

双目标遗传规划应用于行业轮动

华泰研究

2024年5月20日 | 中国内地

深度研究

人工智能 79：应用双目标遗传规划构建周频行业轮动策略

本文提出了双目标遗传规划模型，并将其应用于周频行业轮动。自2022Q3电力设备及新能源行业的行情结束以来，市场主线强度转弱、行业轮动速度加快，月频行业轮动模型难以持续稳定地获取超额收益，构建周频行业轮动策略是应对良策。双目标遗传规划同时以 $|IC|$ 和 $NDCG@k$ 作为适应度函数来评价因子有效性，挖掘出了一批兼顾单调性和多头组表现的因子。我们使用贪心策略将其合成为综合因子。2022-09-30至2024-04-30，综合因子的扣费前年化超额收益为25.74%，夏普比率1.70，最大回撤-21.47%，显著优于单目标遗传规划模型的表现；年化换手为单边约13倍。

双目标遗传规划在进化过程中更易保持种群多样性

遗传规划流程主要包括交叉变异、适应度评估、子代选择等三个步骤，能否维持种群多样性是因子挖掘能否成功的关键。单目标遗传规划不仅因子评价维度不全面，还面临严重的种群拥挤问题。与单目标遗传规划相比，双目标遗传规划在适应度评估阶段选取两个侧重点不同的评价指标来评价因子有效性，在子代选择阶段采用NSGA-II算法，形成了双目标的对抗，种群拥挤的速度自然会慢得多。此外，我们还做了如下努力来尽可能保持种群多样性——在设置种群规模和进化次数的时候要求前者远大于后者，同时将初始种群分为若干小种群，让每一个小种群单独执行遗传规划流程。

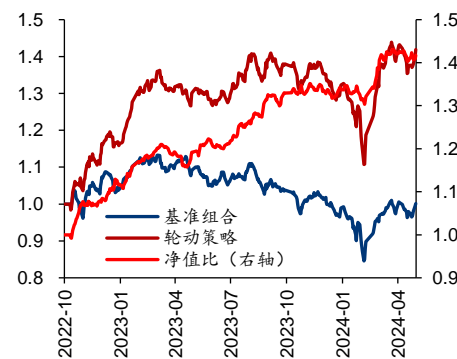
双目标遗传规划模型的表现显著优于单目标遗传规划模型

同时将 $|IC|$ 和 $NDCG@k$ 作为适应度函数来评价因子有效性。其中， $|IC|$ 侧重因子单调性， $NDCG@k$ 侧重多头组表现。考虑到本研究可以拓展至选股场景，故遗传规划算子使用PyTorch实现便于GPU加速，共计64个。我们每隔3个月重新训练模型。每个重训练日，为了降低随机因素的干扰，设置了6个随机数种子执行了6轮因子挖掘，并使用贪心策略将6轮因子挖掘的结果合成为周频行业轮动综合因子。消融实验仅以 $|IC|$ 或 $NDCG@k$ 作为适应度函数开展单目标遗传规划，其他参数细节均保持不变。在相同的回测区间内，双目标遗传规划模型的表现显著优于单目标遗传规划模型。

风险提示：遗传规划在滚动窗口中挖掘历史规律，规律可能在下次重新挖掘之前失效；月频和周频行业轮动模型都有其适用的市场条件，无法保证在任何市场条件下均可取得超额收益；涉及的具体行业不代表任何投资意见，请投资者谨慎理性地看待。

研究员	林晓明
SAC No. S0570516010001	linxiaoming@htsc.com
SFC No. BPY421	+(86) 755 8208 0134
研究员	徐特, PhD
SAC No. S0570523050005	xute@htsc.com
	+(86) 21 6321 1166
研究员	何康, PhD
SAC No. S0570520080004	hegang@htsc.com
SFC No. BRB318	+(86) 21 2897 2039
联系人	卢炯
SAC No. S0570123070272	lujiong@htsc.com
	+(86) 755 8249 2388

双目标遗传规划周频行业轮动净值



注：回测区间为2022-09-30至2024-04-30，底层资产为32个中信行业指数，每周第一个交易日调仓；基准组合为行业等权周度再平衡组合；不考虑交易费用。
资料来源：Wind，华泰研究

正文目录

本文导读.....3

 背景一：市场主线强度不高，月频行业轮动模型表现不稳定.....3

 背景二：用选股因子合成行业轮动因子其有效性无法被保证.....6

双目标遗传规划原理详解.....8

 单目标遗传规划面临的痛点.....8

 NSGA-II 算法重要概念.....9

 遗传规划算子介绍.....10

双目标遗传规划应用于行业轮动.....12

 适应度函数定义.....12

 因子挖掘流程.....13

 实证：同时将|IC|和 NDCG@k 作为适应度函数.....15

 消融实验：仅以|IC|或 NDCG@k 作为适应度函数.....16

 更多思考.....16

 参考文献.....17

 风险提示.....17

附录：重训练日挖掘出来的信息比率前十名因子一览.....18

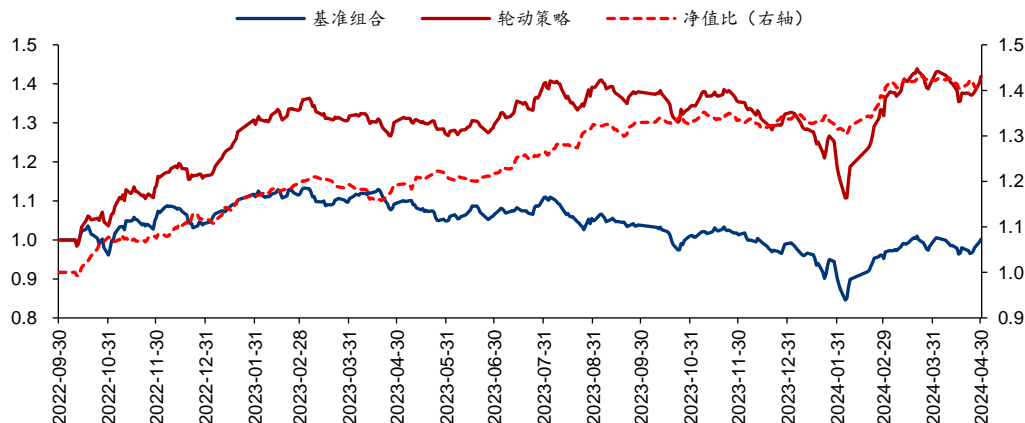
本文导读

物竞天择，适者生存。

行业轮动一直是 A 股投资一个重要的研究课题。大多数买方和卖方机构都布局了由包括基本面和技术面组成的月频行业轮动模型，其中基本面一般起主导作用。但近两年，由于投资者信心不足、市场主线不强、行业轮动速度较快，鲜有月频模型能够持续稳定地获取超额收益，我们的模型也不例外。于是，部分投资者开始着眼周频行业轮动的研究，其中技术面一般起主导作用。然而，根据我们的经验，像 MACD、KDJ 等常见技术指标在周频行业轮动场景中几乎全部失效，寻求人工智能(AI)等高级因子挖掘手段，成了新的探索方向。

由于行业轮动场景具有截面小样本的特点，大家普遍担心 AI 模型潜在的过拟合风险，于是使用选股因子合成行业轮动因子似乎顺理成章。但是，这种做法得到的行业轮动因子是否有效似乎“听天由命”，而且由于要先得到选股因子，工程计算量通常不低。那么，是否有办法直接基于各行业指数的量价数据来挖掘行业轮动因子呢？容易想到的方式包括端到端有监督学习和遗传规划等启发式因子挖掘。对于前者，我们暂时没有取得有效进展。本文将重点介绍我们在遗传规划上取得的成果。全文回测均使用前期报告《经济周期实证、理论及应用》(2024-01-10)使用的中信行业指数，每次调仓选信号前 5 名的行业作为多头组等权配置。遗传规划模型的净值如下图所示：

图表1：遗传规划周频行业轮动模型净值

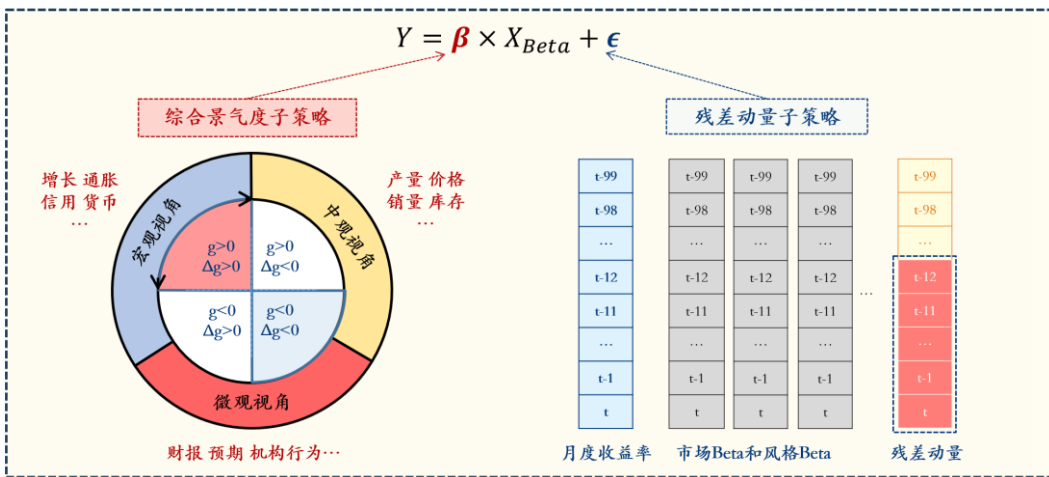


资料来源：Wind，华泰研究

背景一：市场主线强度不高，月频行业轮动模型表现不稳定

我们的月频行业轮动模型从因子定价模型出发。一个行业的表现可以被拆分为由市场和风格 Beta 所解释的部分以及残差。针对 Beta 部分，我们撰写了一系列有关行业景气度的研究报告，从宏观、中观、微观三个视角去捕捉行业盈利能力 g 和盈利能力边际变化 Δg 都大于零的机会，详见前期报告《行业景气投资的顶层设计和落地方案》(2023-09-14)和我们每周发布的观点周报。残差部分是无法被市场和风格 Beta 所解释的部分，可能蕴含着如技术进步、政策支持等行业专属信息。我们采用技术分析的方法构建了残差动量因子来捕捉行业专属信息的异动，详见前期报告《行业残差动量定价能力初探》(2024-02-05)。从统计学原理来说，Beta 部分和残差部分是不相关的；根据实证，综合景气度子策略和残差动量子策略也几乎不相关。因此，我们最后把两个子策略按信号进行等权融合。

