造指数跟踪组合来复制目标指数的业绩, 期望指数产品与目标指数具有高度相关性并取得与目标指数相同的收益率和波动性^[1]. 从长期来看, 消极管理基金不仅不会具有积极管理基金的超额风险, 而且可以获得合理的收益^[2-4]. 因此, 随着证券市场的发展和演化, 指数跟踪方法作为指数化投资的主要管理形式, 日益引起了理论界和实务界的重视.

指数化投资理论发展至今, 产生了各种不同的指数跟踪方法, 按其原理不同, 可以分为完全复制法和优化复制法两大类^[5-6]. 目前我国指数基金大多采用完全复制法进行指数跟踪管理. 完全复制法的优点是管

收稿日期: 2007-11-19

作者简介: 刘磊 (1974-), 女, 河南郑州人, 数量经济学博士, 主要研究方向: 金融工程, 金融风险. E-mail: luckyliulei@126.com; smartliulei@yahoo.cn.

理方式较为简单,易于操作,但完全复制、完全消极被动的指数跟踪方法,需要一个高度成熟有效的资本市场^[7].对于中国这样欠成熟的资本市场,在市场波动性较大,目标指数的编制和更新及其成分证券的质量等因素存在较大欠缺的情况下,采用完全复制法可能导致较高的调整频率和跟踪时滞,带来较高的交易成本和较大的跟踪误差^[8] (Tracking error, TE) 从而抵消直接复制带来的高拟合度.

优化复制法则是使用目标指数成分股的某个子集来复制目标指数,使所构建指数组合的跟踪成本及其与目标指数之间的跟踪误差控制在可以接受的范围内,只需较少数量的证券组合就可以达到指数化投资的目的^[9].在中国缺乏衍生工具的情形下,指数跟踪的关键是根据目标指数的变化尽快调整跟踪组合,对跟踪组合的交易费用和现金流等实行尽可能精细的动态管理,所以采用良好的优化方法来进行指数跟踪管理就显得尤为重要.

此外,目前国内外对指数跟踪的研究^[10-13] 主要是注重跟踪误差这一要素,很少考虑指数复制的超额收益问题,以及这两者之间的权衡关系.如果仅以跟踪误差作为指数跟踪的目标函数,所得跟踪组合不一定能充分拟合指数的走势,因此必须同时考虑指数化投资的收益率,除了要求所构造跟踪组合的跟踪误差尽可能小之外,还要对指数跟踪过程中发生的成本(包括交易成本和现金成本等),以及指数跟踪组合再平衡策略进行管理.

针对完全复制法存在的问题和仅以跟踪误差作为指数跟踪目标函数的不足,考虑实际交易中的交易成本、现金、卖空限制和买卖股票的数量限制等约束,本文以跟踪误差最小化和超额收益最大化两者的权衡作为指数跟踪的目标函数,即遗传算法的适应度函数,采用二进制和实数值混合编码的遗传 BP 网络对指数跟踪管理中的资产优化配置进行研究,使所构建的跟踪组合收益尽可能接近目标指数收益.该算法能同时优化网络结构和权值矢量:网络结构采用二进制编码,对连接权值系数的实数编码进行 Solis&Wets 运算^[14],使新的遗传算法具有进化规划和进化策略^[15-16] 的优点;结合遗传算子和 Solis&Wets 算子生成后代的方法以及最佳个体保留策略使得遗传搜索空间的群体多样性更好,加快了遗传算法的收敛速度;采用动态参数编码^[17] 方法,提高了连接权系数的优化精度.针对深证 100 指数的实证结果表明:应用遗传神经网络算法进行指数跟踪的效果优于完全复制法,并且实现了目标指数的动态跟踪.

2 指数跟踪优化模型的目标和约束条件

作为消极管理的指数跟踪方法,指数化投资研究的内容主要有以下三个方面^[18]:最小化跟踪误差,最大化跟踪组合的超额收益,以及权衡跟踪误差与超额收益之间的关系.指数跟踪的过程,主要是跟踪误差和超额收益的管理过程,关键在于构造并维持一个能尽量拟合指数收益的跟踪组合,并尽量谋求跟踪误差最小化和超额收益率最大化之间的平衡.因此,衡量指数跟踪的目标有两个:跟踪误差和超额收益率^[19].前者是跟踪组合与目标指数收益之间的差异,后者是一定时间内收益偏差的累积效应.这两个标准度量了指数经理人对跟踪组合的成分证券选择能力、跟踪组合的成分证券调整能力以及对各种约束的管理能力,如风险管理能力和成本管理能力等.

