DOI:10.13546/j.cnki.tjyjc.2022.06.036

基于层次聚类的投资策略组合优化研究

宁　涛[1,2]

（1.中泰证券股份有限公司博后工作站,济南　250002;2.清华大学　五道口金融学院,北京　100083）

摘　要:投资策略的组合优化是量化交易体系中的重要环节,传统的均值-方差模型难以满足实际需求,文章提出了一种基于层次聚类的风险平价方法,并针对典型中高频趋势策略的组合优化进行实证研究,分析结果表明:基于层次聚类的风险平价方法在样本内外一致性、最大回撤和投资组合分散度等评价指标上都显著优于基于均值-方差的最大夏普比率方法,且样本外风险调整后的收益显著高于一般风险平价方法和最大夏普比率方法。

关键词:投资组合优化;层次聚类;风险平价

中图分类号:F224　　**文献标识码**:A　　**文章编号**:1002-6487(2022)06-0180-05

0　引言

投资组合管理是业界投资应用中频繁用到的工具,Markowitz(1952)[1]提出的均值-方差理论奠定了现代投资管理的基石,均值-方差模型通过求解基于收益率均值和方差的二次规划问题可以得到最优组合配置,然而以该理论为基础的投资组合优化方法在业界的多年实践过程中暴露了一些明显缺点,如预期收益估计不准、投资权重过于集中、样本外表现差等。Michaud(1989)[2]研究发现均值-方差模型对参数估计误差敏感,投资收益估计的小幅误差便会导致优化结果的较大偏差,De Miguel等(2009)[3]的实证研究也表明较大的估计误差会导致均值-方差的样本外绩效跑输简单的等权重分配方法。为了减小收益估计误差给组合优化样本外绩效带来的负面影响,随之衍生出一类仅基于风险的投资组合优化方法,例如最小方差模型(Minimum Variance)和风险平价模型(Risk Parity)[4]。舍弃对收益的预测在一定程度上可以提高求解的稳定性,但是二次规划问题对协方差正定矩阵的求逆还是会引入数值误差,Bailey和Lopez(2012)[5]研究指出,条件数过大的病态协方差矩阵的求解误差较大,特别是当组合中相关性高的资产越多时,组合分散的必要性越大,然而随着相关性高的资产增多,协方差矩阵的条件数也随之增大,导致优化结果的可靠性变差,Lopez(2016)[6]将这种分散投资的必要性越大时均值-方差模型求解越不准确的现象称作“马科维茨诅咒”(Markowitz's Curse),其进一步从图论(Graph Theory)的角度分析造成这一现象的原因:协方差矩阵等价于不同资产之间的完全图(Complete Graph),这种任意资产两两连通的拓扑结构是导致求解不稳定的一个重要原因,对于病态的协方差矩阵,几条边的估计误差即会导致较大的整体求解误差。Simon(1962)[7]和Mantegna(1999)[8]指出金融市场是一个包含多种层次的复杂系统,而协方差矩阵两两连通的拓扑结构并没有刻画出金融资产中的层次结构,Lopez(2016)[6]提出用层次聚类(Hierarchy Cluster)的方法抛弃金融资产间一些不必要的连接来简化图的结构,具体方法是用资产间的相关系数来表征资产间的距离,将资产划分成层次结构并进行准对角化处理,用二分递归方法逐层根据风险平价进行资金权重分配,数值研究表明,层次风险平价方法(Hierarchical Risk Parity)在样本外波动率显著小于最小方差和传统的风险平价方法。此外,Raffinot(2018)[9]的实证研究也表明基于层次聚类的投资组合优化方法求解更稳定、分散性更好,且相比于传统的方法能获得更好的风险调整后收益。

本文针对国内期货市场中的中高频CTA策略着重研究层次聚类方法在策略资金管理中的应用,主要介绍几种组合优化方法,包括传统的最大夏普比率、一般风险平价方法,并从实践角度出发提出一种基于层次聚类的风险平价组合优化方法。

1　投资组合优化方法

1.1　最大夏普比率方法(Maximum Sharpe Ratio, MSR)

最大夏普比率方法是常用的均值-方差模型,其求解如式(1)所示:

$$\max_{\omega} \frac{\omega^T \mu}{\sqrt{\omega^T \Sigma \omega}} \tag{1}$$

$$\text{s.t.} \sqrt{\omega^T \Sigma \omega} = \sigma, L - \omega^T l = 0, \omega \geq 0$$