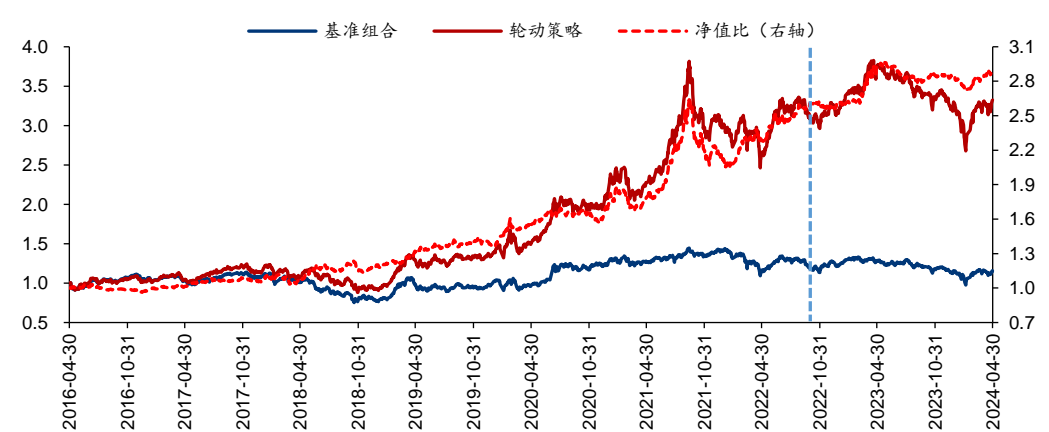
图表2：华泰金工月频行业轮动模型框架



资料来源：华泰研究

取回测区间 2016-04-30 至 2024-04-30，每月末选择综合得分前五名的中信行业指数等权配置，次月第一个交易日按收盘价完成调仓。结果显示，自从 2022Q3 电力设备及新能源行业的行情结束后，月频行业轮动模型的表现就下了一个台阶。2022 年 9 月之前，模型相对于全体行业等权基准的年化超额收益约为 17%；而 2022 年 9 月之后，年化超额收益就下降至了约 7%。而且不难发现，2022 年 9 月之后相较 2022 年 9 月之前，月频行业轮动模型的换手率提升了约 40%，表明一个行业投资机会的持续时间变短了。这可能是由于彼时美联储正在快速加息，全球投资者的风险偏好迅速降低，即使基本面还未被股价充分兑现，更多投资者的心态也可能是“落袋为安”去寻找下一个投资机会，而不愿意对股价可能继续上涨“下重注”，导致市场交易缺乏主线、行业轮动速度较快。

图表3：华泰金工月频行业轮动模型净值



资料来源：Wind，华泰研究

图表4：月频行业轮动模型业绩：2022 年 9 月之前

回测区间：2016-04-30 至 2022-09-30	年化收益	年化波动	夏普比率	最大回撤	卡玛比率	年化换手
月频行业轮动	19.72%	24.17%	0.82	-35.43%	55.65%	单边 3.1 倍
行业等权基准	2.72%	19.19%	0.14	-33.94%	8.02%	/

资料来源：Wind，华泰研究

图表5：月频行业轮动模型业绩：2022 年 9 月之后

回测区间：2022-09-30 至 2024-04-30	年化收益	年化波动	夏普比率	最大回撤	卡玛比率	年化换手
月频行业轮动	6.29%	18.15%	0.35	-30.03%	20.94%	单边 4.3 倍
行业等权基准	-0.70%	15.99%	-0.04	-26.49%	-2.64%	/

资料来源：Wind，华泰研究

对此，我们设计了市场主线强弱指标。该指标的基本思想是尝试寻找在不同期限下收益率均靠前的行业，如果找得到，表明市场主线较强：

- 1) 在每个交易日，计算 N 个行业指数 5/10/20/40/60 日收益率；
- 2) 截面上，计算各行业指数不同期限收益率的升序排名并归一化；如收益率最高的行业就是 N/N ，第二名是 $(N-1)/N$ ，以此类推；
- 3) 在每个交易日，计算各行业指数不同期限收益率归一化排名的均值，得到行业相对强弱指标；如右子图中行业 C 在 5 个期限下的排名是 $N-2/N, N-4/N, N-2/N, N-2/N, N-3/N$ ，则该行业的相对强弱指标等于 $N-2.6/N$ ；
- 4) 截面上，取相对强弱指标前 5 名的行业，计算这 5 个行业相对强弱指标的均值，得到市场主线强弱指标；
- 5) 对日频的市场主线强弱指标开展指数移动平均($\alpha=2/21$)。

左子图是市场主线最强的情况——按照上述计算步骤，对于 $N=31$ 来说，市场主线强度的理论最大值等于 0.935。不过，这种理想情况几乎不可能出现在实际投资中。于是，我们允许前五名行业在各期限收益率下，有平均不超过 2 名的误差，如右子图所示。其市场主线强度不会低于 0.871。这个数是强主线阈值。

我们用月末值来表征该月市场主线强弱，其与当月模型超额收益率相关系数为 +0.36，这说明市场主线越强，月频行业轮动模型越容易获得超额收益。2022 年 9 月之后相较 2022 年 9 月之前，市场主线强弱指标的均值下降了约 0.01，超过强主线阈值月份的比例下降了约 5%，说明市场主线有所变弱，一定程度上导致月频行业轮动模型表现有所变差。最近三个月市场主线都处于较弱水平，我们也无法预测这种状态还将持续多久，只能够通过提升行业轮动策略的调仓频率进行应对。

图表6：市场主线强弱指标计算示例

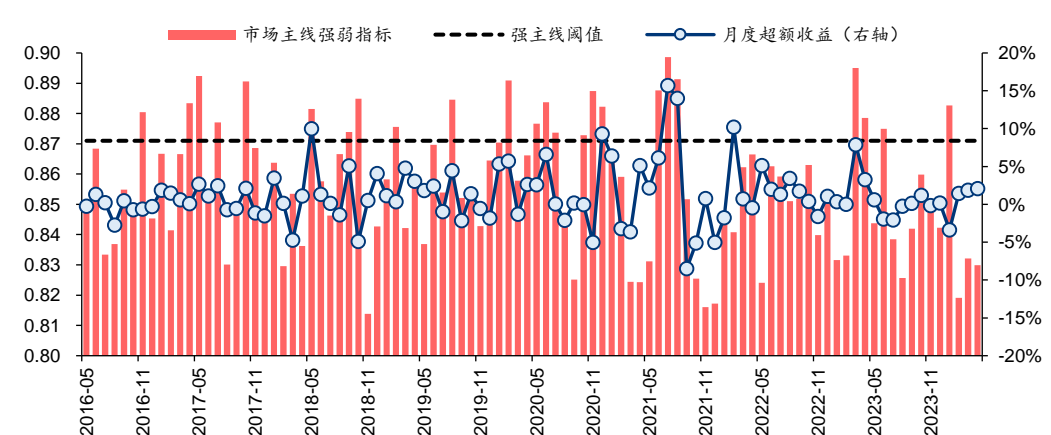
Ret_5d	Ret_10d	Ret_20d	Ret_40d	Ret_60d	Ret_5d	Ret_10d	Ret_20d	Ret_40d	Ret_60d
A	A	A	A	A	B	E	A	B	A
B	B	B	B	B	A	B	B	A	F
$C(N-2/N)$	$C(N-2/N)$	$C(N-2/N)$	$C(N-2/N)$	$C(N-2/N)$	$C(N-2/N)$	H	$C(N-2/N)$	$C(N-2/N)$	D
D	D	D	D	D	D	D	G	D	$C(N-3/N)$
E	E	E	E	E	F	$C(N-4/N)$	E	E	G

市场主线最强的情况： $[N+(N-1)+(N-2)+(N-3)+(N-4)]/5N=0.935$

强市场主线的阈值： $[N+(N-1)+(N-2)+(N-3)+(N-4)-10]/5N=0.871$

资料来源：Wind，华泰研究

图表7：月末市场主线强弱指标与当月模型超额收益率相关系数为+0.36

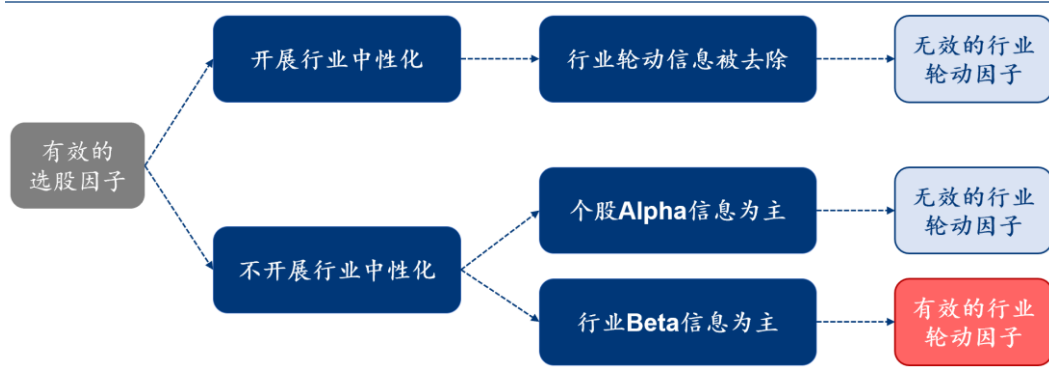


资料来源：Wind，华泰研究

背景二：用选股因子合成行业轮动因子其有效性无法被保证

构建周频行业轮动模型，容易想到的思路是用选股因子合成行业轮动因子。但是，行业轮动因子是否有效一定程度上取决于“运气”，这是因为选股因子中混合着个股 Alpha 信息和行业 Beta 信息。如果对选股因子开展行业中性化处理，那么由于行业轮动信息被去除，行业轮动因子基本上无效。如果不对选股因子开展行业中性化处理，那就得看选股因子中是个股 Alpha 信息为主还是行业 Beta 信息为主，如果是前者，那么行业轮动因子也大概率无效；如果是后者，那么行业轮动因子或许有效。