在本文所建立的指数跟踪优化模型中,设定同时考虑跟踪误差和超额收益两个要素的指数跟踪目标函数,并考虑指数跟踪过程的约束条件,来求解最优跟踪组合.下面对指数跟踪优化模型中的变量进行定义:

N : 可投资的股票总数;

J : 跟踪组合中包含的股票数; ε_i : 在跟踪组合中持有股票 i 的最小比例 ($i = 1, 2, \dots, N$); δ_i : 在跟踪组合中持有股票 i 的最大比例 ($i = 1, 2, \dots, N$); X_i : 当前跟踪组合中股票 i 的份数 ($i = 1, 2, \dots, N$); T : 在该时刻将原跟踪组合 $[X_i]$ 转换为一个新的跟踪组合; V_{it} : 股票 i 在 t 时刻的每股股价 ($t = 0, 1, \dots, T$); R_t : 目标指数在时间 t 内的收益率 ($t = 1, 2, \dots, T$); C_{cash} : T 时刻的跟踪组合中持有的现金量, $C_{\text{cash}} > 0$ 表示可以投资到跟踪组合中的现金, $C_{\text{cash}} < 0$ 表示卖出跟踪组合中的股票得到的现金; C : 在 T 时刻,当前跟踪组合 $[X_i]$ 的总价值,通常,现实中的指数基金扣除交易成本等费用后的资产由目标指数成分证券构成的跟踪组合和现金两部分组成, $C = \sum_{i=1}^N V_{iT} X_i + C_{\text{cash}}$, $C > 0$; $F_i(\zeta, \theta, t)$: t 时刻股票 i 从持有 ζ 股变动到持有 θ 股所发生的交易成本. $\xi > \theta$ 表示卖出 $\zeta - \theta$ 股 i 股票; $\xi < \theta$ 表示买入 $\theta - \zeta$ 股 i 股票; 当 $\zeta = \theta$ 时, $F_i(\zeta, \theta, t) = 0$. 本文采用遗传神经网络算法进行运算时,对交易成本的函数形式没有限制; γ : 交易成本在 C 中所占的比例 ($0 \leq \gamma \leq 1$); x_i : 新的跟踪组合中股票 i 的份数 ($i = 1, 2, \dots, N$); C_{trans} : 在 T 时刻从当前跟踪组合 $[X_i]$ 转换为新的跟踪组合 $[x_i]$ 的总的交易成本; r_t : 新的跟踪组合 $[x_i]$ 在时间 t 内的收益率 ($t = 1, 2, \dots, T$); z_i : 若

新的跟踪组合中持有股票 i , 则 $z_i = 1$, 否则 $z_i = 0$; 此外,

$$C_{\text{trans}} = \sum_{i=1}^N F_i(X_i, x_i, T) \quad (1)$$

$$r_t = \ln \left[\left(\sum_{i=1}^N V_{i,t} x_i \right) / \left(\sum_{i=1}^N V_{i,t-1} x_i \right) \right], \quad t = 1, 2, \dots, T \quad (2)$$

t 时刻、 $t-1$ 时刻新的跟踪组合的总价值分别为 $\sum_{i=1}^N V_{i,t} x_i$ 和 $\sum_{i=1}^N V_{i,t-1} x_i$.

跟踪误差的表达式为:

$$TE = \left[\sum_{i=1}^T |r_t - R_t|^\alpha \right]^{(1/\alpha)} / T \quad (3)$$

在已有的文献中和实际应用中, 普遍采用跟踪组合收益率与目标指数收益率差额的二次项来度量跟踪误差, 本文也采取这种非中心的二阶矩偏差的方法, 取 $\alpha = 2$, 即

$$TE = \left[\sum_{t=1}^T \left| \ln \left(\sum_{i=1}^N V_{i,t} x_i / \sum_{i=1}^N V_{i,t-1} x_i \right) - R_t \right|^2 \right]^{(1/2)} / T \quad (4)$$

超额收益率表示为:

$$r^* = \sum_{i=1}^T (r_t - R_t) / T \quad (5)$$

$$r^* = \sum_{t=1}^T \left[\ln \left(\sum_{i=1}^N V_{i,t} x_i / \sum_{i=1}^N V_{i,t-1} x_i \right) - R_t \right] / T \quad (6)$$

r^* 表示跟踪组合所获的总收益与目标指数总收益的差距.

如果所有的基金都仅仅跟踪指数而不考虑其它目标, 那么这些指数基金之间的区别就在于它们的成本, 但是如果能够在一定程度上牺牲跟踪误差, 就有可能获得超过指数的收益 ($r^* > 0$), 考虑跟踪误差与超额收益两者的权衡关系, 得到指数跟踪的一般目标为:

$$\min \lambda * TE - (1 - \lambda)r^*, \quad 0 \leq \lambda \leq 1 \quad (7)$$

这里 λ 的变化体现了跟踪误差与超额收益率之间隐含的权衡关系, $\lambda = 1$ 对应指数跟踪的目标为最小化跟踪误差, $\lambda = 0$ 对应指数跟踪的目标为最大化超额收益.