其中,ω 和 μ 分别为策略的权重和预期收益向量,Σ 是策略收益率的协方差矩阵,$\sqrt{\omega^T \Sigma \omega}$ 是策略组合的标准差,用来衡量策略组合的风险,l 是策略组合的杠杆向量,组合优化的风险设置为 σ,总杠杆设置为 L。

作者简介:宁　涛(1989—),男,云南宣威人,博士,研究方向:数理金融。

1.2　风险平价方法(Equal Risk Contribution,ERC)

一般风险平价方法要求各个子策略对投资组合风险的贡献相同,该方法的求解公式如下:

$$\min_{\omega}\sum_{i=1}^{N}\left(\omega_i-\frac{\omega^T\Sigma\omega}{(\Sigma\omega)_iN}\right)^2 \tag{2}$$

$$\text{s.t.}\sqrt{\omega^T\Sigma\omega}=\sigma,L-\omega^Tl=0,\omega\geq 0$$

和最大夏普比率方法类似,限定策略组合的波动率和杠杆分别为 σ 和 L 。

1.3　基于层次聚类的风险平价方法(Hierarchical Clustering Based Equal Risk Contribution,HCERC)

针对投资组合求解不稳定的问题,采用分层聚类方法降低资产组合间的拓扑复杂度,本文采用层次聚类的方法将策略组合划分为若干子类。本文采用聚合(Agglomerative)方法自下而上聚类:将每个样本各自分为一个类,将距离最近的两类合并,重复此操作直到所有样本聚合为一个类,如此便能得到层次化的聚类结果,再通过"剪枝"可以得到指定聚类数量的结果。层次聚类需要度量不同样本间的距离,本文参考Lopez(2016)[6]的方法,用相关系数来定义两个策略之间的距离,其定义见式(3):

$$d_{i,j}=\sqrt{\frac{1}{2}\left(1-\rho_{i,j}\right)} \tag{3}$$

其中, $\rho_{i,j}$ 是策略 i 和策略 j 的相关系数,可以证明距离 $d_{i,j}$ 严格满足欧几里得距离度量的非负性、同一性、对称性和直递性[6]。不同类之间的距离称为连接(Linkage),本文采用平均连接(Average Linkage)来定义类 G_p 和 G_q 之间的距离,其计算公式如式(4)所示:

$$D_{p,q}=\frac{1}{n_pn_q}\sum_{x_i\in G_p}\sum_{x_j\in G_q}d_{i,j} \tag{4}$$

其中, x_i 和 x_j 分别为类 G_p 和 G_q 中的样本, n_p 和 n_q 分别为类 G_p 和 G_q 中的样本个数。

本文提出一种兼顾风险和收益的基于层次聚类的风险平价方法,详细步骤如下:

(1)子类中的组合优化方法

整个策略组合的风险主要通过不同子类做分散,对于同一类中相关性较高的策略,考虑以收益风险比为基础调整不同策略的风险额度,即给收益风险比高的策略分配更多的风险预算,给收益风险比低的策略分配更少的风险预算。

传统的夏普比率并没有衡量投资组合的下行风险,且收益率的方差不能精确衡量投资组合在时间序列上的风险特征,投资实践中往往更关心时间序列相关的下行风险,业界常使用投资组合的最大回撤(Maximum Drawdown,MDD)来衡量下行风险,用复合年增长率(Compound Annual Growth Ratio,CAGR)来衡量收益,用卡玛比率(Calmar Ratio,CR)来衡量收益风险比,其定义如式(5)所示:

$$CR=\frac{CAGR}{MDD} \tag{5}$$

但是传统的CAGR和MDD对测试区间的选取较为敏感,本文参考Faith(2007)[10]的研究用更稳健的指标回归年度回报率(Regressed Annual Return,RAR)和长度调整平均最大回撤(Length Adjusted Average Maximum Drawdown,LAAMD)来衡量收益和风险。对投资组合累积收益和时间进行线性回归计算得到回归斜率,通过回归斜率和回测时间区间拟合新的收益曲线,计算得到RAR。此外,用来表征风险的LAAMD指标定义如下:

$$LAAMD=\left(\frac{1}{5}\sum_{i=1}^{5}drawdown_i^{\max}\right)\left(\frac{\frac{1}{5}\sum_{i=1}^{5}length_i^{\max}}{365}\right) \tag{6}$$