图表8：选股因子有效性是行业轮动因子有效性的必要非充分条件

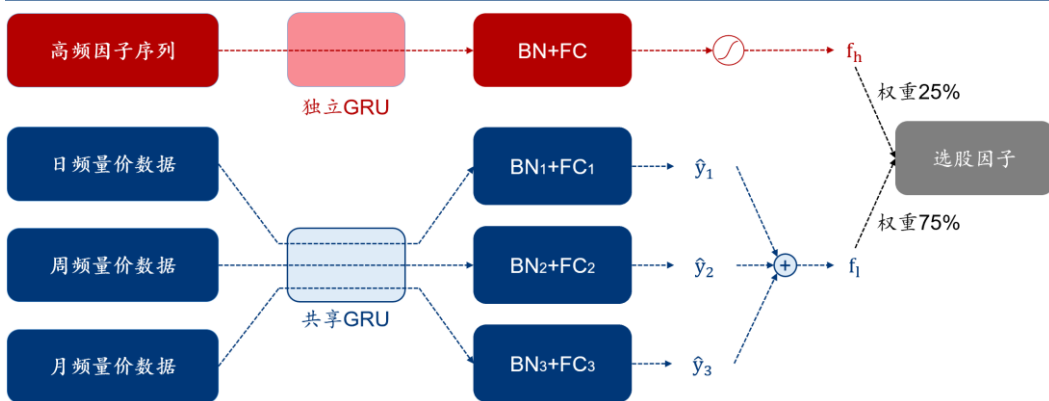


资料来源：华泰研究

基于前期报告《基于全频段量价特征的选股模型》(2023-12-08)，我们就得到了一个有效的周频行业轮动因子。具体来说，我们使用 Level2 和分钟频 K 线等高频数据构建了一系列日频因子，并通过端到端有监督学习，得到了 f_h ；同样采用端到端有监督学习，基于日频、周频、月频量价数据挖掘出了 f_l 。然后，将 f_h 和 f_l 按照 1:3 的权重合成为选股因子。最后，将选股因子按照流通市值加权的方式合成为行业轮动因子。需要注意的是， f_h 和 f_l 都没有开展行业中性化。

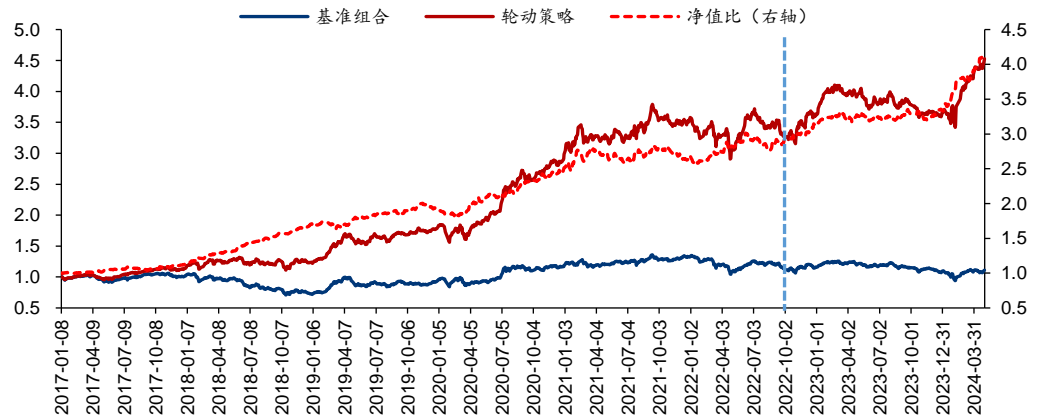
取回测区间 2017-01-08 至 2024-04-30，每周六选择综合得分前五名的中信行业指数等权配置，下周第一个交易日按开盘价、最高价、最低价、收盘价的均价完成调仓。这里需要说明，之所以没有像月频模型那样使用收盘价调仓，是因为调仓价格对于月频模型来说没有像周频模型那样敏感，采用均价更符合实战场景。后文构建的周频行业轮动模型都采用这个规则进行回测。回测结果显示，2022 年 9 月之前，模型相对于全体行业等权基准的年化超额收益约为 22%；2022 年 9 月之后，年化超额收益反而上升至约 24%，显示出周频行业轮动模型在市场主线较弱的环境中更强的适应性。

图表9：全频段量价选股因子的挖掘方法



资料来源：华泰研究

图表10：华泰金工全频段量价周频行业轮动模型净值



资料来源：Wind，华泰研究

图表11：全频段量价周频行业轮动模型业绩：2022年9月之前

回溯区间：2016-12-31 至 2022-09-30	年化收益	年化波动	夏普比率	最大回撤	卡玛比率	年化换手
周频行业轮动	23.70%	16.39%	1.45	-23.37%	1.01	单边约 20 倍
行业等权基准	2.07%	15.75%	0.13	-33.14%	0.06	/

资料来源：Wind，华泰研究

图表12：全频段量价周频行业轮动模型业绩：2022年9月之后

回溯区间：2022-09-30 至 2024-04-30	年化收益	年化波动	夏普比率	最大回撤	卡玛比率	年化换手
周频行业轮动	24.24%	14.33%	1.69	-16.63%	1.46	单边约 22 倍
行业等权基准	0.07%	13.19%	0.01	-25.31%	0.00	/

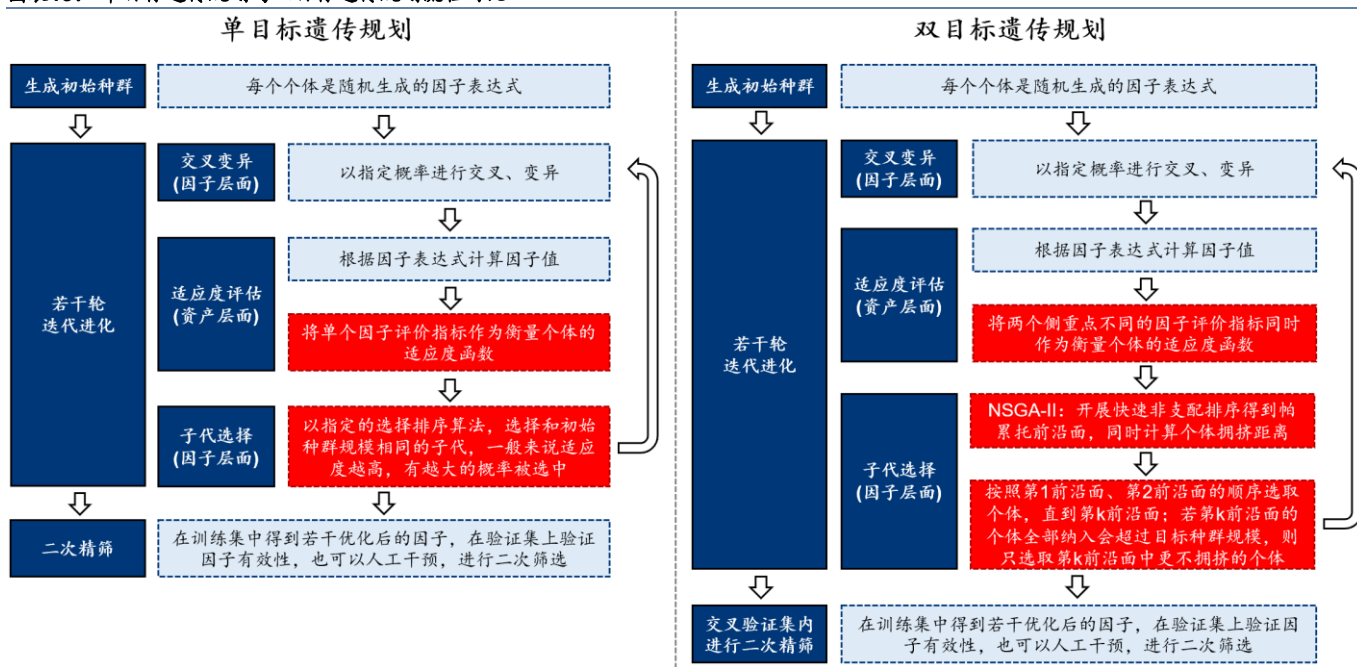
资料来源：Wind，华泰研究

双目标遗传规划原理详解

单目标遗传规划面临的痛点

使用遗传规划开展量价因子挖掘的基本研究范式如左图所示。首先，开展种群初始化，即随机生成规定数量的因子表达式，因子表达式采用“树”的数据结构进行存储。然后，进行规定数量轮迭代进化。每轮迭代进化由交叉变异、适应度评估、子代选择等三个步骤组成。在**交叉变异阶段**，按照一定顺序依次选取两个因子，按照指定的交叉概率交换一部分“基因”即子表达式，再按照指定的变异概率用一个随机生成的子表达式代替已存在的子表达式，相关示例详见前期报告《基于遗传规划的选股因子挖掘》(2019-06-10)。在**适应度评估阶段**，我们根据因子表达式，计算出训练集上每个资产的因子序列，然后在截面上进一步计算因子评价指标，对于单目标遗传规划来说，通常选择 $|IC|$ 或者 $|ICIR|$ 作为适应度函数。在**子代选择阶段**，单目标遗传规划通常采用锦标赛、轮盘赌等选择排序算法来选取子代，一般来说，因子适应度越高，有越大的概率被选中；子代作为新的父代重复上述进化过程。最后，我们将训练集上优化后的因子在验证集上验证其有效性，开展二次筛选，这一步当然也可以人工干预，例如尝试解释一下因子背后的逻辑。

图表13：单目标遗传规划与双目标遗传规划流程对比



在实践中我们发现，单目标遗传规划存在以下痛点。首先，**仅有一个适应度函数作为优化目标，无法全面概括因子的全部特性**。我们的许多前期报告都有提及这一点，这已是量化投资领域的共识。如 $|IC|$ 侧重当期因子值和未来收益率之间的单调性，至于主要是由多头端贡献的还是由空头端贡献的， $|IC|$ 就无能为力了。更重要的是，**单目标遗传规划可能面临严重的种群拥挤问题**。所谓种群拥挤，是指在进化过程中，个体的“基因”趋同，导致种群多样性降低。这就好比明代科举制度和现代选拔制度：明代科举制度采用“八股文”，始于“破题”，末于“束股”，通过考试筛选出来的人才几乎是“清一色”的满腹经纶却只会纸上谈兵；而现代选拔制度考察多门学科、有多种升学途径，各领域人才“百花齐放”。

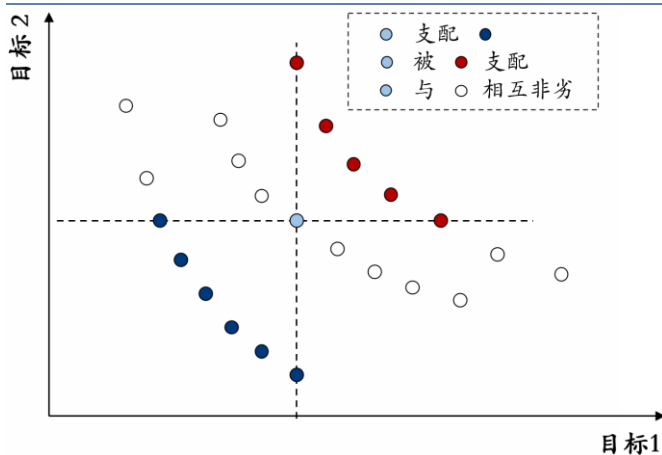
针对上述痛点，本文提出了双目标遗传规划。其基本研究范式与单目标遗传规划相似。不同的是，在**适应度评估阶段**，双目标遗传规划选取两个侧重点不同的评价指标来评价因子有效性。在单目标遗传规划中，通过比较评价指标的大小，就能评价两个因子的优劣；而在双目标遗传规划中，可能会出现双目标矛盾的情况，这就需要我们进一步分析两个因子的支配关系了。在**子代选择阶段**，双目标遗传规划采用NSGA-II算法来选择子代。