由式 (4)、(6) 和 (7), 可得目标函数的具体形式为:

$$\begin{aligned} \min \lambda \left\{ \left[\sum_{t=1}^T \left| \ln \left(\sum_{i=1}^N V_{i,t} x_i / \sum_{i=1}^N V_{i,t-1} x_i \right) - R_t \right|^2 \right]^{(1/2)} / T \right\} - \\ (1 - \lambda) \left\{ \sum_{t=1}^T \left[\ln \left(\sum_{i=1}^N V_{i,t} x_i / \sum_{i=1}^N V_{i,t-1} x_i \right) - R_t \right] / T \right\} \end{aligned} \quad (8)$$

模型的约束条件为:

$$\sum_{i=1}^N V_{i,T} x_i = C - C_{\text{trans}} \quad (9)$$

式 (9) 为现金平衡约束, 表示 T 时刻新的跟踪组合的总价值必须等于 T 时刻原跟踪组合的价值加上预留现金减去总的交易成本.

$$C_{\text{trans}} \leq \gamma C \quad (10)$$

当 γ 值给定时, 式 (10) 为实际的交易成本约束.

$$x_i \geq 0, \quad i = 1, 2, \dots, N \quad (11)$$

考虑到我国存在卖空限制, 式 (11) 为跟踪组合中各股票的卖空限制约束.

$$\varepsilon_i z_i \leq V_{i,T} x_i / C \leq \delta_i z_i \quad (12)$$

式 (12) 为买卖股票 i 的交易成本在 C 中所占比例, 当新的跟踪组合不持有股票 i ($z_i = 0$) 时, $x_i = 0$; 当新的跟踪组合持有股票 i ($z_i = 1$) 时, x_i 的取值受到适当限制 ($\varepsilon_i \leq V_{i,T} x_i / C \leq \delta_i$).

3 算法介绍

3.1 混合编码下的遗传算法

遗传算法由于其简单通用、鲁棒性强、适用于并行处理等特点,已在许多领域得到应用,Shapcott^[20]在 1992 年首次将遗传算法应用于指数基金构建领域,但经典的 GA 存在搜索后期效率低和缺乏局部优化能力等情况,为此,人们尝试了许多改进方法^[21-22],本文采用混合编码下的遗传算法来克服这些缺陷.下面结合指数跟踪的资产优化配置对该算法作简单介绍:

1) 编码:本文中神经网络的结构采用二进制编码,权值系数(即跟踪组合中各股票所占比例)采用实数值编码.下面以图 1 所示的网络为例说明这种混合编码的方法.

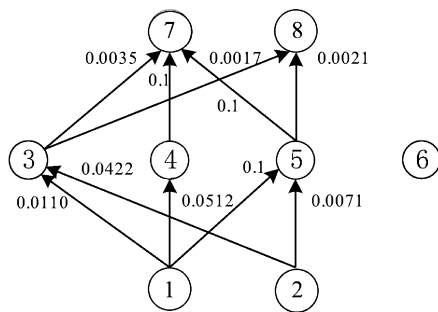


图 1 网络结构和权值的混合编码

第 1 号神经元只与 3, 4 和 5 号相连,结构编码是按神经元编号顺序以及与之相连的神经元编号顺序来编排的,1 表示相连接(即新的跟踪组合中持有股票 i , $z_i = 1$), 0 表示不连接(即新的跟踪组合中不包括股票 i , $z_i = 0$),因此编码为 1110.为使表示结构的二进制码与表示权值的实数码一一对应,对于不连接的神经元,令其权值为 0,因此与第 1 号神经元结构编码对应的实数值编码为 0.0110, 0.0512, 0.1, 0.因此图 2 所示网络混合编码结果为:输入层结构编码: 1110 1010,隐含层结构编码: 11 10 11 00,对应的权值为 0.0110 0.0512, 0.1, 0, 0.0422, 0, 0.0071, 0, 0.0035, 0.0017, 0.1, 0, 0.1, 0.0021, 0, 0.

本文选取深证 100 指数作为目标指数,采用 100-15-1 的网络结构,初始群体(初始跟踪组合)每个混合编码串有 3030 个元素组成,前 1515 个为随机生成的、值为 0 或 1 的二进制码,后 1515 个为随机生成的表示网络权值的实数码串.

2) 初始群体(初始跟踪组合):根据初始样本空间中的备选股票数目,染色体(个体)上的每 1 位(基因)代表 1 只股票,1 条染色体就代表 1 种投资组合;混合遗传算法的初始群体的每个码串是随机产生的,但实数值编码被限定在可能的连接权值系数取值范围内,考虑到指数跟踪组合中各股票所占实际比例一般不超过 10%,本文取值范围为 $[0, 0.1]$,且所有权值和为 1.