其中, $drawdown_i^{\max}$ ($i=1,2,\cdots,5$)表示前五次最大回撤, $length_i^{\max}$ ($i=1,2,\cdots,5$)表示前五次最大回撤对应的回撤持续时间。基于更加稳健的回归年度回报率和长度调整平均最大回撤,可以构建更稳健的收益风险比指标R立方(Rcubic),具体如下:

$$Rcubic=\frac{RAR}{LAAMD} \tag{7}$$

基于风险调整后的收益指标R立方对子类中不同策略的风险预算进行分配,具体计算如式(8)所示:

$$\min_{\omega^p}\sum_{i=1}^{N^p}\left(\omega_{p,i}-\frac{\omega_p^T\Sigma\omega_p}{\left(\Sigma\omega_p\right)_i}\cdot\frac{Rcubic_i^{normal}}{\sum_{j=1}^{N_p}Rcubic_j^{normal}}\right)^2 \tag{8}$$

$$\text{s.t.}\sqrt{\omega_p^T\Sigma\omega_p}=\sigma,L-\omega_p^Tl=0,\omega\geq 0,p=1,2,\cdots,M$$

其中, ω_p 是子类 p 的权重向量,投资组合中共有 M 个子类, N^p 是子类 p 中的策略总数, $Rcubic_i^{normal}$ 是策略 i 标准化后的R立方指标,其计算见式(9), α 用来调整策略组合中风险预算分配的离散程度, α 越大则风险预算在子类中的分配差异就越大, α 为0时等价于一般的风险平价方法。

$$Rcubic_i^{normal}=1-\alpha+2\alpha\frac{Rcubic_i-\min(Rcubic)}{\max(Rcubic)-\min(Rcubic)},$$

$$\alpha\geq 0,\ i=1,2,\cdots,N \tag{9}$$

(2)子类间的组合优化方法

不同子类组合策略之间的权重分配用一般的风险平价方法,具体如式(10)所示:

$$\min_{\omega}\sum_{i=1}^{M}\left(\omega_i-\frac{\omega^T\Sigma\omega}{(\Sigma\omega)_iM}\right)^2 \tag{10}$$

$$\text{s.t.}\ \sqrt{\omega^T\Sigma\omega}=\sigma,L-\omega^Tl=0,\omega\geq 0$$

2　投资策略构建

2.1　数据选取

本文针对国内四个期货交易所流动性较好的30个期货品种展开研究,具体见下页表1。本文的测试数据为消

除换月跳空之后的主力连续合约数据(在换月节点做差值前复权处理),覆盖的时间范围为2015年5月4日至2020年4月3日。

表1　　期货品种列表

交易所	期货品种
中国金融期货交易所(CFFEX)	中证500股指期货(IC)、沪深300股指期货(IF)、上证50股指期货(IH)、5年期国债期货(TF)、10年期国债期货(T)
上海期货交易所(SHFE)	铝(AL)、石油沥青(BU)、铜(CU)、镍(NI)、铅(PB)、螺纹钢(RB)、天然橡胶(RU)、锡(SN)、锌(ZN)
大连商品交易所(DCE)	黄大豆1号(A)、玉米淀粉(CS)、焦炭(J)、鸡蛋(JD)、焦煤(JM)、聚乙烯(L)、棕榈油(P)、聚丙烯(PP)、豆油(Y)
郑州商品交易所(CZCE)	棉花(CF)、玻璃(FG)、甲醇(MA)、菜籽油(OI)、菜籽粕(RM)、白糖(SR)、精对苯二甲酸(TA)

2.2 投资策略类的构建方法

本文通过测试不同品种、不同类型的策略给后文的投资组合优化提供足够的研究样本,为了更接近业界的实际应用场景,本文测试的策略主要是趋势策略类型,在指标类型上涵盖常见的均线、通道两种形式的趋势策略,在策略持仓频率类型上包括日内和日间两种策略类型,具体如表2所示,不同策略的具体构建方法见下文。

表2　　策略类型

策略	指标类型	持仓类型	回测数据粒度
指数移动双均线日内策略(EMA Intraday Strategy)	均线	日内	1分钟
开盘区间突破策略(Open Range Breaker Strategy)	突破	日内	1分钟
指数移动双均线日间策略(EMA Strategy)	均线	日间	10分钟
波动率通道突破策略(ATR Channel Breaker Strategy)	突破	日间	10分钟

