

基于时间加权 SVM 的指数优化复制模型与实证分析

胡春萍¹, 薛宏刚², 徐凤敏³

(1. 西安交通大学 公共政策与管理学院, 西安 710049; 2. 西安交通大学 经济与金融学院, 西安 710061;
3. 西安交通大学 数学与统计学院, 西安 710049)

摘要 本文研究了指数型基金管理中最核心的指数追踪问题. 依据结构风险最小化思想, 首先构造了以追踪误差为基础的损失函数, 接着考虑了时间因素对历史追踪误差的影响, 采用指数加权的方法将时间因素纳入追踪误差的分析中, 利用 SVM 理论给出了指数复制问题的结构风险的形式, 在此基础上构建了各种市场实际约束下最小化结构风险的时间加权 SVM 的指数优化复制模型, 最后利用 OR-Library 中的 5 个市场指数历史数据进行了实证检验. 结果表明本文提出的时间加权 SVM 模型能够显著提高样本外的追踪效果. 同时使得追踪组合的鲁棒性有明显的改善, 从而表明本文提出的时间加权 SVM 指数复制模型具有较高的理论和应用价值.

关键词 指数复制; 时间加权; 结构风险; SVM

An stock index replicating model based on time weighted SVM and it's empirical analysis

HU Chun-ping¹, XUE Hong-gang², XU Feng-min³

(1. School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an 710049, China; 2. School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710061, China; 3. School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract In this paper, we study the index tracking problem which plays core role in index fund management. According to minimizing structure risk, a loss function based on tracking error is constructed firstly, then time is considered in history tracking error, and exponential weighted method is used to handle it, so the structure risk is proposed on the basis of SVM, a model is generated to deal with index tracking problem by minimizing structure risk constrained some real market conditions, lastly, some empirical tests are conducted using five index historical data from OR-Library. The numerical results show that our model can improve the out-of-sample performance, at the same time, the robust of the tracking portfolio is improved evidently, thus it has good theoretical and applied value.

Keywords index tracking; time weight; structure risk; SVM

1 引言

指数复制 (也称指数追踪) 是指利用资本市场上若干个金融资产的组合来追踪市场上某一指数的表现, 例如用若干只股票的组合追踪某个股票指数. 由于指数追踪技术是指数型基金管理领域的核心技术, 追踪效果的优劣和基金管理成本的大小通常作为基金管理水平的评价指标, 因此研究具有高精度、低交易成本且能保证追踪组合的高流动性的指数追踪技术对于提高我国指数型基金管理水平的理论价值与现实意义.

指数复制方法依据理论假设的不同可分为两类: 一类是假设历史可以重演的指数复制方法, 主要思路是假设未来追踪期内追踪组合的表现和过去一段时间的表现相同, 因此利用历史数据构造历史追踪误差最小的

收稿日期: 2012-11-30

资助项目: 国家自然科学基金 (71171158); 教育部人文社会科学研究项目 (10YJCZH043, 09XJAZH005); 教育部新世纪优秀人才支持计划 (NCET-10-0646)

作者简介: 胡春萍 (1977-), 女, 汉, 陕西汉中, 讲师, 管理学博士, 研究方向: 绩效评价与金融工程, E-mail: hcp@mail.xjtu.edu.cn; 通讯作者: 薛宏刚 (1974-), 男, 汉, 陕西岐山人, 教授, 博士生导师, 研究方向: 金融工程与风险管理, E-mail: xhg@mail.xjtu.edu.cn.

追踪组合,且认为该组合就是未来追踪期内的最优追踪组合,目前大部分指数复制模型属于这一类;另一类不对追踪组合的未来表现做任何假设,而是寻找与目标指数具有协整关系的股票组合,从协整理论的角度保证追踪组合与目标指数在未来追踪期内的表现与过去一段时间的表现尽可能一致.从复制指数需要的股票数量角度,指数复制技术主要分为完全复制和不完全复制两类.完全复制是指按照目标指数成分证券在目标指数所占权重来购买所有成分证券,构建追踪组合对指数进行追踪.这种方法由于成本高,管理复杂,理论上是不科学的,实践中在指数型基金发展的初期被使用过,但逐渐被不完全复制法替代.不完全复制是指通过利用所有成分证券的子集中包含的证券按照一定比例构成的组合来追踪指数,包括优化复制和分层抽样复制.优化复制是直接利用优化方法确定进入追踪组合内的成分证券及其投资权重,而分层抽样复制则是先按行业、流通市值、换手率等指标人为确定进入追踪组合的证券,再通过优化方法来确定各成分证券的权重.显然,依据经验的分层抽样确定的追踪组合在样本内外追踪效果无法从理论上得到保证,因此,优化复制技术得到了研究者和实际工作者的青睐.