3) 适应度计算和遗传算法的选择算子:适应度函数选取式(8)所示的目标函数,同时满足式(9)-(12)所示的约束条件.选择算子采用每代中所有个体适应度的平均值,低于平均适应度的个体被淘汰,在进行下一次交叉之前用随机生成的码串补齐被淘汰的个体.

4) 交叉算子和变异算子:本文提出的遗传算法可对编码串中二进制码串和实数码串同时进行标准遗传算法中的交叉和变异运算.在进行交叉运算时,将对应的实数值编码进行对等的交叉来完成整个码串的交叉算子的运算,本文采用单点交叉,交叉概率取 0.8.在进行变异运算时,对于表示神经网络权值系数的实数值编码进行相对应的变化,如果某两个神经元之间的连接被变异算子取消,则对应的权值系数的实数值编码也被取消,若变异算子增加了某两个神经元之间的连接,则也要相应的增加其权值系数的实数值编码,编码值随机确定,且不超过 0.1,变异率取 0.01.对于最佳个体(最优跟踪组合)不进行交叉和变异操作.

这种对混合编码进行的遗传变异算子运算,对表示神经网络结构的二进制码串部分产生了较强的变异效果,而对表示神经网络权值系数的实数值编码变异效果较弱.因此,本文提出了包含 Solis & Wets 运算的混合搜索法来加强网络权值系数的实数值编码的变异效果.

3.2 Solis & Wets 算子

Solis & Wets 算法也是一种全局寻优算法^[23],该算法随机生成预优化的数列,数列中元素数目等于预优化的变量数目.该算法以随机生成的数列为初始数列,在随机搜索过程中根据目标函数值的上升或下降来自适应调整搜索的方向和步长,最终获得全局最优解.下面结合本文情况对该算法作简要介绍.

将混合编码中表示神经网络每一个权值系数的实数编码分别作为一个分量 w_i ,而整个实值码串就是一个数列 W .首先选取阈值参数 K_{\max} , $scnt$, $fcnt$, σ_{ub} , σ_{lb} , ex , ct 且初始化变量 $k = 0$, $scnt = 0$, $fcnt = 0$, $\sigma_0 = 1$, $\mu_0 = 0$.其中, k 为搜索次数, K_{\max} 为最大搜索次数,本文中取 1000, $scnt$ 和 $fcnt$ 为循环搜索过程

中的暂态变量, 分别表示在循环中目标函数下降时连续成功和失败的次数, $Scnt, Fcnt$ 分别为 $scnt, fcnt$ 的比较阈值, 本文分别取 5 和 3. σ_0 为 Gaussian 随机变量的初始方差, σ_{ub}, σ_{lb} 分别为方差比较阈值的上下限, 对应取 3 和 0.1. ex, ct 是方差步进系数, 即目标函数下降时连续成功或失败的次数达到或超过设定的相应阈值时对方差 σ_{k-1} 所进行的扩大或缩小的操作系数, 本文分别取 2 和 0.5, μ_0 为 Gaussian 的初始均值. 将数列 W 作为初始搜索数列 W_0 , Solis & Wets 运算过程如下:

1) 设置方差变量为

$$\sigma_k = \begin{cases} ex \cdot \sigma_{k-1}, & \text{若 } scnt > Scnt, \\ ct \cdot \sigma_{k-1}, & \text{若 } fcnt > Fcnt, \\ \sigma_{ub}, & \text{若 } \sigma_{k-1} < \sigma_{lb}, \\ \sigma_{k-1}, & \text{其他.} \end{cases}$$

2) 生成一个多维 Gaussian 随机变量: $\xi_k = N(\mu_k, \sigma_k)$.

3) 计算目标函数值

(a) 如果 $f(W_k) < f(\xi_k)$, 则:

$$W_{k+1} = \xi_k, \quad \mu_{k+1} = 0.2\mu_k + 0.4(\xi_k - W_k), \quad scnt = scnt + 1, \quad fcnt = 0;$$

(b) 如果 $f(\xi_k) \leq f(W_k) < f(2W_k - \xi_k)$, 则:

$$W_{k+1} = 2W_k - \xi_k, \quad \mu_{k+1} = \mu_k - 0.4(\xi_k - W_k), \quad scnt = scnt + 1, \quad fcnt = 0;$$

(c) 否则: $W_{k+1} = W_k, \mu_{k+1} = 0.5\mu_k, fcnt = fcnt + 1, scnt = 0$.

4) 如果 $k = k_{max}$, 停止 Solis&Wets 运算. 否则 $k = k + 1$, 返回到 1).

需说明一点, 上述迭代过程中 $f(W_k)$ 表示数列 W_k 解码后所对应的 BP 网络的目标函数值, 即遗传算法中的适应度函数值, 依此类推.