2.2.1 指数移动双均线日内策略

指数移动双均线日内策略是均线类型策略,盈利来源是日内的较大波动率,其参数有两个:短期均线 $length_{fast}$ 和长期均线 $length_{slow}$,迭代计算过程见式(11)和式(12)。

$$\begin{aligned}&ema_{fast,1}=close_1\\&\alpha_{fast}=\frac{2}{length_{fast}+1}\\&ema_{fast,t}=\left(1-\alpha_{fast}\right)\cdot ema_{fast,t-1}+\alpha_{fast}\cdot close_t\ (t=2,\cdots,T)\end{aligned}\tag{11}$$

$$\begin{aligned}&ema_{slow,1}=close_1\\&\alpha_{slow}=\frac{2}{length_{slow}+1}\\&ema_{slow,t}=\left(1-\alpha_{slow}\right)\cdot ema_{slow,t-1}+\alpha_{slow}\cdot close_t\ (t=2,\cdots,T)\end{aligned}\tag{12}$$

其中,α_{fast} 和 α_{slow} 代表短期均线和长期均线的指数加权权重系数,$length_{fast}$ 和 $length_{slow}$ 分别是短期均线和长期均线的时间周期,ema_{fast} 和 ema_{slow} 分别是短期均线和长期均线计算得到的指数加权均值。

交易信号的计算规则如下:

(1) $ema_{fast,t-1}\leqslant ema_{slow,t-1}$ 且 $ema_{fast,t}>ema_{slow,t}$,若策略在 t 时刻无持仓,则触发买入信号;若策略在 t 时刻持多仓,则维持仓位不变;若在 t 时刻为持空仓,则触发平仓和买入信号,平空头仓位并开多头仓位。

(2) $ema_{fast,t-1}\geqslant ema_{slow,t-1}$ 且 $ema_{fast,t}<ema_{slow,t}$,若策略在 t 时刻无持仓,则触发卖出信号;若策略在 t 时刻持空仓,则维持仓位不变;若在 t 时刻为多头持仓,则触发平仓和卖出信号,平多头仓位并开空头仓位。

(3)收盘前进行平仓操作以保证策略不留有隔夜仓位,具体实现为:在每个交易日的倒数第二根K线检查策略的持仓状态,若有持仓,则做平仓操作。

2.2.2 开盘区间突破策略

本文用5日指数加权日均波幅(Exponential Moving Average True Range,EMATR)作为当前交易日的波动幅值估计,设置一定倍数的波动率作为开平仓基准,该策略有四个参数:多头开仓阈值(long Entry Num)、空头开仓阈值(short Entry Num)、多头平仓阈值(long Exit Num)、空头平仓阈值(short Exit Num)。策略决策中的有关指标计算如式(13)至式(16)所示:

$$upperBand_{entry,i}=open_i+longEntryNum\cdot EMATR_i,\quad i=1,2,\cdots,T\tag{13}$$

$$lowerBand_{entry,i}=open_i-shortEntryNum\cdot EMATR_i,\quad i=1,2,\cdots,T\tag{14}$$

$$upperBand_{exit,i}=upperBand_{entry,i}-longExitNum\cdot EMATR_i,\quad i=1,2,\cdots,T\tag{15}$$

$$lowerBand_{exit,i}=lowerBand_{entry,i}+shortExitNum\cdot EMATR_i,\quad i=1,2,\cdots,T\tag{16}$$

其中,$open_i$ 表示第 i 日的开盘价,$EMATR_i$ 表示用第 i 个交易日之前日波动计算得到的指数加权日均波幅,$upperBand_{entry,i}$ 和 $lowerBand_{entry,i}$ 分别是多头和空头的开仓阈值,$upperBand_{exit,i}$ 和 $lowerBand_{exit,i}$ 分别是多头和空头的平仓阈值,在单个交易日内这些开、平仓阈值保持不变。

开盘区间突破策略的交易信号计算规则具体如下:

(1) $close_{t-1}\leqslant upperBand_{entry,t-1}$ 且 $close_t>upperBand_{entry,t}$,若策略在 t 时刻无持仓,则触发买入信号;若策略在 t 时刻持多仓,则维持仓位不变;若在 t 时刻为持空仓,则触发平仓和买入信号,平空头仓位并开多头仓位。