NSGA-II 算法重要概念

这一小节，我们来解释一下 NSGA-II 算法(Deb et al, 2002)涉及的重要概念：

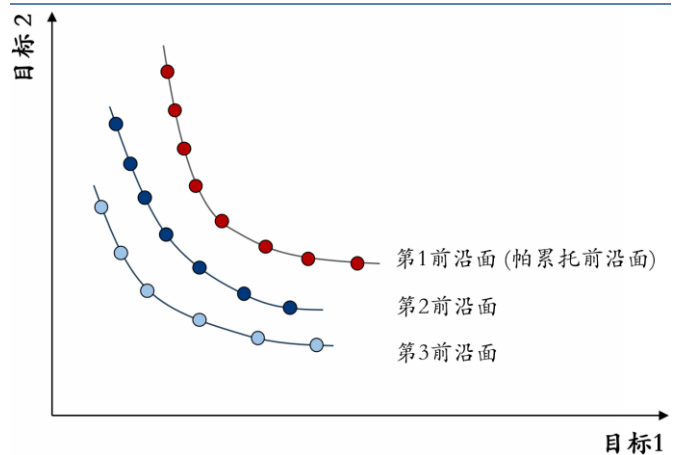
- 1) **支配与非劣**：在双目标遗传规划中，每个样本均存在两个评价维度。当个体 p 的一个评价指标优于个体 q ，而另一个评价指标不比个体 q 差，就称为个体 p 支配个体 q ；顾名思义，支配就是一个个体绝对占优于另一个个体。当个体 p 的一个评价指标优于个体 q 而另一个评价指标劣于个体 q ，就称个体 p 和个体 q 相互非劣；非劣是指两个个体势均力敌、互不支配。
- 2) **非支配前沿面**：通过快速非支配排序，可以将种群划分为如右图所示的分层状态。同一层内任何两个个体之间都是相互非劣的，所以这些分层也被称作非支配前沿面。显然，在一个种群中，存在最优的非支配前沿面，其中的个体是完全不受种群中其他任何个体支配的，这一前沿面被称为帕累托前沿面或第 1 前沿面。往后依次为第 2 前沿面、第 3 前沿面等。因篇幅原因，快速非支配排序的原理详见 NSGA-II 算法原文。

图表 14：重要概念示意图之支配与非劣



资料来源：华泰研究

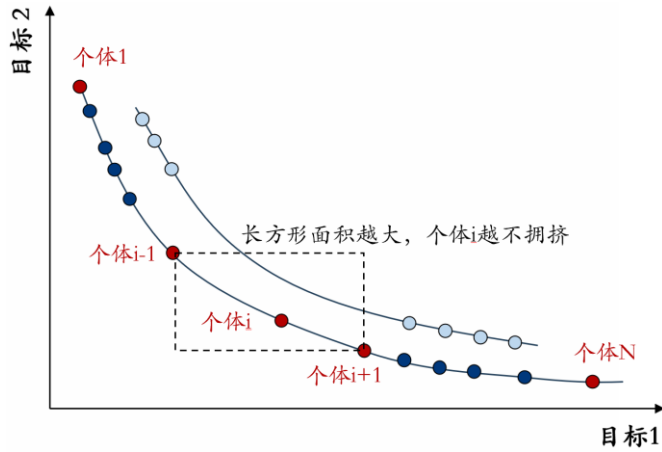
图表 15：重要概念示意图之非支配前沿面



资料来源：华泰研究

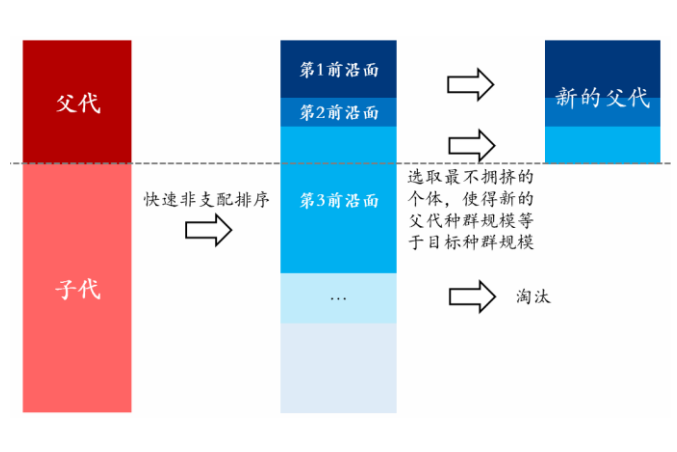
- 3) **拥挤距离**：若两个个体处于不同前沿面，可以直接根据前沿面顺序比较孰优孰劣，跟拥挤距离关系不大；但如果两个个体处于同一个前沿面，那就需要使用拥挤距离来选出更好的个体。具体来说，对于个体 i ，我们首先确定同一前沿面上其“邻居”，即个体 $i-1$ 和个体 $i+1$ ；个体 $i-1$ 和个体 $i+1$ 在坐标轴上包围形成长方形，其面积就是拥挤距离。拥挤距离越大，表明个体越不拥挤，对维持种群多样性来说越重要。处于前沿面两端的个体，是同一前沿面上差异最大的一对个体，其拥挤距离设为正无穷。
- 4) **精英选择策略**：是双目标遗传规划中独特的用于优选个体的方法。首先，将父代种群与子代种群合并，子代种群规模不应小于父代。然后，对合并后的种群开展快速非支配排序，得到非支配前沿面。接着，按照第 1 前沿面、第 2 前沿面的顺序依次选取个体，直至达到目标种群规模。如果纳入第 k 前沿面的全部个体会造成种群规模超过目标种群规模，则只纳入第 k 前沿面中最不拥挤的个体，以尽量维持种群多样性。

图表16：重要概念示意图之拥挤距离



资料来源：华泰研究

图表17：重要概念示意图之精英选择策略



资料来源：华泰研究

遗传规划算子介绍

我们在前期报告《遗传规划因子挖掘的 GPU 加速》(2024-02-19)的基础上，进一步将算子补充到 64 个，以提升因子的表达能力。新增的算子函数同样使用 **PyTorch** 实现。虽然行业轮动场景仅涉及 30 余个行业指数，截面上样本数不及选股场景的百分之一，没有必要使用 GPU 加速，但本研究是有可能拓展到选股场景的，故用 PyTorch 实现以未雨绸缪。

图表18：元素级运算算子（20 个）

算子函数	算子含义	参数范围
add_const(X, c)	$X+c$	c: 1,2,...,10
mul_const(X, c)	$X*c$	c: 1,2,...,10
add(X, Y)	$X+Y$	/
sub(X, Y)	$X-Y$	/
mul(X, Y)	$X*Y$	/
div(X, Y)	X/Y	/
log_torch(X)	$\ln(\max(X, 1.0e-12))$	/
exp_torch(X)	$\exp(X)$	/
sqrt_torch(X)	$\sqrt{\max(X, 0)}$	/
square_torch(X)	X^2	/
sin_torch(X)	$\sin(X)$	/
cos_torch(X)	$\cos(X)$	/
neg(X)	$-X$	/
inv(X)	$1/X$	/
sign_torch(X)	X 的正负性	/
abs_torch(X)	$ X $	/
sigmoid_torch(X)	对应 pytorch 中的 Sigmoid 激活函数	/
hardsigmoid_torch(X)	对应 pytorch 中的 HardSigmoid 激活函数	/
leakyrelu_torch(X, α)	对应 pytorch 中的 LeakyReLU 激活函数	α : 0,0.1,...,0.9
gelu_torch(X)	对应 pytorch 中的 GELU 激活函数	/

资料来源：华泰研究

图表19：截面运算算子（8个）

算子函数	算子含义	参数范围
xs_cutquartile_torch(X, α)	采用四分位数法拉回截面上的离群值	α -四分位距倍数: 1.5,2.0,2.5,3.0
xs_cutzscore_torch(X, α)	采用 zscore 法拉回截面上的离群值	α -离群值阈值: 1.5,2.0,2.5,3.0
rank_pct_torch(X)	截面上分位数	/
xs_regres_torch(X, Y)	截面上 Y 对 X 开展一元线性回归的残差	/
xs_sortreverse_torch(X, n, mode)	X 截面前/后 n 名对应的 X 乘以-1	n: 1,2,...,10
xs_zscorereverse_torch(X, α , mode)	X 截面 z 值大于/小于阈值对应的 X 乘以-1	α : 1.5,2.0,2.5,3.0
xs_grouping_sortreverse_torch(X, Y, n, mode)	Y 截面前/后 n 名对应的 X 乘以-1	mode: 0-前端, 1-后端, 2-两端
xs_grouping_zscorereverse_torch(X, Y, α , mode)	Y 截面 z 值大于/小于阈值对应的 X 乘以-1	