关于指数复制的文献非常丰富,许多文献对指数复制问题的相关文献作了较详尽的综述,如薛宏刚对 2003 年以前国内外关于指数追踪的文献进行了详细分析^[1].近几年以来这方面的研究又有很大进展,Kyong 给出了一个新的分层抽样复制方法,该方法依据成份股的 β 系数,成交量和资本市场化率来确定追踪组合,然后应用作者设计的遗传算法来确定追踪组合的权重^[2].Lobo 考虑了固定交易费用的指数追踪问题,给出了求解该问题的启发式松弛算法^[3],Beasley 将指数追踪技术的测试数据库 OR-Library 从 5 个数据集扩充到 8 个,并给出了一个混合整数规划模型,并考虑了交易费用对追踪效果的影响^[4].Ruiz-Torrubian 设计了一个新的编码方法,基于此给出了新的遗传算法,用来搜索可能的追踪组合,然后利用二次规划确定追踪组合的权重,通过对 OR-Library 中的数据集的测试,表明算法的效率比较高^[5].Barro 给出了一个动态指数追踪问题的随机优化解决方案,在模型中考虑了交易费用和流动性的影响,并对 MSCI Euro index 进行了测试,结果表明动态模型优于静态模型^[6].Wang 将对于追踪组合的风险指标 CVaR 的限制纳入追踪组合的约束中,提出了一个基于 CVaR 风险控制的指数复制模型^[7].Guastaroba 提出了一个核搜索技术用来求解增强型指数复制问题^[8].Scozzari 考虑了 UCITS 交易规则约束下指数复制问题,给出了一个混合二次规划模型来确定最优追踪组合^[9].这些模型均属于假设历史可以重演的指数复制方法类.基于协整理论的指数复制的这几年的文献主要有 Thomas 进行的关于优化复制方法与协整方法的对比,结果表明在短的追踪期限内,优化复制方法的效果优于基于协整理论的指数复制方法,但在一个较长的追踪期内,后者的追踪效果反而更好一些^[10].

从已有文献可以看出,对指数追踪问题的研究成果尽管丰富,但研究角度主要是基于样本内追踪误差(也称经验风险)最小,然后假设市场是有效的,因此认为依据样本内经验风险最小构建的追踪组合在样本外的追踪误差也是最小的.已有的研究结论表明我国证券市场在 1994 年以前是非有效的,此后为弱有效的^[11-14].但事实上有效市场理论近年来受到了行为金融学的挑战,市场是否有效值得怀疑.因此在不对市场有效性做任何假设的前提下,研究如何构建基于样本外视角的指数追踪模型是一个非常有益的课题.另外,金融数据具有典型的时序性,越接近现在的数据对于未来的作用相对于更久远的历史数据更大,因此在指数复制问题中将时间因素考虑进去是非常有必要的.基于此,本文提出了基于时间加权的支持向量机(简称 SVM)的指数追踪模型,并利用 OR-Library 中的 5 个指数测试数据集对提出的模型进行了检验,结果表明该模型具有良好的泛化能力,很大程度上降低了样本外追踪误差,同时提高了模型的鲁棒性.

本文第 2 节建立了基于加权支持向量机的指数优化复制模型,第 3 节利用 OR-Library 中的数据集对提出的模型进行了实证检验,第 4 节总结了本文主要研究结论.

2 基于加权支持向量机的指数优化复制模型

2.1 指数优化复制问题

股票市场指数优化复制问题是指从股票市场上的 N 只股票(一般是目标指数的成份股票,也可能包含部分有代表性的非成份股)中选出 k 只股票构建一个投资组合($k < N$),并使得该组合相对标的指数在未来一段时间内的表现尽可能接近.

构建指数追踪组合首先是选取目标指数和备选股票的历史数据, 并将其分为样本内和样本外, 即 $t = 0, 1, \dots, T$ 的数据为样本内数据, $t = T + 1, \dots, T + L$ 的数据为样本外数据, 用机器学习的语言分别称为训练集和测试集. 然后通过样本内的数据, 依据某一具体的指数复制模型来确定追踪组合. 最后利用样本外数据评价追踪组合的追踪效果.

为了方便, 下面对文中用到的变量给出说明.

参数

p_{it} : 第 i 只股票在 t 时刻的价格;

I_t : 股票指数在 t 时刻的值;

η_i : 如果第 i 只股票进入追踪组合, 它在追踪组合中的最小投资比例;

δ_i : 如果第 i 只股票进入追踪组合, 它在追踪组合中的最大投资比例;

决策变量

c_{it} : 在第 t 个周期内如果第 i 只股票进入追踪组合, 持有股票 i 的数量;

z_i : 判断第 i 只股票是否在追踪组合中的变量, 取 1 表示在追踪组合中, 取 0 表示不在追踪组合中.

2.2 模型构建

第 i 只股票和目标指数在第 t 个周期内的算术收益率分别为

$$r_{it} = \frac{p_{i,t+1} - p_{it}}{p_{it}}, \quad R_t = \frac{I_{t+1} - I_t}{I_t}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N.$$

t 时刻第 i 只股票占追踪组合的总价值的比例为

$$w_{it} = \frac{c_{it} p_{it}}{\sum_{j=1}^N c_{jt} p_{jt}}, \quad t = 1, 2, \dots, T, \quad i = 1, 2, \dots, N.$$

因此追踪组合在第 t 个周期内的算术收益率为

$$R_{pt} = \frac{\sum_{j=1}^N c_{j,t+1} p_{j,t+1} - \sum_{j=1}^N c_{jt} p_{jt}}{\sum_{j=1}^N c_{jt} p_{jt}} = \sum_{j=1}^N w_{jt} r_{jt} = \mathbf{w}_t^T \mathbf{r}_t, \quad t = 1, 2, \dots, T.$$

其中 $\mathbf{w}_t = (w_{1t}, w_{2t}, \dots, w_{Nt})^T$ 表示 N 个备选股票 t 时刻在追踪组合中的比例向量, $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{Nt})^T$ 表示 N 个备选股票第 t 个周期内的收益向量.