(2) $close_{t-1}\geqslant lowerBand_{entry,t-1}$ 且 $close_{t-1}<lowerBand_{entry,t-1}$,若策略在 t 时刻无持仓,则触发卖出信号;若策略在 t 时刻持空仓,则维持仓位不变;若在 t 时刻为多头持仓,则触发平仓和卖出信号,平多头仓位并开空头仓位。

(3) $close_{t-1}\geqslant upperBand_{exit,t-1}$ 且 $close_{t-1}<upperBand_{exit,t-1}$,若策略在 t 时刻为多头持仓,则触发平仓信号,卖平多仓。

(4) $close_{t-1}\leqslant lowerBand_{exit,t-1}$ 且 $close_{t-1}>lowerBand_{exit,t-1}$,若策略在 t 时刻为空头持仓,则触发平仓信号,买平空仓。

2.2.3 指数移动双均线日间策略

指数移动双均线的日间策略和日内策略的指标计算公式是相同的,此处不再赘述,两者的区别是日间策略在收盘前不做日内平仓操作,指数移动双均线日间策略的收益来源是较大的日间波动率。

2.2.4 波动率通道突破策略

波动率通道突破策略的基本思路是以均线为中轨，以一定倍数的波动率确定价格通道的上下轨，通过价格对通道的突破来确定买卖信号，是典型的日间趋势跟踪策略。本文以前5个交易日的日均波幅（Average True Range，ATR）作为当前交易日波动的估计，该策略有两个控制参数：均线周期（*length*）、ATR倍数（*AtrNum*）。策略信号计算相关的指标如下：

$$upperBand_i = close_i + AtrNum \cdot ATR_i, i = 1, 2, \cdots, T \quad (17)$$

$$lowerBand_i = close_i - AtrNum \cdot ATR_i, i = 1, 2, \cdots, T \quad (18)$$

波动率通道突破策略的交易信号计算规则具体如下：

（1）$close_{t-1} \leqslant upperBand_{t-1}$ 且 $close_t > upperBand_i$，若策略在 t 时刻无持仓，则触发买入信号；若策略在 t 时刻持多仓，则维持仓位不变；若在 t 时刻为持空仓，则触发平仓和买入信号，平空头仓位并开多头仓位。

（2）$close_{t-1} \geqslant lowerBand_{t-1}$ 且 $close_{t-1} < lowerBand_{t-1}$，若策略在 t 时刻无持仓，则触发卖出信号；若策略在 t 时刻持空仓，则维持仓位不变；若在 t 时刻为多头持仓，则触发平仓和卖出信号，平多头仓位并开空头仓位。

2.3 开仓手数的计算方法

参考Faith（2007）[10]用交易品种的平均真实波幅来衡量交易风险的做法，本文也采用指数移动平均真实波幅（Exponential Moving Average True Range，EMATR）来作为波动率的度量指标，用TR的5日指数加权均值来表示指数移动平均真实波幅，每个交易日后更新EMATR，在交易日中EMATR的值不更新。开仓手数的计算如式(19)所示：

$$lots = \frac{initialCapital \cdot riskRate}{EMATR \cdot size} \quad (19)$$

其中，*initialCapital* 为初始本金，本文设置为固定值1000万，*riskRate* 表示各个策略的风险额度比例，本文统一设置为2%，*size* 表示期货品种的合约乘数，用 $EMATR \cdot size$ 来表征交易1手期货合约的风险。

2.4 回测参数的设置

回测的时间范围是2015年5月4日到2020年4月3日，且回测时按每笔交易的开始结束时间来区分平今和平昨，不考虑日内锁仓，即日内的平仓操作均按平今处理，计入交易所的实际开平仓手续费，暂不考虑下单算法的优化，假设每笔交易的成交价为触发信号时刻的中间价 $(BidPrice\_1 + AskPrice\_1)/2$。

3 策略投资组合的实证

针对上文中的四种策略和30个期货品种，本文共生成120个子策略作为投资组合优化的输入数据，下文对三种投资组合方法进行实证研究。其中，从2016年1月开始构建投资组合，预留半年历史数据做投资组合计算的“预热”，每间隔二十个交易日做一次策略权重的重新计算。为避免未来数据给组合绩效的统计带来前视偏差（Look-ahead bias），计算涉及的收益率和协方差矩阵仅采用再分配时间点之前的历史数据进行计算。