资料来源：华泰研究

图表20：时序运算算子（28个）

算子函数	算子含义	参数范围
ts_cutquartile_torch(X, α)	采用四分位数法拉回时序扩展窗口上的离群值	α : 1.5,2.0,2.5,3.0
ts_cutzscore_torch(X, α)	采用 zscore 法拉回时序扩展窗口上的离群值	α : 1.5,2.0,2.5,3.0
ts_delay_torch(X, d)	时间序列延后 d 期	d: 1,2,...,10
ts_delta_torch(X, d)	时间序列的 d 期差分	d: 1,2,...,10
ts_pctchange_torch(X, d)	时间序列的 d 期百分比变化	d: 1,2,...,10
ts_rollrank_torch(X, d)	滚动 d 日的分位数	d: 10,15,...,60
ts_rollzscore_torch(X, d)	滚动 d 日的 z 值	d: 10,15,...,60
ts_rollcorr_torch(X, Y, d)	X 和 Y 滚动 d 日的相关系数	d: 10,15,...,60
ts_rankcorr_torch(X, Y, d)	X 和 Y 滚动 d 日的秩相关系数	d: 10,15,...,60
ts_covariance_torch(X, Y, d)	X 和 Y 滚动 d 日的协方差	d: 10,15,...,60
ts_max_torch(X, d)	滚动 d 日最大值	d: 2,3,...,20
ts_min_torch(X, d)	滚动 d 日最小值	d: 2,3,...,20
ts_argmax_torch(X, d)	滚动 d 日最大值的索引除以 d	d: 2,3,...,20
ts_argmin_torch(X, d)	滚动 d 日最小值的索引除以 d	d: 2,3,...,20
ts_sum_torch(X, d)	滚动 d 日之和	d: 2,3,...,20
ts_mean_torch(X, d)	滚动 d 日均值	d: 2,3,...,20
ts_decay_linear_mean_torch(X, d)	滚动 d 日按时间索引作为权重求均值	d: 2,3,...,20
ts_rollweighted_mean_torch(X, Y, d)	滚动 d 日中，以归一化的 Y 为权重，求 X 均值	d: 2,3,...,20
ts_median_torch(X, d)	滚动 d 日中位数	d: 5,6,...,20
ts_var_torch(X, d)	滚动 d 日方差	d: 5,6,...,20
ts_stddev_torch(X, d)	滚动 d 日标准差	d: 5,6,...,20
ts_avgdev_torch(X, d)	滚动 d 日平均偏差	d: 5,6,...,20
ts_kurt_torch(X, d)	滚动 d 日峰度	d: 5,6,...,20
ts_skew_torch(X, d)	滚动 d 日偏度	d: 5,6,...,20
ts_regbeta_torch(X, Y, d)	滚动 d 日中，Y 对 X 开展一元线性回归的斜率	d: 10,15,...,60
ts_regalpha_torch(X, Y, d)	滚动 d 日中，Y 对 X 开展一元线性回归的截距	d: 10,15,...,60
ts_regres_torch(X, Y, d)	滚动 d 日中，Y 对 X 开展一元线性回归的最近一期残差	d: 10,15,...,60
ts_fftpeak_torch(X, d)	滚动 d 日中，对 X 做傅里叶变换，取频谱图峰值	d: 10,15,...,60

资料来源：华泰研究

图表21：时序切割算子（8个）

算子函数	算子含义	参数范围
ts_ascsortcut_torch(X, d, n, mode)	滚动 d 日中，将 X 最小的 n 天切割出来	d: 10,15,...,60 n: 1,2,...,10 α : 1.5,2.0,2.5,3.0
ts_descsortcut_torch(X, d, n, mode)	滚动 d 日中，将 X 最大的 n 天切割出来	
ts_asczscorecut_torch(X, d, α , mode)	滚动 d 日中，将 X 的 z 值小于 $-\alpha$ 的时间切割出来；若不存在，则切割出 X 最小的 1 天	
ts_deczscorecut_torch(X, d, α , mode)	滚动 d 日中，将 X 的 z 值大于 α 的时间切割出来；若不存在，则切割出 X 最大的 1 天	mode: 1-切割部分求 X 之和, 2-全序列 X 之和减去切割部分的 X 之和, 3-切割部分 X 乘以-1 后再求全序列 X 之和
ts_grouping_ascsortcut_torch(X, Y, d, n, mode)	滚动 d 日中，将 Y 最小的 n 天切割出来	
ts_grouping_descsortcut_torch(X, Y, d, n, mode)	滚动 d 日中，将 Y 最大的 n 天切割出来	
ts_grouping_asczscorecut_torch(X, Y, d, α , mode)	滚动 d 日中，将 Y 的 z 值小于 $-\alpha$ 的时间切割出来；若不存在，则切割出 Y 最小的 1 天	
ts_grouping_deczscorecut_torch(X, Y, d, α , mode)	滚动 d 日中，将 Y 的 z 值大于 α 的时间切割出来；若不存在，则切割出 Y 最大的 1 天	

资料来源：华泰研究

双目标遗传规划应用于行业轮动

适应度函数定义

|IC|是量化投资中最常用的因子评价指标，它等于截面上各行业当期因子值排名与未来一段时期内涨跌幅排名相关系数的绝对值。但在行业轮动场景中，只有**|IC|**高是不够的。这是因为实践行业轮动的投资者通常更关心多头组能否跑赢基准，而**|IC|**侧重单调性，高**|IC|**也有可能是空头组显著跑输基准带来的，无法和多头组表现出色划上等号。

归一化折损累计增益(NDCG)在评价因子有效性，会给予多头组表现更高的关注度。其计算过程同样使用截面上各行业当期因子值和下期收益率——计算当期因子值降序排名 i ，如因子值最大的行业 $i=1$ ，因子值最小的行业 $i=N$ ；计算下期收益率的分组得分 r ，在全体行业分 6 组的场景下，表现最好的一组中各行业得分均为 5，表现最差的一组中各行业得分均为 0；组间按升序排名，与因子值的排序方向相反；处于同一组的行业不做更精细区分，是因为多头组行业会被一揽子等权买入，组内排序的差异并不会影响策略表现。

每个行业都有了 r 和 i ，我们首先计算每个行业的 DCG_i ：

$$DCG_i = \frac{2^r - 1}{\log_2(i + 1)}$$

然后，取因子值前 k 名即 $i \leq k$ 的行业，取其 DCG_i 之和，得到 $DCG@k$ ：

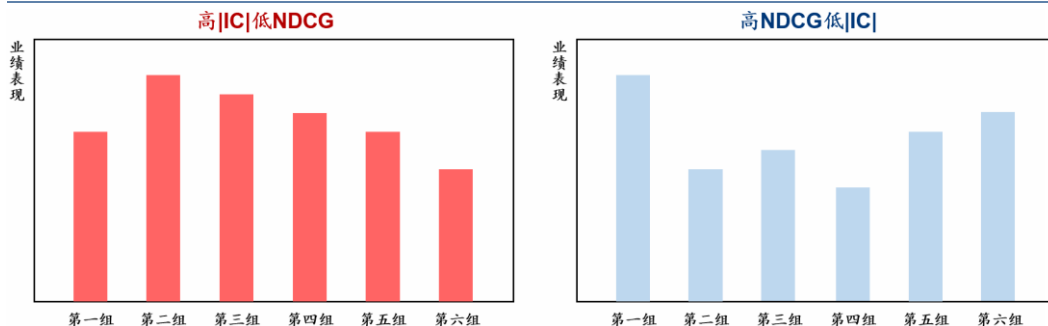
$$DCG@k = \sum_{i=1}^k DCG_i$$

对于因子值前 k 名的行业， i 是给定的即 1 至 k ，所以 DCG_i 的分母也是给定的，其大小完全取决于分子中的 r 。 r 越大即模型预测越准确，那么 DCG_i 越大， $DCG@k$ 也越大。当然 $DCG@k$ 的数量级会受到行业数量、分组数量、 k 的影响。本研究 $k=5$ 。为了统一量纲，我们可以计算出理想情况即真实收益率排名完全预测准确时的 $DCG@k$ ，记作 $IDCG@k$ 。然后，我们可以使用 $IDCG@k$ 对 $DCG@k$ 进行去量纲化：

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

|IC|和 $NDCG@k$ 的取值范围都是 0 至 1，数值越大，因子越有效。我们希望遗传规划得到的因子同时具备较高的**|IC|**和较高的 $NDCG@k$ ，因为如果 $NDCG@k$ 较低，意味着多头组表现不佳；如果**|IC|**较低，意味着因子单调性不好，会给多因子线性合成造成麻烦。

图表22：|IC|和 NDCG 的区别



资料来源：华泰研究

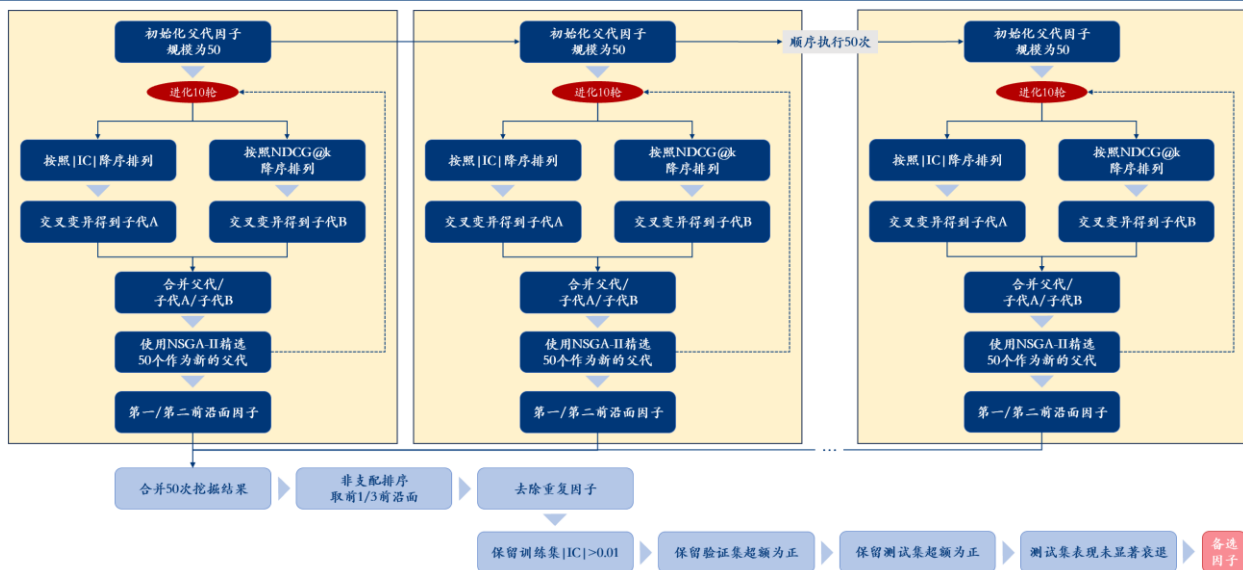
因子挖掘流程

在因子挖掘流程中能否维持种群多样性是因子挖掘能否成功的关键。前文已经说过，相较于单目标遗传规划，双目标遗传规划由于双目标之间的对抗，种群拥挤的速度自然会慢得多。但这不等于不会拥挤，毕竟随着多轮优胜劣汰，优秀“基因”的复制将呈现指数级增长，而新的优秀“基因”的出现概率却仅近似于一个较小的常数。一方面，除非交叉概率或变异概率等于1，否则父代因子总有一定概率既没有参与交叉又没有参与变异，而原封不动地被保留到了子代，这里不可能产生新的优秀“基因”。另一方面，为了能够让优秀“基因”传承下去，交叉概率一般设置得比较大而变异概率一般设置得比较小，像本研究交叉概率和变异概率分别设置为0.8、0.3；而且，变异“基因”不一定有原“基因”优秀。

对此，我们还做了如下努力来尽可能阻止种群拥挤。首先，在设置种群规模和进化次数的时候，我们延续前期报告《遗传规划因子挖掘的GPU加速》(2024-02-19)中的做法，要求前者远大于后者。换句话说，由种群规模来决定初始种群多样性，并且在有限的进化次数内，即使发生了极端的种群拥挤，种群多样性依然能够保持在一个可以接受的水平。在本研究中，种群规模设置为2500，进化次数设置为10。