3.3 混合编码下遗传神经网络算法

本文提出的混合编码下的遗传神经网络算法描述如下:

1) 随机产生一组有 50 个个体的初始群体, 按前述编码方法对其进行混合编码, 进化代数置为 1;

2) 对一部分个体通过对其二进制码串部分进行标准遗传算法生成后代, 其他个体则通过对其实数值码串部分进行 Solis & Wets 运算生成后代;

3) 计算所有个体的适应度, 若适应度满足中止条件, 转 4), 否则转 7);

4) 选择 m 个具有全局性的进化解, 选择时应避免相似个体;

5) 分别以这些进化解为初始解, 计算神经网络的输出最优解;

6) 计算 m 个最优解的均值和方差, 以方差最小的解作为最优解输出, 结束;

7) 利用选择算子生成新的父代, 进化代数加 1, 转 2).

步骤 4) 中的全局性是由 GA 搜索机理提供的, GA 能以较大概率进化至全局解区域, 但不能明确是哪个解进化至该区域. 因此不能简单地认为第 k 代群体中适应值最高的个体就是全局性进化解, 选择 m 个适应值较高而彼此又不相似的个体就增加了包含全局性进化解的概率^[8].

整个遗传神经网络算法的流程图如图 2 所示.

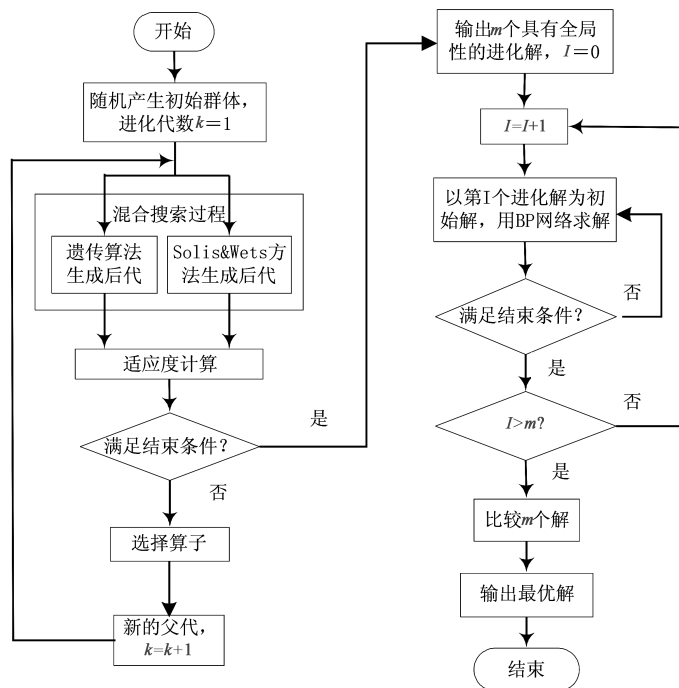


图 2 遗传神经网络算法流程图

4 基于遗传神经网络的指数跟踪实证分析

本文选择深证 100 指数为目标指数, 以深证 100 指数的 100 只成分股作为备选股票, 构造跟踪组合形成指数基金. 由于深证 100 指数的调整频率为每半年调整一次 (调整时间分别为 5 月和 11 月), 每次调整的比例一般情况不超过 10%, 除此之外无临时性调整. 为充分考虑每次调整对目标指数的影响, 选取研究时段为 2006 年 10 月 9 日到 2007 年 8 月 31 日共计 227 个工作日, 在这段时间内目标指数进行过 2 次调整. 以 2006 年 10 月 9 日到 2007 年 3 月 16 日的 107 天作为训练期, 以 2007 年 3 月 19 日到 2007 年 8 月 31 日的 115 天为检验期, 训练期和检验期各有一次调整. 实证研究数据来源于 CSMAR 数据库, 并且所有的收益率是考虑现金红利再投资的日个股收益率. 此外, 在不考虑冲击成本的前提下, 本文假定两种方法下的跟踪组合都在实施调整的前一天以收盘价完成调整和构造, 其中采用完全复制法的跟踪组合再平衡的时间、频率和成分股调整情况与目标指数相同.

4.1 实证研究中的具体约束条件

指数跟踪组合的资产配置过程本质上是一个在约束条件下的最优化过程, 本文采用遗传神经网络算法和完全复制法进行指数跟踪时, 构造和调整指数跟踪组合的共同约束为:

1) 初始总资产的价值为 C , 本文取初始总资产为 10 亿元 ($C = 10^8$). 考虑到指数基金的总资产由指数跟踪组合的价值和现金两部分组成, 初始跟踪组合和调整跟踪组合中用于应付购买和赎回的预留现金比例设为 4%, 即 $C_{\text{cash}} = 0.04C$;

采用遗传神经网络算法进行指数跟踪的其它约束条件为:

2) 初始跟踪组合包括 100 支股票, 即 $J=100$;

3) $\varepsilon_i = 0, \delta_i = 0.1$, 即跟踪组合中任何一只股票投资比例下限为 0, 上限为 10%;

4) $F_i(\zeta, \theta, t) = 0.005|\zeta - \theta|V_{it}$, 即交易费用为所购入或卖出股票价值的 0.5%;

5) 取 $\gamma = 0.003$, 即跟踪组合的交易成本比例为 0.3%;

6) λ 的初始值为 1.