得到收益率的表示形式之后, 指数复制可以表示为如下的线性回归问题 (Rudolf^[15])

$$R_t = R_{pt} + \varepsilon_t = \sum_{i=1}^N w_{it} r_{it} + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (1)$$

其中 ε_t 表示追踪组合的收益与目标指数的收益在第 t 个周期内的偏差.

如果追踪组合中第 i 只股票的权重 w_{it} 与时间有关, 那么在各种约束下的线性回归问题的处理是非常困难的. 因此一般有两种简化方法, 一种是采取购买持有策略, 即令 $c_{it} = c_i$, $t = 0, 1, \dots, T - 1$, 从而样本内不调整追踪组合; 另一种是采取固定比例策略, 即令 $w_{it} = w_i$, $t = 0, 1, \dots, T - 1$, 从而在每个时刻 t 调整组合, 使得每一只股票在追踪组合中所占比例保持不变. 采取固定比例策略的优势一是在市场由牛市转为熊市前后调整追踪组合时, 涨幅越大的股票必然超过指数的涨幅, 因此将其持有的比例减小; 反之, 在市场由熊市转为牛市前后调整追踪组合时, 跌幅越大的股票, 持有的比例越大, 这样在市场转向之后追踪组合就会尽可能与股指保持一致; 二是由于固定比例策略是一种追跌杀涨的策略, 因此能够保证追踪组合具有较好的流动性; 三是可以将本文的指数追踪模型简化为一个二次规划模型, 能够有效求解, 因此常采取这种方法. 基于此, 本文在构建追踪组合时也采用固定比例策略, 即 $\mathbf{w}_t = \mathbf{w} = (w_1, w_2, \dots, w_N)^T$, 如此指数复制问题可以看作在训练集 $\{(\mathbf{r}_1, R_1), (\mathbf{r}_2, R_2), \dots, (\mathbf{r}_T, R_T)\}$ 上对下面的 (2) 式进行多元线性回归:

$$R_t = R_{pt} + \varepsilon_t = \sum_{i=1}^N w_i r_{it} + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (2)$$

对于线性回归问题的处理有多种方法, 一个经典的线性回归方法是普通最小二乘回归, 其处理方式是回归的均方误差最小化, 这里指数复制问题的均方误差为 (Torrubiano 定义的追踪误差)

$$MSE = \frac{1}{T} \sum_{t=1}^T (R_{pt} - R_t)^2 = \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^N w_i r_{it} - R_t \right)^2 \quad (3)$$

依据统计学习理论, (3) 式表示了学习问题的经验风险, 其中损失函数为样本内第 t 个周期内追踪组合收益与目标指数收益的偏差的平方

$$L(R_t, R_{pt}) = (R_{pt} - R_t)^2 \quad (4)$$

传统的统计方法认为只要找到能很好拟合已知数据 (训练集) 的函数, 即令“经验风险”最小, 就能保证所得的数学模型预报能力最强. 但这一假设并无严格的理论依据. 统计学习理论已证明: “经验风险”最小不等于数学模型的实际预报风险最小, 因而会导致“过拟合”(overfitting) 现象发生, 普通最小二乘方法 (OLS) 作为传统线性回归的主要方法也存在这个问题.

对于多元线性回归问题, 除了经典的普通最小二乘方法, 目前统计学习理论也提供了一些方法, 一个典型的方法就是基于支持向量机 (SVM) 的线性回归方法. SVM 的主要优点是能够使回归方程的泛化能力 (即模型的预测能力) 得到提高, 因此本文采取结构风险最小化原则, 建立基于支持向量机的指数跟踪模型. 为此首先定义损失函数为下面的 ε 不敏感函数

$$L_\varepsilon(R_t, R_{pt}) = \begin{cases} |R_{pt} - R_t| - \varepsilon, & \text{if } |R_{pt} - R_t| > \varepsilon \\ 0, & \text{else} \end{cases} \quad (5)$$

其中 ε 为一个比较小的正数, 利用它来控制支持向量的数目.

(5) 式表示在第 t 个周期内, 如果追踪组合的收益与目标指数的收益偏差不超过预先指定的小数 ε , 那么不认为追踪组合偏离了目标指数, 即误差忽略不计, 只有这种偏差超过 ε 时, 才将超过 ε 的部分记为追踪组合的偏差.