其次，我们将初始种群分为50个小种群，每一个小种群单独执行遗传规划流程。我们主要担心种群中可能会出现“超强基因”，导致其他的优秀“基因”被替换掉。一旦初始种群被分割成若干份，即使出现一个“超强基因”，它也只能被限制在其中一个小种群中，其他小种群中的优秀“基因”依然有机会得以保留。这种做法的思想有点像大陆分裂，如果地球一直处于盘古大陆状态，像澳大利亚的袋鼠就不一定会出现了。

图表23：一轮因子挖掘流程图

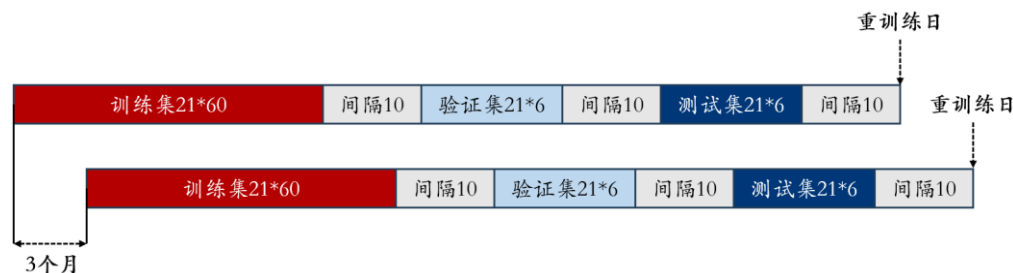


资料来源：华泰研究

一轮因子挖掘流程的其他细节相见上图。因子挖掘从2022-09-30开始，即专门针对月频行业轮动模型效果变差的阶段建模，我们每隔3个月重新挖掘一遍因子。为了降低随机因素的干扰，我们设置了6个不同的随机数种子，执行6轮因子挖掘。一个滚动窗口的长度略长于6年，其中约5年的数据为训练集，用于因子挖掘，约半年的数据用于验证集，约半年的数据用于测试集。滚动窗口中之所以既设置验证集又设置测试集，是为了模拟因子样本外跟踪半年没有失效才被纳入因子库的场景。

在验证集和测试集中，我们不再采用 $|IC|$ 、 $NDCG@k$ 等评价指标，而是直接对多头组开展周频调仓回测，使用多头组相对于行业等权基准的超额收益作为评价指标。我们不仅要求验证集和测试集上单因子超额收益都为正，还要求测试集表现相对于验证集表现未显著衰退。具体来说，我们使用验证集和测试集的日超额收益序列开展备择假设为“后者均值小于前者均值”的单边 t 检验，p 值越接近于 0 表明因子失效越严重，显著性水平取 0.05。因为计算 $|IC|$ 和 $NDCG@k$ 时均采用了未来 10 个交易日收益率，所以训练集、验证集、测试集、重训练日之间都要间隔 10 个交易日，防止信息泄露。

图表24：一个滚动窗口中训练集、验证集、测试集划分示意图



资料来源：华泰研究

我们对因子挖掘流程使用的输入变量和涉及的参数细节进行了总结：

图表25：因子挖掘流程使用的输入变量（26个）

输入变量	变量含义	输入变量	变量含义
close_ori	收盘价原始值	close_st	收盘价/收盘价滚动 60 日均值
open_ori	开盘价原始值	open_st	开盘价/收盘价滚动 60 日均值
high_ori	最高价原始值	high_st	最高价/收盘价滚动 60 日均值
low_ori	最低价原始值	low_st	最低价/收盘价滚动 60 日均值
volume_ori	成交额原始值	volume_st	成交额/成交额滚动 60 日均值
turn_ori	换手率原始值	turn_qt	换手率 5 年分位数
wopen_st	滚动 5 日开盘价/收盘价滚动 60 日均值	mopen_st	滚动 20 日开盘价/收盘价滚动 60 日均值
whigh_st	滚动 5 日最高价/收盘价滚动 60 日均值	mhigh_st	滚动 20 日最高价/收盘价滚动 60 日均值
wlow_st	滚动 5 日最低价/收盘价滚动 60 日均值	mlow_st	滚动 20 日最低价/收盘价滚动 60 日均值
wvolume_st	滚动 5 日成交额/成交额滚动 60 日均值	mvolume_st	滚动 20 日成交额/成交额滚动 60 日均值
wturn_qt	换手率 5 日均值的滚动 5 年分位数	mturn_qt	换手率 20 日均值的滚动 5 年分位数
daily_ret	日收益率	daily_exc	相对于行业等权基准的日超额收益率
pblf_ori	市净率原始值	pblf_qt	市净率滚动 4 年分位数

注：滚动 5 日收盘价和滚动 10 日收盘价就是每日收盘价，故没有 wclose_st 和 mclose_st 这两个变量

资料来源：Wind，华泰研究

图表26：因子挖掘流程涉及的参数细节

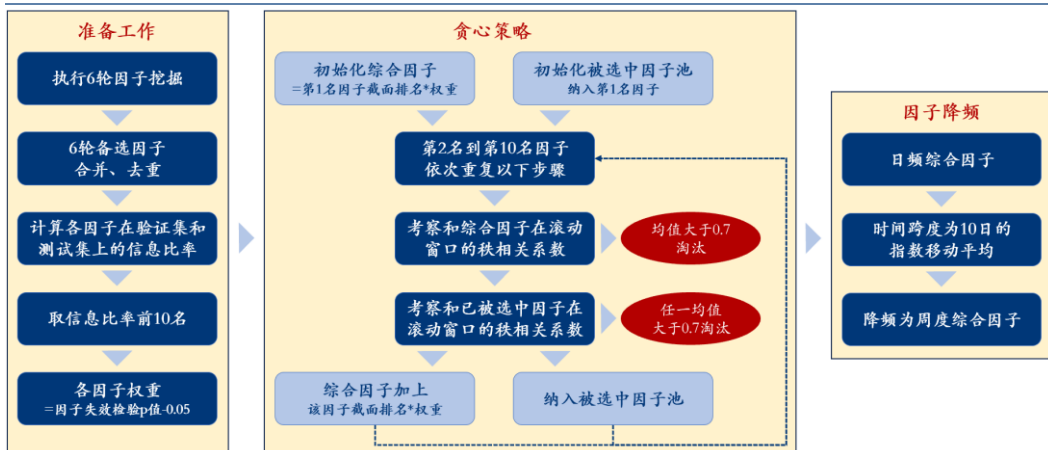
参数细节名称	参数细节说明
目标种群规模	2500，分成 50 个小种群，每个小种群独立执行遗传规划
进化次数	10
因子表达式最低阶数	1
因子表达式最高阶数	3
交叉概率	0.8
变异概率	0.3
随机试验次数	6
训练集规模	21*60
训练集标签	未来 10 个交易日收益率
训练集评价指标	$ IC $ 和 $NDCG@k$ ， $k=5$
验证集规模	21*6
验证集评价指标	周频调仓，多头组相对行业等权基准的超额收益
测试集规模	21*6
测试集评价指标	周频调仓，多头组相对行业等权基准的超额收益
因子失效标准	单边 t 检验 p 值 ≤ 0.05 ，t 检验的备择假设为测试集超额收益小于验证集超额收益

资料来源：华泰研究

实证：同时将 $|IC|$ 和 $NDCG@k$ 作为适应度函数

在每个重训练日，执行了 6 轮因子挖掘流程之后，我们得到了若干备选因子。我们按照各因子在验证集和测试集上的信息比率降序排名，取前 10 名因子，其表达式及在滚动窗口中的表现详见附录。我们采用贪心策略将这 10 个因子合成为综合因子。贪心策略的过程主要考察目标因子和已被选中因子之间的相关性；因子合成采用截面排名加权求和的方式，权重等于上一小节提及的因子失效检验 p 值减显著性水平 0.05。通过贪心策略得到日频综合因子后，我们进一步对其开展时间跨度为 10 个交易日的指数移动平均，再降频为周度综合因子。之所以开展指数移动平均，一方面是有利于降低因子换手率，另一方面是因为训练集标签是未来 10 个交易日收益率，所以近 10 个交易日的因子可能都含有关于行业未来涨跌幅的信息。具体步骤和细节详见下图：

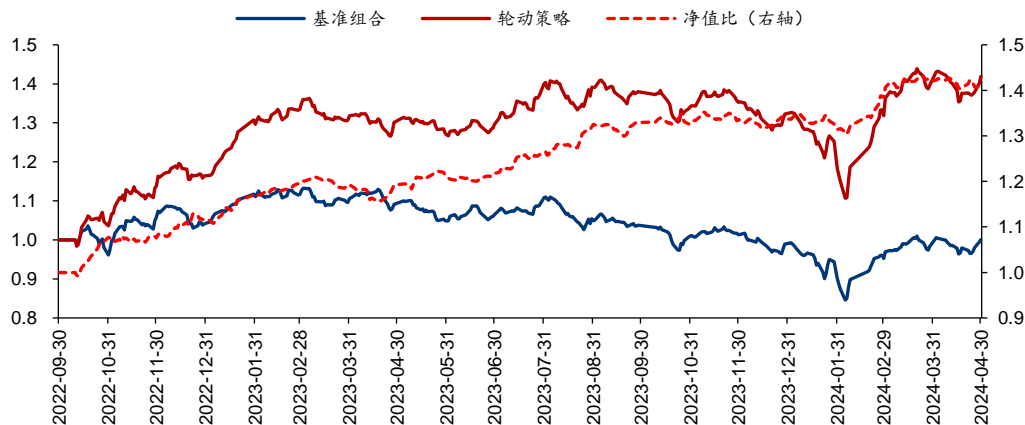
图表27：多因子合成流程



资料来源：华泰研究

取回测区间 2022-09-30 至 2024-04-30，每周六选择综合因子前五名的中信行业指数等权配置，下周第一个交易日按开盘价、最高价、最低价、收盘价的均价完成调仓。回测结果显示，双目标遗传规划周频行业轮动模型的扣费前年化超额收益为 25.74%，和全频段量价周频行业轮动模型的表现不相上下，前者的优势在低换手，后者的优势在低回撤。