4.2 实证方法和实证结果

为验证遗传神经网络算法的优劣, 本文将采用该算法获得的指数跟踪效果与完全复制法的跟踪效果进行对比. 在完全复制法下, 以 2006 年 10 月 9 日深证 100 指数的 100 只成分股作为备选股票, 构造初始跟踪组合, 同时满足前述的第一个约束条件, 并且跟踪组合的变动始终与深证 100 指数的调整情况保持一致. 以 2007 年 3 月 19 日到 2007 年 8 月 31 日 115 天的日数据为例来检验该方法的跟踪性能, 获得如图 3 所示的跟踪效果. 采用遗传神经网络算法时, 以 2006 年 10 月 9 日到 2007 年 3 月 16 日 107 天的日数据作为训练样本, 采用第二部分所确定的目标函数, 根据前述的具体约束条件和构建、调整跟踪组合的方法对神经网络进行训练¹. 然后同样以 2007 年 3 月 19 日到 2007 年 8 月 31 日 115 天的日数据为检验数据, 验证该算法的性能, 获得如图 4 所示的跟踪效果.

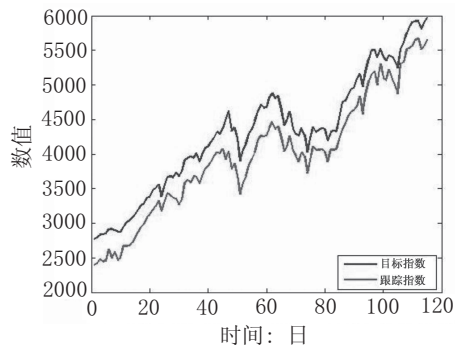


图 3 完全复制法下的指数跟踪情况

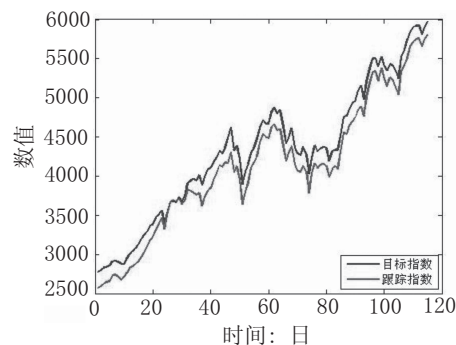


图 4 基于遗传神经网络算法的指数跟踪情况

对指数基金管理者和投资者而言, 除了考虑指数基金跟踪误差尽可能小之外, 也希望同时能够获得较高的超额收益率. 从图 3、图 4 可以看出, 当用跟踪误差和超额收益率两者的权衡作为指数跟踪目标时, 运用遗传神经网络算法进行优化复制, 得到的跟踪组合的指数跟踪效果明显优于完全复制法下得到的跟踪效果. 由于中国证券市场处于高速成长期, 在目标指数波动较大, 有周期性调整时, 采用遗传神经网络算法能够迅速

1. 本文使用 Matlab 7.0 软件进行实证分析, 实证结果中的图例也由 Matlab 7.0 相关程序来实现.

地调整跟踪组合,无论在大市下跌时还是上涨时都能获得与目标指数较为一致的收益率和波动性;而完全复制法由于不包括任何积极管理的成分,跟踪迅速上涨的深证 100 指数显得比较滞后.此外,每次指数调整时,和完全复制法相比,遗传神经网络算法都能获得一个更优的跟踪组合,但很少超越目标指数.这也从另一个角度说明,指数基金的优势在于低成本和保证获得市场平均收益,而不是为了跑赢大盘.采用遗传神经网络算法进行指数跟踪优化的目的也在于此,即在较长期的投资过程中,能够在充分考虑各种实际约束的同时提高指数跟踪的效果,忽略短期市场波动,从而分享证券市场成长的长期平均收益.

为进一步考察采用以上两种方法进行指数化投资的具体效果,本文比较了这两种方法所获跟踪组合的跟踪误差、超额收益率和其它性能指标.首先分别按照式(4)和式(6)计算得到跟踪误差和跟踪组合的超额收益率,其中跟踪误差为日平均跟踪误差,超额收益率为日平均超额收益率.其次,由于 Sharp 指数²能够同时考虑包括系统风险和非系统风险在内的总风险,而且反映了基金经理分散和降低非系统风险的能力;信息比率³则是用跟踪误差来衡量投资组合的风险水平的一种风险调整收益率指标,反映了单位跟踪误差所产生平均超额收益率,所以这两个指标可以较好地反映风险调整后的指数基金的业绩.因此本文选用这两个指标来

衡量以上两种方法的业绩表现,分别计算了两种方法下的 Sharp 指数和信息比率,结果如表 1 所示.