由于金融数据序列的不稳定和高噪声等特性, 距离现在越近的数据对未来的影响越强烈, 能够提供的信息量也越大, 因此, 在度量经验风险时由远及近给每个时间段的差异赋予逐渐增加的权重是一个有效的时间序列处理方法, 基于此本文建立了基于不等权重支持向量机的指数跟踪模型. 为此首先对不同时间的损失函数进行时间加权, 然后在综合经验风险和预测能力的基础上, 给出下面的统计学习问题的结构风险表达式

$$\frac{1}{2} \|w\|^2 + C \sum_{t=1}^T \lambda_t L_\varepsilon(R_t, R_{pt}) \quad (6)$$

在 (6) 式中, 第一项度量结构风险函数的平坦度, 对它的最小化等于最大化数据点的分隔空间, 也就是最大化回归方程的预测能力. 第二项代表由 ε 不敏感函数决定的经验风险损失, 类似于 (4) 式中的误差平方. 常数 $C (C > 0)$ 是平衡经验风险和预测能力的调节参数, λ_t 是对第 t 个周期内的损失函数的时间加权参数.

关于金融数据时间加重的研究主要包括 Morgan 在其金融风险度量系统 RiskMetrics 中提出 EWMA 模型^[16], 其中利用历史数据的加权来估计方差, 以及 Tay 等在对金融时间序列的支持向量机建模时提出了线性加权和指数加权^[17]. 由于本文主要讨论支持向量机中数据时间加权的问题, 并且 Tay 等指出线性加重的效果不如指数加权方法, 因此下面只给出他们提出的指数加权方法. 在 Tay 等提出的指数加权方法中, 时间权重

$$\lambda_t = \frac{2}{1 + e^{\alpha - 2\alpha t/T}}, \quad t = 1, 2, \dots, T \quad (7)$$

其中 $\alpha \geq 0$ 称为调节参数或折扣率. 图 1 是 $T = 200$ 时, 不同的 α 对应的权重 λ_t 随时间 t 的变化图.

从图 1 可以看到, 在 $\alpha = 0$ 时, 权重 λ_t 为常数 1, 即时间对权重无影响; 当 $\alpha > 0$ 时, 距离现在越近权重越大, 并且 α 的取值越大时间对权重的影响越明显. α 的选取可采用交叉验证方法, Tay 等建议在利用 SVM 做金融数据分析时, α 取 1 比较合适, 因此本文设定 $\alpha = 1$.

在指数复制问题中, 通常需要考虑下面的几个约束 (Ruiz-Torrubiano):

一是资本预算约束:

$$\sum_{i=1}^N w_i = 1 \quad (8)$$

二是投资比例上下界约束:

$$z_i \eta_i \leq w_i \leq z_i \delta_i, \quad i = 1, 2, \dots, N \quad (9)$$

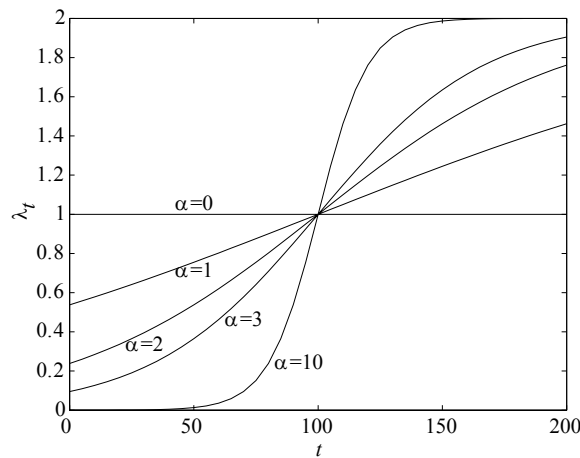


图 1 α 对权重 λ_t 随时间 t 的变化影响

其中下界 η_i 反映了对进入追踪组合的股票的比例的最低要求, 通常取一个比较小的正数, 以体现降低交易成本的要求. 上界 δ_i 反映了对进入追踪组合的股票的比例的最大限制, 通常取小于 1 的正数.

三是对进入追踪组合的股票总数量的限制 (这里总数量 k 是远小于市场上所有可能用来追踪指数的股票数量 N), 主要目的是降低追踪组合的规模, 减小管理成本和难度, 也称势约束:

$$\sum_{i=1}^N z_i = k \quad (10)$$

四是对变量 z_i 的约束:

$$z_i = 0 \text{ or } 1, \quad i = 1, 2, \dots, N \quad (11)$$

当 $z_i = 1$ 时, 表示第 i 只股票进入追踪组合, 否则被排除在外, 该变量的取值体现了选股的过程.

在对变量施加 (8)~(11) 式的约束下, 最小化结构风险 (6) 式, 可得本文提出的基于加权支持向量机的指数优化复制新模型

$$\begin{aligned} & \min \frac{1}{2} \|w\|^2 + C \sum_{t=1}^T \lambda_t L_\varepsilon(R_t, R_{pt}) \\ & \text{s.t.} \quad \sum_{i=1}^N w_i = 1 \\ & \quad z_i \eta_i \leq w_i \leq z_i \delta_i, \quad i = 1, 2, \dots, N \\ & \quad \sum_{i=1}^N z_i = k \\ & \quad z_i = 0 \text{ or } 1, \quad i = 1, 2, \dots, N \end{aligned} \quad (12)$$

3 实证分析

3.1 数据集与评价指标

本节利用国际通用的指数跟踪问题金融数据库 OR-Library 中的数据对本文提出的指数优化复制模型进行实证分析, 通过与 Ruiz-Torrubiano 提出的指数复制模型, 以及不考虑时间加权的基于支持向量机的指数复制模型的实证结果的对比, 可以看出本文提出的基于时间加权的 SVM 的指数优化复制模型无论是样本外的追踪效果, 还是样本内外的一致性都优于前两个模型.