图表28：双目标遗传规划周频行业轮动模型净值



资料来源：Wind，华泰研究

图表29：双目标遗传规划周频行业轮动模型业绩

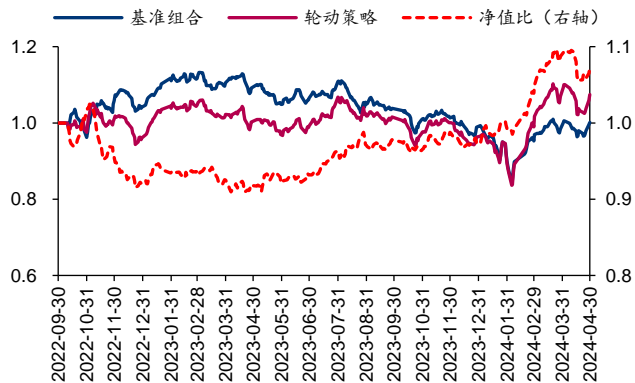
回测区间：2022-09-30 至 2024-04-30	年化收益	年化波动	夏普比率	最大回撤	卡玛比率	年化换手
周频行业轮动	25.81%	15.17%	1.70	-21.47%	1.20	单边约 13 倍
行业等权基准	0.07%	13.19%	0.01	-25.31%	0.00	/

资料来源：Wind，华泰研究

消融实验：仅以 $|IC|$ 或 NDCG@k 作为适应度函数

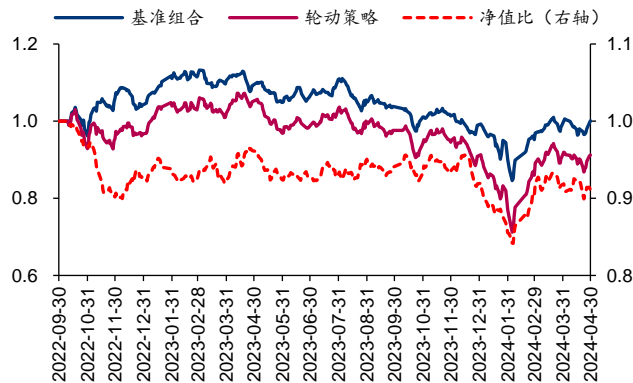
在消融实验中，每一轮因子挖掘退化成了仅以 $|IC|$ 或 NDCG@k 为适应度函数的单目标遗传规划，多因子合成流程、遗传规划和因子合成流程中的其他参数细节均保持不变。在同样的回测区间里，无论是以 $|IC|$ 还是以 NDCG@k 为适应度函数，单目标遗传规划模型的表现与双目标遗传规划模型的表现均呈现天壤之别，其中以 NDCG@k 为适应度函数的模型截至报告发布日仍跑输行业等权基准。

图表30：仅以 $|IC|$ 为目标的遗传规划模型净值



资料来源：Wind，华泰研究

图表31：仅以 NDCG@k 为目标的遗传规划模型净值



资料来源：Wind，华泰研究

更多思考

从双目标遗传规划和单目标遗传规划的结果对比来看，双目标对抗确实可能将因子挖掘引入了一条与单目标遗传规划截然不同的路径，尽管双目标遗传规划的优势还需要更多场景如选股因子挖掘的验证。我们不禁好奇，对于另一类常见的因子挖掘范式——端到端有监督学习，引入双目标对抗的机制或许也能够提升模型表现，不过实现难度似乎不低。

从挖掘出来的因子来看，时序切割因子、成交额相关变量被识别为“优秀”基因。回测区间内共有 7 个重训练日，信息比率前 10 名因子共有 70 个，其中有 52 个因子表达式用到了时序切割因子。时序切割的思想是“分域建模”，这也是量化投资领域重要的课题之一。至于成交额相关变量的频繁出现，可能是因为回测区间内 A 股市场缺乏增量资金，量对于把握行业投资机会可能更为重要，所谓“得量者得势”。

双目标遗传规划仍然存在工程上的改进空间。比如，目前随机数种子只能控制种群的初始化，还无法控制交叉和变异，在一定程度上会影响因子的可复现性。我们暂时还无法彻底解决这个问题，但可以通过重复执行因子挖掘流程来尽量降低随机因素的干扰。再如，加入更多的输入数据，如基本面数据，或者实现更多不同功能的算子，对提升种群多样性可能也是有益的。以及，我们也可以为算子增加一些限制性条件，让其尽量不出现匪夷所思的表达式，如自己除以自己，来缩小无效搜索空间。

本研究在一台配置普通的计算机上完成——内存为 24G，CPU 型号为 i7-7700，GPU 型号为 GTX1660-6G。一个随机数种子需要启动 7 个 Console，执行 7 个重训练日的因子挖掘任务，不到 8 个小时能够运行完毕。如果算力允许的话，适当提高重训练频率可能有助于应对因子过快失效的风险。

参考文献

Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2):182-197.

风险提示

遗传规划在滚动窗口中挖掘历史规律，规律可能在下次重新挖掘之前失效；月频和周频行业轮动模型都有其适用的市场条件，无法保证在任何市场条件下均可取得超额收益；涉及的具体行业不代表任何投资意见，请投资者谨慎理性地看待。

附录：重训练日挖掘出来的信息比率前十名因子一览

图表32：重训练日 2022-09-30 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
div(ts_decsortcut_torch(exp_torch(wvolume_st), 60, 2, 2), sqrt_torch(pblf_qt))	0.087	0.389	4.16	1.71	0.08
ts_grouping_asczscorecut_torch(add(mlow_st, pblf_qt), ts_cutzscore_torch(open_ori, 1.5), 10, 3.0, 1)	0.089	0.386	4.76	0.89	0.06
div(ts_deczscorecut_torch(ts_argmax_torch(low_ori, 8), 10, 3.0, 0), volume_st)	0.094	0.320	3.33	1.88	0.20
div(pblf_qt, xs_zscorereverse_torch(ts_decsortcut_torch(volume_st, 60, 5, 2), 2.0, 1))	0.072	0.393	2.49	2.04	0.35
ts_regalpha_torch(close_st, ts_cutquartile_torch(close_st, 2.0), 55)	0.103	0.363	2.39	2.01	0.33
div(ts_cutzscore_torch(pblf_qt, 2.0), ts_grouping_asczscorecut_torch(mvolume_st, neg(close_ori), 50, 2.0, 2))	0.074	0.389	2.04	1.66	0.41
ts_covariance_torch(log_torch(close_st), ts_cutzscore_torch(low_ori, 1.5), 50)	0.092	0.389	1.16	2.50	0.73
div(ts_decsortcut_torch(volume_st, 60, 2, 2), pblf_qt)	0.076	0.389	1.67	1.83	0.49
div(add(ts_rollrank_torch(high_ori, 55), wlow_st), ts_sum_torch(ts_sum_torch(wvolume_st, 2), 12))	0.105	0.339	2.57	0.93	0.19
div(add(ts_rollrank_torch(high_ori, 50), wlow_st), ts_sum_torch(ts_sum_torch(wvolume_st, 5), 12))	0.106	0.338	3.26	0.22	0.05

注：高亮的因子用到了时序切割算子，下同

资料来源：Wind，华泰研究

图表33：重训练日 2022-12-31 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
div(ts_cutzscore_torch(square_torch(close_st), 2.0), square_torch(close_st))	0.087	0.367	4.84	1.66	0.11
exp_torch(div(ts_grouping_decsortcut_torch(mvolume_st, mvolume_st, 20, 7, 0), ts_cutzscore_torch(whigh_st, 2.5)))	0.087	0.327	3.48	2.05	0.22
ts_grouping_deczscorecut_torch(ts_covariance_torch(mopen_st, abs_torch(wlow_st), 20), abs_torch(low_ori), 10, 2.5, 2)	0.096	0.353	1.79	3.67	0.76
xs_regres_torch(ts_rollzscore_torch(open_ori, 60), ts_decsortcut_torch(wvolume_st, 30, 5, 1))	0.103	0.336	3.08	1.84	0.24
ts_grouping_asczscorecut_torch(div(wvolume_st, wvolume_st), open_ori, 45, 2.5, 0)	0.085	0.341	2.77	1.94	0.32
ts_grouping_decsortcut_torch(exp_torch(volume_st), div(ts_rollrank_torch(pblf_qt, 50), ts_delay_torch(pblf_qt, 9)), 50, 9, 2)	0.089	0.332	2.17	2.45	0.46
div(mvolume_st, sigmoid_torch(ts_rollrank_torch(high_ori, 55)))	0.109	0.320	3.25	1.34	0.15
ts_regalpha_torch(ts_rollzscore_torch(close_ori, 45), add_const(sqrt_torch(wvolume_st), 5), 40)	0.087	0.336	3.23	1.33	0.20
xs_regres_torch(ts_rollzscore_torch(open_ori, 60), ts_decsortcut_torch(wvolume_st, 30, 2, 1))	0.107	0.335	3.37	1.10	0.13
ts_grouping_ascsortcut_torch(wvolume_st, leakyrelu_torch(leakyrelu_torch(low_ori, 0.7), 0.7), 55, 10, 0)	0.086	0.345	2.60	1.87	0.28

资料来源：Wind，华泰研究

图表34：重训练日 2023-03-31 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
div(ts_cutzscore_torch(abs_torch(low_st), 2.0), low_st)	0.081	0.361	4.41	3.35	0.50
ts_grouping_ascsortcut_torch(wvolume_st, ts_rollweighted_mean_torch(ts_decsortcut_torch(high_ori, 25, 6, 0), mopen_st, 4), 45, 5, 0)	0.103	0.354	2.62	4.63	0.76
ts_covariance_torch(inv(wlow_st), mlow_st, 45)	0.068	0.359	4.18	1.73	0.07
ts_covariance_torch(mopen_st, wlow_st, 25)	0.086	0.346	2.30	3.36	0.51
sigmoid_torch(ts_grouping_ascsortcut_torch(sqrt_torch(volume_st), sigmoid_torch(mopen_st), 45, 8, 0))	0.094	0.333	2.45	3.10	0.58
sigmoid_torch(ts_grouping_ascsortcut_torch(sqrt_torch(volume_st), sigmoid_torch(mopen_st), 40, 7, 0))	0.088	0.341	1.87	3.63	0.77
sigmoid_torch(ts_grouping_ascsortcut_torch(sqrt_torch(volume_st), sigmoid_torch(mopen_st), 45, 7, 0))	0.094	0.339	1.94	3.52	0.79
ts_covariance_torch(ts_min_torch(pblf_ori, 16), high_st, 45)	0.054	0.382	3.19	1.57	0.14
ts_regalpha_torch(ts_rollweighted_mean_torch(sqrt_torch(high_ori), high_ori, 18), ts_cutzscore_torch(high_ori, 1.5), 60)	0.099	0.358	2.41	2.30	0.52
xs_grouping_zscorereverse_torch(ts_max_torch(mvolume_st, 11), high_ori, 2.0, 0)	0.072	0.357	1.69	2.90	0.67