表 1 指数跟踪目标及业绩指标比较

跟踪组合	跟踪误差	超额收益率 (%)	Sharp 指数	信息比率 (IR)
遗传神经网络算法	0.00247	3.338	1.2904	0.2762
完全复制法	0.00313	-1.545	0.6131	-0.0926

由表 1 可知: 1) 从跟踪误差的角度看,运用遗传神经网络算法进行优化复制的跟踪精度稍好,完全复制法的跟踪精度略差,但两种方法下跟踪误差的差别不大; 2) 从超额收益率的角度看,优化复制法所获跟踪组合的超额收益率比完全复制法下的超额收益率大的多,这说明构造跟踪组合时所运用的优化方法对指数基金的资产配置以及收益率都有非常大的影响,能够极大地提高指数化投资获取超额收益率的能力; 3) 与完全复制法相比,遗传神经网络算法所获跟踪组合的 Sharp 指数和信息比率都明显较大.由于 Sharp 指数表示单位总风险所获的超额收益率,较大的 Sharp 指数表示较好的业绩;而且较大的信息比率表示投资组合获得了较高的风险补偿水平,具有较高的投资绩效,由此说明遗传神经网络算法下的实际收益率高于相同风险水平下完全复制法所获跟踪组合的实际收益率,遗传神经网络算法的优势较为明显.

通过对表 1 和图 3、图 4 的分析可知,和完全复制法相比,采用遗传神经网络算法进行指数基金的资产配置可以在保持跟踪误差较小的情况下使指数基金的超额收益率得到大幅度提高,表明该算法对优化指数跟踪,提高指数基金业绩方面具有明显的优势.

5 结论

由于现实的中国股票市场是有摩擦的市场,并且新股不断上市,股票市场处于高度成长期.这些原因不仅使股票市场的波动性增大,而且造成了指数成分股市值结构的巨大变化,给指数跟踪带来了很大困难.因此,在我国证券市场的不断发展和深化的情况下,寻求良好的指数跟踪方法,考虑指数跟踪实务中存在的约束条件,在较小的跟踪误差和较高的超额收益范围内用少量的成分证券来实现对整个目标指数的优化复制,对于进行消极资产配置和构造指数跟踪组合非常重要,对投资者和基金管理者也具有重要的意义.

针对使用完全复制法进行指数跟踪的缺点和仅以跟踪误差最小化作为指数跟踪目标的不足,本文提出了基于遗传神经网络的指数跟踪优化方法.该方法的优点在于:

1) 运用该方法可以创建最初的投资组合,即解决创建问题,也可以在目标指数变动时,构建新的跟踪组合,并且考虑到交易成本、预留现金等约束,即解决再平衡问题.

2) 将指数化投资中跟踪组合创建或再平衡问题转化为最优规划问题,充分利用该算法非线性和多维的特点,综合考虑了交易费用和现金约束、个股最大最小投资比例和卖空限制等,不仅在投资组合空间中能快速搜索出符合条件的满意解,避免搜索过程陷入局部最优解,而且在具体求解最优跟踪组合时,提高了投资权值向量的适应性,得到了鲁棒性较高的跟踪组合,减少了跟踪组合再平衡的次数.

3) 通过调整交易成本在跟踪组合总价值中所占比例,实现跟踪误差最小化和超额收益率最大化两者之

2. Sharp 指数 (S_p) 用基金投资组合的平均超额收益除以这个时期收益率的标准差, $S_p = (R_p - R_f) / \sigma(R)_p$, R_p 为基金在样本期间的平均收益率, R_f 为同期 1 年期银行定期存款利率按连续复利法折算的周无风险收益率, $\sigma(R)_p$ 为基金收益率的标准差.
3. 信息比率 (IR) 的计算公式为: $IR = (R_p - R_b) / TE$, 其中 R_p 表示基金在样本期间的平均收益率, R_b 表示目标指数在样本期间的平均收益率, TE 表示跟踪组合的跟踪误差.

间的权衡.

4) 指数跟踪的目标函数是跟踪误差和超额收益率的线性函数, 而且其权值是不确定的, 本文采用的遗传神经网络算法, 能够动态地确定两部分的权值, 反映在不同跟踪时段, 这两部分的重要程度不同.