实证分析的数据集包括 OR-Library 中的 Hang Seng 指数 (香港), DAX 指数 (德国), FTSE 指数 (英国), S&P 指数 (美国) 以及 Nikkei 指数 (日本) 及其成份股票 1992 年至 1997 年的 291 个周收盘价构成的 5 个数据集.

在实证分析过程中, 首先利用这 5 个指数和其成份股票的 291 个收盘价得到 290 个几何收益率, 将其中

的前 145 个作为训练集 (样本内), 用于确定跟踪组合, 将后 145 个作为样本外, 再分别计算所确定的跟踪组合在样本内和样本外对指数的跟踪效果.

在样本内和样本外, 用 Ruiz-Torrubiano 模型中的均方误差定义跟踪误差来度量跟踪效果, 将第 i 个数复制模型所得跟踪组合在样本内和样本外的跟踪误差分别为 $InErr_i$ 与 $OutErr_i$.

为比较不同模型的跟踪效果, 本文定义了如下的指标:

指标 1: 模型的鲁棒性

$$Robust_i = |InErr_i - OutErr_i| \quad (13)$$

基金管理者希望在样本内跟踪误差小的跟踪组合在样本外的跟踪误差也小, 也就是跟踪组合在样本内外的跟踪效果尽可能的一致. 该指标反映了一个模型所得跟踪组合在样本内和样本外跟踪效果的一致性, 显然越小越好.

指标 2: 样本外优越性

$$OutSup = \frac{OutErr_j - OutErr_i}{OutErr_i} \quad (14)$$

在指数复制中, 跟踪组合在样本外的跟踪效果是人们最关心的. 这个指标用于比较两个模型在样本外的预测效果, 大于零时表示模型 j 确定的跟踪组合在样本外的跟踪误差比较大, 小于零时表示模型 i 确定的跟踪组合在样本外的跟踪误差比较小, 等于零时表示两个模型确定的跟踪组合在样本外的跟踪效果一样.

指标 3: 超鲁棒性

$$ExRobust = \frac{Robust_i - Robust_j}{Robust_i} \quad (15)$$

该指标用于比较两个模型的鲁棒性的优劣, 大于零时表示模型 j 确定的跟踪组合的鲁棒性优于模型 i , 小于零时表示模型 j 确定的跟踪组合的鲁棒性劣于模型 i , 等于零时表示两个模型确定的跟踪组合的鲁棒性相同.

3.2 实证分析结果

实证分析时模型中设定股票 i 在跟踪组合中所占权重的下界 $\eta_i = 0.01$, 上界 $\delta_i = 0.5$, 在 5 个数据集上本文分别考察了跟踪组合规模 $k = 5, 6, 7, 8, 9, 10$ 六种情形下的跟踪效果.

在求解本文提出的指数复制模型时, 首先利用 Ruiz-Torrubiano 设计的遗传算法对其建立的指数复制模型求解, 确定跟踪组合; 之后选取同样的股票, 用本文提出的模型 (12) 对所选股票重新优化权重最终得到新的跟踪组合. 对于模型 (12) 中的参数 C, ε 以及时间加权参数 α , 本文并没有直接采用交叉验证法确定, 而是考虑到时间顺序的重要性, 采取了下面的顺序验证法确定: 将训练集分成 5 等份, 利用前 4 等份轮流做训练集, 用最后 1 等份做测试集, 取测试误差最小的 C 和 ε . 下面的表 1 列出了不同条件下 C 和 ε 的值.

表 1 时间加权 SVM 指数复制模型中参数 C 和 ε 的值

指数	参数	$k = 5$	6	7	8	9	10
Hang Seng($N = 31$)	C	20	15	20	16	18	18
	ε	1.50E-3	1.50E-3	5.50E-4	1.00E-8	9.00E-4	3.00E-8
DAX($N = 85$)	C	10	10	10	10	3	10
	ε	2.00E-3	5.00E-4	1.50E-5	8.00E-4	5.50E-4	1.50E-4
FTSE($N = 89$)	C	0.2	0.2	0.2	0.1	0.1	0.2
	ε	7.00E-7	2.50E-3	1.00E-3	3.00E-3	1.50E-3	3.00E-3
S&P($N = 98$)	C	0.1	0.5	2	2	0.4	0.3
	ε	1.50E-5	1.50E-5	1.50E-5	2.50E-3	2.50E-3	2.50E-3
Nikkei($N = 225$)	C	20	10	10	1.00E-3	1.00E-3	10
	ε	2.50E-4	2.00E-3	2.50E-4	9.00E-5	2.50E-3	2.50E-4

在确定了参数 C 和 ε 之后, 依据前面定义的三个指标 (12), (13) 和 (14), 对本文模型 (12) 和 Ruiz-Torrubiano 的模型从鲁棒性、样本外的追踪效果两方面进行了对比, 结果见表 2. 为了行文方便, 表 2 中把本文模型 (12) 记为模型 2, Ruiz-Torrubiano 的模型记为模型 1.