3.1 策略投资组合的评价指标

本文为了具体比较不同策略投资组合方法的差异，对如下指标进行对比分析：信息比率（Information Ratio，IR）、投资组合权重平方和（the Sum of Squared Portfolio Weights，SSPW）、权重再平衡的平均换手率（the Average Turnover Per Rebalancing，ATPR）、夏普比率（Sharpe Ratio，SR）、卡玛比率（Calmar Ratio，CR）、R立方（R-cubic）。

3.1.1 信息比率

本文对策略组合权重按一定的时间间隔进行优化，并进行动态分配。为了比较策略投资组合样本内和样本外的差别，参考Grinold和Kahn（2014）[11]的信息比率，以投资组合样本外的绩效为比较基准，构造样本内绩效相对于样本外绩效的超额收益，进而对比两种方法的优劣，计算方法如下：

$$r_{volume-bars} = \beta \cdot r_{time-bars} + r_{\varepsilon} \quad (20)$$

$$IR = \frac{E(r_{\varepsilon})}{\sigma(r_{\varepsilon})} \quad (21)$$

其中，$r_{volume-bars}$ 和 $r_{time-bars}$ 分别代表样本内绩效和样本外绩效的收益率序列，r_{ε} 是残差收益率序列，$E(r_{\varepsilon})$ 和 $\sigma(r_{\varepsilon})$ 分别表示残差收益率期望和残差风险。信息比率越大，样本外相对于样本内表现就越差。

3.1.2 投资组合权重平方和

参考Goetzmann和Kumar（2008）[12]对投资组合分散度的衡量方法，本文用投资权重平方和来表征投资组合的分散程度，其定义如下：

$$SSPW = \frac{1}{F}\sum_{t=1}^{F}\sum_{i=1}^{N}\omega_{i,t}^{2} \quad (22)$$

其中，$\omega_{i,t}$ 是第 t 次再平衡时策略 i 的配置权重，$\sum_{i=1}^{N}\omega_{i,t}^{2}$ 表示第 t 次进行策略权重再平衡时的投资组合权重平方和，组合调整的再平衡总次数为 F。$SSPW$ 的取值范围为0~1，1对应着最集中的投资组合，$SSPW$ 越小，投资组合就越分散。

3.1.3 组合再平衡的平均换手率

投资组合权重的再平衡调整会额外增加交易逻辑之外的换手率，参考Raffinot（2018）[9]的方法定义组合再平衡的换手率如下：

$$ATPR = \frac{1}{F}\sum_{t=2}^{F}\sum_{i=1}^{N}\left|\omega_{i,t} - \omega_{i,t-1}\right| \quad (23)$$

其中，$\left|\omega_{i,t} - \omega_{i,t-1}\right|$ 代表第 t 次再平衡时策略 i 的换手率，组合调整的再平衡总次数为 F，需要做组合配置的策略总数为 N。

3.1.4 夏普比率

夏普比率是用来衡量收益风险比的常用指标，其定义如下：

$$SR = \frac{\mu(r_p)}{\sigma(r_p)} \quad (24)$$

其中，投资组合收益率序列的均值 $\mu(r_p)$ 用作衡量收益，投资组合收益率序列的标准差 $\sigma(r_p)$ 用作衡量风险。

3.2 实证结果分析

在杠杆和波动一定的情况下对比分析不同的策略投资组合优化方法，限制组合的总体杠杆不大于3，并设定年化波动为10%，针对基于层次聚类的风险平价方法HCERC，在子类中按R立方分配风险时设置系数 α 为0.5，聚类的数量 M 设置为20。策略组合的协方差矩阵的条件数影响输出结果的稳定性，协方差矩阵的条件数定义见式(25)，其中 λ 是协方差矩阵 Σ 对应的特征值向量。

$$conditionNum=\max(\lambda)/\min(\lambda) \tag{25}$$

本文每隔20天做一次组合优化，一共做 F 次策略权重的再平衡，所有策略集合的平均条件数的定义如下：

$$aveConditonNumber=\frac{1}{F}\sum_{i=1}^{F}conditionNum_i \tag{26}$$

本文基于层次聚类的风险平价方法分两个层次进行组合优化，第一层是对子类中的策略集合进行组合优化，第二层是对第一层中每个类的组合策略进行再次组合。针对第二层的组合优化可以统计子类平均条件数如下：

$$aveConditonNumber_{cluster}=\frac{1}{F\cdot N_{cluster}}\sum_{i=1}^{F}\sum_{j}^{N_{cluster}}conditionNum_{i,j} \tag{27}$$