资料来源：Wind，华泰研究

图表35：重训练日 2023-06-30 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
xs_zscorereverse_torch(ts_grouping_ascsortcut_torch(mvolume_st, ts_cutzscore_torch(mhigh_st, 1.5), 50, 10, 0), 2.0, 1)	0.076	0.340	2.49	3.76	0.75
ts_cutquartile_torch(ts_grouping_asczscorecut_torch(wvolume_st, pblf_qt, 50, 3.0, 0), 1.5)	0.070	0.355	3.82	1.96	0.11
ts_grouping_ascsortcut_torch(add_const(mvolume_st, 8), mopen_st, 40, 8, 0)	0.094	0.325	3.45	1.84	0.17
ts_grouping_ascsortcut_torch(mvolume_st, wlow_st, 50, 8, 0)	0.082	0.344	2.69	2.01	0.27
ts_grouping_ascsortcut_torch(ts_grouping_deczscorecut_torch(wvolume_st, volume_st, 15, 3.0, 1), wlow_st, 50, 4, 0)	0.075	0.344	2.82	1.76	0.24
ts_grouping_deczscorecut_torch(wvolume_st, neg(high_ori), 50, 3.0, 0)	0.086	0.350	3.09	1.40	0.16
ts_grouping_ascsortcut_torch(mvolume_st, wlow_st, 50, 3, 0)	0.074	0.347	2.16	2.32	0.49
ts_grouping_ascsortcut_torch(volume_st, leakyrelu_torch(low_ori, 0.9), 45, 8, 0)	0.105	0.345	3.64	0.82	0.07
ts_grouping_ascsortcut_torch(volume_st, leakyrelu_torch(leakyrelu_torch(low_ori, 0.9), 0.9), 45, 8, 0)	0.105	0.345	3.64	0.82	0.07
ts_grouping_decsortcut_torch(wvolume_st, ts_rollweighted_mean_torch(open_ori, open_ori, 14), 40, 4, 2)	0.084	0.326	2.25	2.09	0.43

资料来源：Wind，华泰研究

图表36：重训练日 2023-09-30 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
xs_grouping_zscorereverse_torch(ts_decsortcut_torch(mvolume_st, 15, 5, 0), high_ori, 3.0, 0)	0.077	0.346	3.43	2.81	0.36
xs_grouping_zscorereverse_torch(add_const(ts_grouping_deczscorecut_torch(wvolume_st, wvolume_st, 40, 3.0, 2), 2), ts_min_torch(daily_ret, 20), 2.5, 1)	0.094	0.345	3.09	3.09	0.47
ts_grouping_asczscorecut_torch(volume_st, ts_rankcorr_torch(ts_grouping_ascsortcut_torch(wturn_qt, turn_ori, 55, 7, 1), ts_asczscorecut_torch(open_ori, 35, 2.0, 0), 35), 40, 2.0, 1)	0.092	0.332	2.42	3.76	0.63
ts_delay_torch(ts_grouping_decsortcut_torch(wvolume_st, mopen_st, 25, 6, 2), 1)	0.097	0.330	2.05	4.10	0.77
xs_grouping_zscorereverse_torch(ts_decsortcut_torch(mvolume_st, 15, 2, 0), high_ori, 3.0, 0)	0.078	0.344	2.93	3.16	0.52
ts_grouping_ascsortcut_torch(add(mvolume_st, open_st), ts_decsortcut_torch(open_ori, 20, 3, 2), 35, 9, 0)	0.098	0.329	3.63	2.37	0.20
ts_grouping_ascsortcut_torch(abs_torch(mvolume_st), add(pblf_qt, whigh_st), 50, 2, 0)	0.092	0.350	0.57	5.33	0.97
ts_regalpha_torch(ts_rollrank_torch(mopen_st, 15), wvolume_st, 45)	0.097	0.335	3.07	2.68	0.32
ts_min_torch(add_const(xs_regres_torch(mhigh_st, mvolume_st), 4), 20)	0.075	0.355	2.06	3.63	0.84
ts_regalpha_torch(ts_rollrank_torch(mopen_st, 45), wvolume_st, 30)	0.089	0.341	0.74	4.90	0.94

资料来源：Wind，华泰研究

图表37：重训练日 2023-12-31 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
ts_grouping_ascsortcut_torch(neg(sqrt_torch(mvolume_st)),	0.102	0.334	1.05	5.38	0.97
ts_grouping_ascsortcut_torch(xs_zscorereverse_torch(close_ori, 2.0, 1), close_ori, 10, 8, 1), 50, 7, 0)					
ts_grouping_decsortcut_torch(wvolume_st, mopen_st, 30, 8, 2)	0.103	0.333	1.80	4.60	0.86
ts_grouping_decsortcut_torch(wvolume_st, mopen_st, 25, 10, 2)	0.103	0.331	1.90	3.93	0.80
ts_grouping_ascsortcut_torch(xs_cutquartile_torch(mvolume_st, 1.5), open_ori, 50, 7, 0)	0.100	0.342	0.99	4.63	0.92
ts_grouping_ascsortcut_torch(mvolume_st, open_ori, 55, 8, 0)	0.098	0.342	0.45	4.79	0.97
ts_grouping_ascsortcut_torch(mvolume_st, open_st, 55, 10, 0)	0.097	0.350	1.16	4.01	0.89
ts_grouping_ascsortcut_torch(mvolume_st, high_ori, 50, 9, 0)	0.098	0.338	0.40	4.73	0.96
ts_grouping_ascsortcut_torch(ts_decay_linear_mean_torch(wvolume_st, 4), mopen_st, 40, 10, 0)	0.097	0.340	1.92	3.13	0.69
ts_grouping_decsortcut_torch(wvolume_st, mopen_st, 25, 8, 2)	0.106	0.336	2.01	2.99	0.64
ts_grouping_ascsortcut_torch(wvolume_st, open_ori, 50, 10, 0)	0.109	0.340	0.88	4.06	0.90

资料来源：Wind，华泰研究

图表38：重训练日 2024-03-31 的前十名因子

因子表达式	训练集 IC	训练集 NDCG	验证集 信息比率	测试集 信息比率	失效检验 p 值
ts_grouping_decsortcut_torch(mul_const(xs_cutzscore_torch(mvolume_st, 1.5), 2), xs_grouping_sortreverse_torch(ts_ascsortcut_torch(high_ori, 30, 9, 2), ts_rollzscore_torch(daily_ret, 15), 7, 0), 20, 6, 2)	0.094	0.330	5.32	4.18	0.55
ts_grouping_decsortcut_torch(mul_const(xs_cutzscore_torch(mvolume_st, 2.0), 2), xs_grouping_sortreverse_torch(ts_ascsortcut_torch(high_ori, 30, 9, 2), ts_rollzscore_torch(daily_ret, 15), 7, 0), 20, 6, 2)	0.095	0.334	4.86	3.96	0.57
ts_grouping_ascsortcut_torch(wvolume_st, inv(ts_decsortcut_torch(daily_exc, 40, 6, 2)), 35, 3, 2)	0.092	0.330	4.49	2.64	0.40
ts_grouping_ascsortcut_torch(wvolume_st, mopen_st, 45, 10, 0)	0.090	0.331	3.07	3.88	0.78
xs_grouping_zscorereverse_torch(ts_max_torch(leakyrelu_torch(mvolume_st, 0.4), 11), ts_mean_torch(wlow_st, 11), 2.5, 0)	0.089	0.338	3.32	3.57	0.66
ts_grouping_ascsortcut_torch(mvolume_st, mopen_st, 40, 4, 0)	0.095	0.330	2.26	4.59	0.88
ts_grouping_deczscorecut_torch(volume_st, ts_fftpeak_torch(volume_st, 15), 40, 2.0, 2)	0.094	0.305	3.42	3.43	0.71
ts_regalpha_torch(log_torch(high_st), mvolume_st, 50)	0.095	0.334	1.42	5.02	0.97
ts_ascsortcut_torch(xs_grouping_zscorereverse_torch(wvolume_st, pblf_ori, 2.5, 0), 40, 2, 2)	0.088	0.343	2.66	3.73	0.82
ts_rollweighted_mean_torch(ts_grouping_ascsortcut_torch(mvolume_st, mopen_st, 25, 9, 0), mvolume_st, 2)	0.097	0.336	3.61	2.75	0.45

资料来源：Wind，华泰研究

免责声明

分析师声明

本人，林晓明、徐特、何康，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并无就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934年证券交易法》（修订版）第15a-6条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师林晓明、徐特、何康本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

新加坡

华泰证券（新加坡）有限公司持有新加坡金融管理局颁发的资本市场服务许可证，可从事资本市场产品交易，包括证券、集体投资计划中的单位、交易所交易的衍生品合约和场外衍生品合约，并且是《财务顾问法》规定的豁免财务顾问，就投资产品向他人提供建议，包括发布或公布研究分析或研究报告。华泰证券（新加坡）有限公司可能会根据《财务顾问条例》第32C条的规定分发其在华泰内的外国附属公司各自制作的信息/研究。本报告仅供认可投资者、专家投资者或机构投资者使用，华泰证券（新加坡）有限公司不对本报告内容承担法律责任。如果您是非预期接收者，请您立即通知并直接将本报告返回给华泰证券（新加坡）有限公司。本报告的新加坡接收者应联系您的华泰证券（新加坡）有限公司关系经理或客户主管，了解来自或所分发的信息相关的事宜。

评级说明

投资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500指数，台湾市场基准为台湾加权指数，日本市场基准为日经225指数，新加坡市场基准为海峡时报指数，韩国市场基准为韩国有价证券指数），具体如下：

行业评级

- 增持：**预计行业股票指数超越基准
中性：预计行业股票指数基本与基准持平
减持：预计行业股票指数明显弱于基准

公司评级

- 买入：**预计股价超越基准15%以上
增持：预计股价超越基准5%~15%
持有：预计股价相对基准波动在-15%~5%之间
卖出：预计股价弱于基准15%以上
暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策
无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：91320000704041011J

香港: 华泰金融控股(香港)有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员，具有在美国开展经纪交易商业业务的资格，经营业务许可编号为：CRD#:298809/SEC#:8-70231

新加坡: 华泰证券(新加坡)有限公司具有新加坡金融管理局颁发的资本市场服务许可证，并且是豁免财务顾问。公司注册号：202233398E

华泰证券股份有限公司

南京

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码：518017

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心53楼

电话：+852-3658-6000/传真：+852-2567-6123

电子邮件：research@htsc.com

http://www.htsc.com.hk

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话：+212-763-8160/传真：+917-725-9702

电子邮件：Huatai@htsc-us.com

http://www.htsc-us.com

华泰证券(新加坡)有限公司

滨海湾金融中心1号大厦，#08-02，新加坡 018981

电话：+65 68603600

传真：+65 65091183

©版权所有2024年华泰证券股份有限公司