从表 2 中的第 3 列和第 6 列可以看到, 对 5 个不同的指数, 无论是本文模型还是 Ruiz-Torrubiano 的模型, 随着跟踪组合的规模 k 的逐渐增大样本内的跟踪误差在逐渐减小. 这种现象的出现是必然的, 因为跟

踪组合规模越大, 构建的跟踪组合包含的指数成份股越多, 从而能更好地复制指数的表现, 使得跟踪误差变小. 由此说明本文建立的指数复制模型是合理的. 另外, 在样本内本文模型的跟踪误差均比 Ruiz-Torrubiano 的模型的大, 这与后者只注重经验风险最小化, 而本文模型牺牲掉一部分样本内的追踪效果是一致的. 从表 2 的 9, 10 两列可以看出, 在总的 30 组跟踪组合中, 本文模型有 23 组的鲁棒性, 20 组的样本外优越性超过了 Ruiz-Torrubiano 的模型, 也就是说本文模型通过结构风险最小化牺牲了一部分样本内的跟踪效果在大部分情况下获得了模型鲁棒性和样本外跟踪效果的提高. 特别是对于 FTSE 指数, 在 6 种不同的跟踪组合规模下模型的鲁棒性和样本外的跟踪效果都比 Ruiz-Torrubiano 的模型明显提高. 因此本文建立的时间加权 SVM 模型与 Ruiz-Torrubiano 的模型相比, 能够获得更好的跟踪组合, 提高了对指数的跟踪效果.

表 2 时间加权 SVM 指数复制模型与 Ruiz-Torrubiano 模型的对比分析

指数	k	$InErr_1$	$OutErr_1$	$Robust_1$	$InErr_2$	$OutErr_2$	$Robust_2$	$ExRobust$	$OutSup$
Hang Seng ($N = 31$)	5	4.135E-5	7.218E-5	3.083E-5	4.295E-5	8.073E-5	3.778E-5	-22.54%	-11.84%
	6	3.031E-5	4.755E-5	1.724E-5	3.149E-5	4.350E-5	1.201E-5	30.33%	8.52%
	7	2.371E-5	3.810E-5	1.439E-5	2.522E-5	3.798E-5	1.276E-5	11.35%	0.32%
	8	1.907E-5	2.899E-5	9.920E-6	2.058E-5	3.021E-5	9.623E-6	2.99%	-4.19%
	9	1.622E-5	2.581E-5	9.590E-6	1.734E-5	2.480E-5	7.466E-6	22.15%	3.91%
	10	1.346E-5	2.057E-5	7.110E-6	1.455E-5	2.002E-5	5.471E-6	23.05%	2.65%
DAX ($N = 85$)	5	2.211E-5	1.018E-4	7.969E-5	2.283E-5	1.021E-4	7.927E-5	0.52%	-0.29%
	6	1.764E-5	8.938E-5	7.174E-5	1.795E-5	8.785E-5	6.989E-5	2.57%	1.72%
	7	1.371E-5	8.459E-5	7.088E-5	1.431E-5	8.416E-5	6.985E-5	1.45%	0.50%
	8	1.110E-5	7.928E-5	6.818E-5	1.182E-5	7.894E-5	6.712E-5	1.55%	0.43%
	9	9.216E-6	7.776E-5	6.854E-5	1.161E-5	8.093E-5	6.486E-5	-1.13%	-4.07%
	10	8.084E-6	7.482E-5	6.674E-5	9.713E-6	7.477E-5	6.506E-5	2.51%	0.07%
FTSE ($N = 89$)	5	6.417E-5	1.581E-4	9.393E-5	6.652E-5	1.526E-4	8.609E-5	8.35%	3.47%
	6	4.961E-5	1.119E-4	6.229E-5	5.573E-5	1.030E-4	4.727E-5	24.11%	7.95%
	7	3.828E-5	9.069E-5	5.241E-5	4.120E-5	8.231E-5	4.111E-5	21.57%	9.24%
	8	2.903E-5	9.662E-5	6.759E-5	3.507E-5	9.369E-5	5.862E-5	13.27%	3.03%
	9	2.486E-5	8.592E-5	6.106E-5	2.709E-5	7.756E-5	5.047E-5	17.34%	9.73%
	10	2.184E-5	8.009E-5	5.825E-5	2.438E-5	7.201E-5	4.764E-5	18.22%	10.08%
S&P ($N = 98$)	5	4.497E-5	1.142E-4	6.923E-5	4.725E-5	1.244E-4	7.713E-5	-11.41%	-8.91%
	6	3.373E-5	1.007E-4	6.697E-5	3.451E-5	1.054E-4	7.084E-5	-5.78%	-4.62%
	7	2.761E-5	7.798E-5	5.037E-5	2.900E-5	7.696E-5	4.796E-5	4.79%	1.31%
	8	2.274E-5	6.764E-5	4.490E-5	2.562E-5	6.243E-5	3.681E-5	18.02%	7.71%
	9	1.939E-5	5.910E-5	3.971E-5	2.180E-5	5.333E-5	3.153E-5	20.61%	9.77%
	10	1.657E-5	5.546E-5	3.889E-5	1.826E-5	5.271E-5	3.445E-5	11.43%	4.96%
Nikkei ($N = 225$)	5	5.456E-5	1.629E-4	1.083E-4	5.734E-5	1.720E-4	1.146E-4	-5.79%	-5.56%
	6	4.008E-5	1.468E-4	1.067E-4	4.099E-5	1.527E-4	1.117E-4	-4.70%	-4.04%
	7	3.356E-5	1.324E-4	9.884E-5	3.718E-5	1.336E-4	9.645E-5	2.42%	-0.93%
	8	2.601E-5	1.100E-4	8.399E-5	2.940E-5	1.039E-4	7.452E-5	11.27%	5.53%
	9	2.125E-5	9.803E-5	7.678E-5	2.749E-5	9.252E-5	6.503E-5	15.30%	5.62%
	10	1.797E-5	6.471E-5	4.674E-5	2.122E-5	6.597E-5	4.475E-5	4.25%	-1.95%