因此, 采用遗传神经网络算法不仅使指数跟踪优化模型更具真实性和客观性, 能够有效地控制跟踪组合相对目标指数的偏离风险, 较好地拟合了目标指数的增长, 实现了指数基金构建的动态跟踪. 对深证 100 指数进行跟踪的实证结果表明: 这种方法能够使指数跟踪管理中的选股过程和资金配置过程具有随目标指数变化而自适应调整的能力, 因此所构建的跟踪组合收益能够尽可能接近目标指数收益, 指数跟踪效果显著优于完全复制法. 这一良好的指数跟踪优化方法不仅对提高指数基金的业绩具有重要作用, 而且对指数基金管理也有重要的借鉴意义.

参考文献

- [1] Frino A, Gallagher D R, Neubert A. Index design and implications for index tracking[J]. *The Journal of Portfolio Management*, 2004, 30(2): 89–95.
- [2] Larsen J G A, Resnick B G. Empirical insights on indexing[J]. *Journal of Portfolio Management*, 1998, 25(1): 51–60.
- [3] Browne S. Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark[J]. *Finance and Stochastics*, 1999, 3(3): 275–294.
- [4] Sorenson E H, Miller K L, Samak V. Allocating between active and passive management[J]. *Financial Analysts Journal*, 1998, 54(5): 18–31.
- [5] Olama A. Implementing Equity Index Portfolios[M]. *Indexing for Maximum Investment Results* GPCo Publishers, 1998.
- [6] Frino A, Gallagher D R. Is index performance achievable? An analysis of Australian equity index funds[J]. *ABACUS*, 2002, 38(2): 200–214.
- [7] 高见, 杨丹. 指数化投资复制方法的比较分析 [J]. *金融研究*, 2006, (8): 31–40.
Gao J, Yang D. Comparative analysis of replication approach of index investment[J]. *Journal of Financial Study*, 2006, (8): 31–40.
- [8] Masters S J. The problem with emerging markets indexes[J]. *The Journal of Portfolio Management*, 1998, 24(2): 93–100.
- [9] Bogle J C. Selecting equity mutual funds[J]. *The Journal of Portfolio Management*, 1992, 18(2): 94–100.
- [10] Baestaens D E, Van den Bergh W M. Tracking the Amsterdam stock index using neural networks[J]. *Neural Networks in Capital Markets*, 1995, (5): 149–161.
- [11] Jansen R, Dijk D V. Optimal benchmark tracking with small portfolios[J]. *Journal of Portfolio Management*, 2002, 28(2): 33–39.
- [12] 谭治国, 袁靖. 遗传算法构建我国指数基金自动生成模型的实证分析 [J]. *石家庄经济学院学报*, 2007, (4): 31–34.
Tan Z G, Yuan J. The empirical analysis of automatic index fund construction model in China based on genetic algorithm[J]. *Journal of Shijiazhuang University of Economics*, 2007, (4): 31–34.
- [13] 劳剑勇. 指数跟踪问题的广义双线性规划模型 [J]. *应用数学与计算数学学报*, 2004, (6): 85–91.
Lao J Y. Generalized bilinear programming model for index tracking problem[J]. *Communication on Applied Mathematics and Computation*, 2004, (6): 85–91.
- [14] Francisco J S, Roger J B W. Minimization by random search techniques[J]. *Mathematics of Operations Research*, 1981, (6): 19–30.
- [15] Peter J A, Gregory M S, Jordan B P. An evolutionary algorithm that constructs recurrent neural networks[J]. *IEEE Transactions on Neural Networks*, 1994, (5): 54–65.
- [16] David B F. An introduction to simulated evolutionary optimization[J]. *IEEE Transactions on Neural Networks*, 1994, (5): 3–14.
- [17] Nicol N S, Richard K B. Dynamic parameter encoding for genetic algorithms[J]. *Machine Learning*, 1992, (9): 9–21.
- [18] Wang M Y. Multiple-benchmark and multiple-portfolio optimization[J]. *Financial Analysts Journal*, 1999, 51(1): 63–72.
- [19] Beasley J E, Meade N, Chang T J. An evolutionary heuristic for the index tracking problem[J]. *European Journal of Operational Research*, 2003, 148(3): 621–643.
- [20] Shapcott J. Index tracking: Genetic algorithms for investment portfolio selection[EB/OL]. [http://www.smartquant.com/references/Portfolio Selection/portfselect2.pdf](http://www.smartquant.com/references/Portfolio%20Selection/portfselect2.pdf).
- [21] Yao X. A review of evolutionary artificial neural networks[J]. *International Journal of Intelligent Systems*, 1993, (8): 539–567.
- [22] John R K, James P R. Genetic generation of both the weights and architecture for a neural network[J]. *International Journal of Intelligent Systems*, UCNN-91-Seattle, 1991: 397–404.
- [23] Francisco J S, Roger J B W. Minimization by random search techniques[J]. *Mathematics of Operations Research*, 1981, (6): 19–30.