其中，$N_{cluster}$ 是策略的聚类数量，$conditionNum_{i,j}$ 为第 i 次组合优化时子类 j 的条件数。

本文研究的所有策略集合的平均条件数 $aveConditonNumber$ 的统计值为175，子类的平均条件数 $aveConditonNumber_{cluster}$ 的统计值为46，HCERC在第二层子类间的组合优化的数值特性有明显的改善。不同投资组合方法的评价指标对比见表3，可以看到MSR的信息比率IR高达14.1，即样本外的表现显著差于样本内，样本内外的巨大差异会导致实际应用中优化结果的可信度较低，优化输出的稳定性较差，而HCERC在考虑收益的情况下信息系数IR仅为0.4，仅考虑风险的ERC的IR为0.2，HCERC和ERC的样本内外表现基本保持一致，优化结果鲁棒性更好，且HCERC和ERC的最大回撤也小于MSR。实践中均值-方差模型被人诟病的另外一个原因是资产分配的集中度较高，较为集中的持仓会削弱组合在极端市场环境下的抗风险能力，由投资组合权重平方和SSPW的统计结果可知，ERC和HCERC的策略分散效果显著优于MSR。此外，从夏普比率、卡玛比率和R立方这三个衡量收益风险比的指标统计结果可以看到，HCERC能获得最高的收益风险比，ERC也略好于MSR。

表3 不同投资组合方法的评价指标对比

	IR	SSPW	ATPR	SR	MDD	CR	RAR	Rcubic
MSR	14.1	0.24	0.42	1.30	13.5%	0.99	6.4%	2.13
ERC	0.2	0.05	0.61	1.34	10.2%	1.28	5.9%	2.17
HCERC	0.4	0.08	0.57	1.66	9.8%	1.65	9.4%	4.78

4 结论

投资策略的组合优化是量化交易体系中不可或缺的环节，而传统的均值-方差模型难以满足实战应用的需求，本文从实际应用的角度出发，提出了一种基于层次聚类的风险平价方法：先通过层次聚类对策略集合进行分类，然后通过两个层次进行组合优化以兼顾收益和风险，第一步在子类中考虑收益风险比分配风险预算，第二步用风险平价做子类间的组合优化。本文针对国内期货市场上的典型中高频趋势策略的组合优化进行了实证分析，研究表明：基于层次聚类的风险平价方法在样本内外的表现基本保持一致，求解的鲁棒性、风险控制和投资分散度都显著优于基于均值-方差的最大夏普比率方法；此外，一般风险平价方法样本外的风险调整后收益优于最大夏普比率方法，而基于层次聚类的风险平价方法能够得到最高的样本外收益风险比，对实践投资有更好的指导意义。

参考文献：

[1]Markowitz H. Portfolio Selection [J].Journal of Finance,1952,7(1).

[2]Michaud R. The Markowitz Optimization Enigma: Is Optimized Optimal [J].Financial Analysts Journal,1989,45(1).

[3]De Miguel V, Garlappi L, Uppal R. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? [J].The Review of Financial Studies,2009,22(5).

[4]Roncalli T. Introduction to Risk Parity and Budgeting [M].London: Chapman & Hall,2013.

[5]Bailey D, Lopez P M. Balanced Baskets: A New Approach to Trading and Hedging Risks [J].Journal of Investment Strategies,2012,1(4).

[6]Lopez P M. Building Diversified Portfolios That Outperform Out of Sample [J].Journal of Portfolio Management,2016,42(4).

[7]Simon H A. The Architecture of Complexity [J].Proceedings of the American Philosophical Society,1962,106(6).

[8]Mantegna R N. Hierarchical Structure in Financial Markets [J].The European Physical Journal B-Condensed Matter and Complex Systems,1999,11(1).

[9]Raffinot T. Hierarchical Clustering-based Asset Allocation [J].Journal of Portfolio Management,2018,44(2).

[10]Faith C. Wayofthe Turtle: The Secret Methods That Turned Ordinary People Into Legendary Trader [M].Singapore: McGraw-Hill,2007.

[11][美]格林诺德,[美]卡恩.主动投资组合管理:创造高收益并控制风险的量化投资方法(原书第2版)[M].李腾,杨柯敏,刘震,译.北京:机械工业出版社,2014.

[12]Goetzmann W N, Kumar A. Equity Portfolio Diversification [J].Review of Finance,2008,12(3).

（责任编辑/刘柳青）