为了考察时间加权的效应, 本文在时间加权 SVM 模型的基础上简化权重系数的确定, 特别地取参数 $\alpha = 0$, 这样权重 $\lambda_t = 1$. 如此所得的无时间加权的 SVM 指数复制模型记为模型 1, 时间加权 SVM 指数复制模型仍记为模型 2. 两个模型在 5 个数据集上的实证分析的结果见表 3.

从表 3 可以看出, 考虑时间加权的模型得到的追踪组合与无时间加权的模型的结果相比, 在 30 组追踪组合中, 尽管只有 15 个组合的样本外追踪误差有所减小, 但是从超鲁棒性的指标来看, 有 21 个组合的鲁棒性得到明显改善, 因此依据样本内数据所得追踪组合对于未来的追踪效果会更加稳定, 这说明了时间加权的 SVM 指数复制模型是优于无时间加权的模型.

表 3 时间加权 SVM 指数复制模型与无时间加权的模型的对比分析

指数	k	$InErr_1$	$OutErr_1$	$Robust_1$	$InErr_2$	$OutErr_2$	$Robust_2$	$ExRobust$	$OutSup$
Hang Seng ($N = 31$)	5	4.25E-05	7.26E-05	3.01E-05	4.30E-05	8.07E-05	3.78E-05	-25.56%	11.21%
	6	3.06E-05	4.32E-05	1.26E-05	3.15E-05	4.35E-05	1.20E-05	4.76%	0.67%
	7	2.46E-05	3.90E-05	1.45E-05	2.52E-05	3.80E-05	1.28E-05	12.00%	-2.72%
	8	1.96E-05	2.86E-05	9.05E-06	2.06E-05	3.02E-05	9.62E-06	-6.38%	5.63%
	9	1.66E-05	2.48E-05	8.22E-06	1.73E-05	2.48E-05	7.47E-06	9.21%	-0.08%
	10	1.36E-05	2.03E-05	6.68E-06	1.46E-05	2.00E-05	5.47E-06	18.05%	-1.33%
DAX ($N = 85$)	5	2.31E-05	1.03E-04	7.97E-05	2.28E-05	1.02E-04	7.93E-05	0.49%	-0.68%
	6	1.80E-05	8.78E-05	6.98E-05	1.80E-05	8.79E-05	6.99E-05	-0.09%	0.06%
	7	1.39E-05	8.41E-05	7.02E-05	1.43E-05	8.42E-05	6.99E-05	0.54%	0.02%
	8	1.16E-05	7.84E-05	6.68E-05	1.18E-05	7.89E-05	6.71E-05	-0.52%	0.69%
	9	9.67E-06	7.79E-05	6.82E-05	1.16E-05	8.09E-05	6.49E-05	4.95%	3.88%
	10	8.17E-06	7.46E-05	6.64E-05	9.71E-06	7.48E-05	6.51E-05	2.06%	0.23%
FTSE ($N = 89$)	5	6.49E-05	1.51E-04	8.62E-05	6.65E-05	1.53E-04	8.61E-05	0.13%	0.99%
	6	5.02E-05	1.10E-04	5.98E-05	5.57E-05	1.03E-04	4.73E-05	20.95%	-6.36%
	7	3.94E-05	8.50E-05	4.57E-05	4.12E-05	8.23E-05	4.11E-05	9.96%	-3.19%
	8	3.03E-05	8.93E-05	5.90E-05	3.51E-05	9.37E-05	5.86E-05	0.66%	4.95%
	9	2.59E-05	7.92E-05	5.33E-05	2.71E-05	7.76E-05	5.05E-05	5.36%	-2.06%
	10	2.26E-05	7.58E-05	5.32E-05	2.44E-05	7.20E-05	4.76E-05	10.37%	-4.94%
S&P ($N = 98$)	5	4.63E-05	1.21E-04	7.50E-05	4.73E-05	1.24E-04	7.71E-05	-2.80%	2.56%
	6	3.42E-05	1.04E-04	6.94E-05	3.45E-05	1.05E-04	7.08E-05	-2.03%	1.74%
	7	2.81E-05	7.90E-05	5.08E-05	2.90E-05	7.70E-05	4.80E-05	5.65%	-2.55%
	8	2.30E-05	6.70E-05	4.40E-05	2.56E-05	6.24E-05	3.68E-05	16.30%	-6.75%
	9	1.99E-05	5.74E-05	3.74E-05	2.18E-05	5.33E-05	3.15E-05	15.79%	-7.06%
	10	1.70E-05	5.30E-05	3.60E-05	1.83E-05	5.27E-05	3.45E-05	4.25%	-0.51%
Nikkei ($N = 225$)	5	5.46E-05	1.63E-04	1.08E-04	5.73E-05	1.72E-04	1.15E-04	-6.21%	5.78%
	6	4.08E-05	1.52E-04	1.11E-04	4.10E-05	1.53E-04	1.12E-04	-0.81%	0.79%
	7	3.50E-05	1.31E-04	9.56E-05	3.72E-05	1.34E-04	9.65E-05	-0.87%	2.30%
	8	2.62E-05	1.09E-04	8.28E-05	2.94E-05	1.04E-04	7.45E-05	10.00%	-4.68%
	9	2.18E-05	9.88E-05	7.70E-05	2.75E-05	9.25E-05	6.50E-05	15.57%	-6.36%
	10	1.85E-05	6.66E-05	4.81E-05	2.12E-05	6.60E-05	4.48E-05	7.00%	-0.89%

4 结论

本文从提高样本外追踪效果和追踪组合鲁棒性的角度, 借鉴机器学习理论中的 SVM 方法, 通过应用指数加权将时间因素引入基于支持向量机的指数追踪问题中, 通过最小化结构风险得到了时间加权 SVM 指数优化复制模型. 并利用国际上标准测试数据集 OR-Library 中的 5 个指数以及成份股历史数据对本文模型进行了实证检验, 并和 Ruiz-Torrubiano 的模型以及无时间加权的 SVM 指数复制模型进行了对比分析. 实证结果表明本文提出的模型具有两个明显的优点: 一是能够提高样本外的追踪效果; 二是能够提高追踪组合在样本内外表现的一致性, 即模型具有良好的鲁棒性, 这一特点说明了本文给出的指数追踪新技术具有较高的理论意义和实际应用价值.

参考文献

- [1] 薛宏刚, 徐成贤, 徐凤敏. 股指期货 —— 投资, 套利与套期保值 [J]. 北京: 科学出版社, 2008.
- [2] Kyong J O, Tae Y K, Sungky M. Using genetic algorithm to support portfolio optimization for index fund management[J]. Expert Systems with Applications, 2005(28): 371-379.
- [3] Lobo M, Fazel M, Boyd S. Portfolio optimization with linear and fixed transaction costs[J]. Annals of Operation Research, 2007, 152(1): 341-365.
- [4] Canakgoz N A, Beasley J E. Mixed-integer programming approaches for index tracking and enhanced indexation[J]. European Journal of Operational Research, 2008, 196: 384-399.

- [5] Ruiz-Torrubiano R, Suarez A. A hybrid optimization approach to index tracking[J]. *Annals of Operation Research*, 2009, 166: 57–71.
- [6] Barro D, Canestrelli E. Tracking error: A multistage portfolio model[J]. *Annals of Operation Research*, 2009, 165: 47–66.
- [7] Wang M H, Xu C X, Xu F M, et al. A mixed 0-1 LP for index tracking problem with CVaR risk constraints[J]. *Annals of Operations Research*, 2012, 196(1): 591–609.
- [8] Guastaroba G, Speranza M G. Kernel search: An application to the index tracking problem[J]. *European Journal of Operational Research*, 2012(217): 54–68.
- [9] Scozzari A, Tardella F, Paterlini S, et al. Exact and heuristic approaches for the index tracking problem with UCITS constrains[R]. 2012, <http://ssrn.com/abstract=2066672>.
- [10] Thomas A M. Cointegration in finance: An application to index tracking[R]. 2008, <http://ssrn.com/abstract=1586997>.
- [11] 乜堪雄. 对我国股票市场有效性的探讨 [J]. *重庆交通学院学报: 社科版*, 2006, 6(1): 76–79.
- [12] 马向前, 任若恩. 基于市场效率的中国股市波动和发展阶段划分 [J]. *经济科学*, 2002(1): 66–72.
- [13] 易荣华, 达庆利. 市场效率计量方法及我国证券市场效率实证研究 [J]. *中国软科学*, 2004(3): 144–156.
Yi Ronghua, Da Qingli. A measuring method of market efficiency and empirical research on market efficiency of Chinese security market[J]. *China Soft Science*, 2004(3): 144–156.
- [14] 张月飞, 史震涛, 陈耀光. 香港与大陆股市有效性比较研究 [J]. *金融研究*, 2006(6): 33–40.
- [15] Rudolf M, Wolter H J, Zimmermann H. A linear model for tracking error minimization[J]. *Journal of Banking & Finance*, 1999, 23(1): 85–103.
- [16] Morgan J P. Risk metrics-technical document [R]. 4th ed. NewYork: Morgan Trust Company, 1996.
- [17] Tay F E H, Cao L J. Modified support vector machines in financial time series forecasting[J]. *Neurocomputing*, 2002(48): 847–861.