金融工程

证券研究报告
2020 年 02 月 20 日

基于基因表达式规划的价量因子挖掘

作者

吴先兴 分析师
SAC 执业证书编号：S1110516120001
wuxianxing@tfzq.com
18616029821

杨怡玲 联系人
yangyiling@tfzq.com

基因表达式规划下的价量因子挖掘

多因子模型能够持续改进的核心是持续有效地发现有显著选股能力的因子。基因表达式规划是一种启发式算法，其借鉴生物基因进化的思想，能够通过不断变异与进化来发现更好的解。因此，本报告中我们基于基因表达式规划来挖掘有效的价量因子。

在短周期价量数据构造的因子中，大部分因子选股效果的单调性并不显著，还有很多因子的多头没有超额收益，因此在设置因子有效性筛选指标时，我们结合了因子的 ICIR、多头超额收益以及分组收益的单调性，综合考察因子的选股效果，实际挖掘出的因也具有较为单调和显著的选股效果。

基于挖掘因子构建指数增强组合

我们将挖掘得到的周频价量因子与传统基于基本面的因子结合来构建多头组合与中证 500 指数增强组合，相比于不带挖掘因子的传统基本面因子组合，组合的收益提升显著。

- 周频调仓的多头 Top50 等权组合，年化收益 43.3%，相对于中证 500 指数年化超额 41.6%，每年都能跑赢中证 500 指数 18%以上。
- 周频调仓的中证 500 指数增强组合，年化超额收益 28.4%，信息比 5.2，每年都能跑赢中证 500 指数 14%以上。
- 日频调仓的中证 500 指数增强组合，年化超额收益 32.9%，信息比 5.7，每年都能跑赢中证 500 指数 18%以上。

风险提示：市场系统性风险，有效因子变动风险。

相关报告

1 《基于投资者偏好的组合构建》2019-12-06
2 《如何刻画因子对收益的真实预测效果》2019-09-25
3 《市场微观结构探析系列之二：订单簿上的 alpha》2019-09-05
4 《机构业绩增强，巨人肩膀上的 alpha》2019-07-04
5 《和时间赛跑-利用实时财务信息增强组合收益》2019-05-24
6 《A 股公司治理类因子解析》2019-03-27
7 《利用交易型 alpha 捕获低频模型短期收益》2019-03-18
8 《短周期视角下的指数增强策略》2019-02-11
9 《股息率因子全解析》2018-12-07
10 《基于预期业绩季度化分解的超预期 30 组合》2018-09-07
11 《基于自适应风险控制的指数增强策略》2018-07-05
12 《基本面量化视角下的医药行业选股研究》2018-06-15
13 《基于基础数据的分析师一致预期指标构建》2018-04-10
14 《哪些行业应该单独选股？——基于动态因子筛选的行业内选股实证研究》2018-02-23
15 《因子正交全攻略——理论、框架与实践》2017-10-30
16 《基于动态风险控制的组合优化模型》2017-09-21
17 《MHKQ 因子择时模型在 A 股中的应用》2017-08-15

内容目录

图表目录

1. 引言

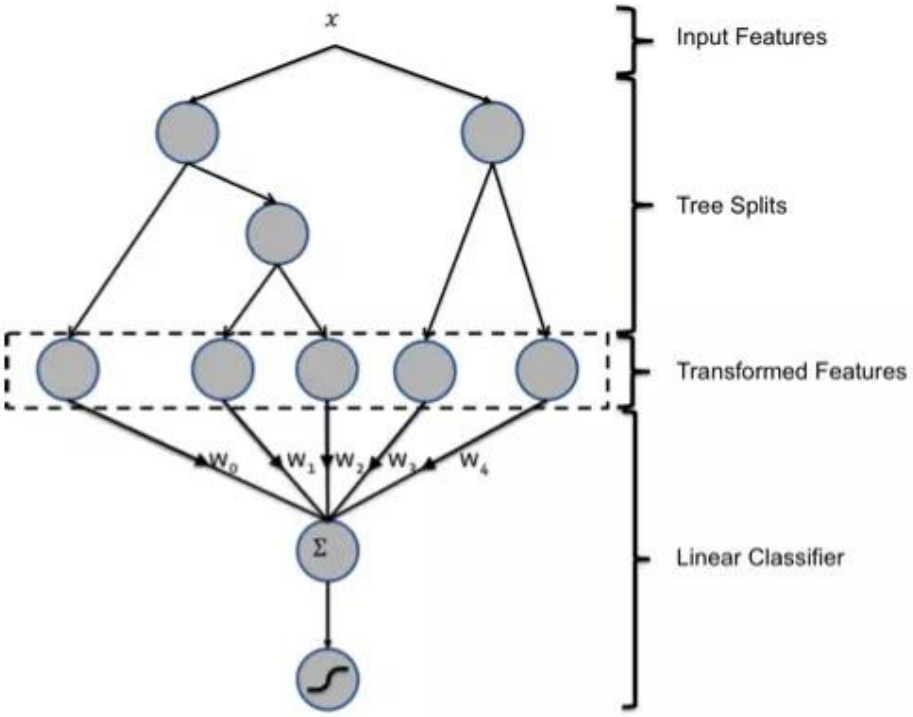
在我们之前的报告《短周期视角下的指数增强策略》(20190211) 中，我们实现了 100+ 个短周期的价量因子并构建了指数增强组合。模型能够持续改进的核心是持续挖掘和发现有效的价量因子。人工挖掘因子受限于人的知识结构与市场认知水平而效率较低，因此我们希望能够找到自动化挖掘有效因子的方法。自动化因子挖掘通常有以下几种做法。

第一种简单的做法就是对一些原始输入数据例如价格、成交量等指标进行暴力组合，典型的代表就是 sklean 包中的 PolynomialFeatures，其能够自动构建特征变量之间的多项式组合。然而这种做法可能会浪费大量的时间来检验一些无用的因子。

自然地我们会想采用一些启发式算法来提高因子挖掘的效率，典型的代表就是遗传规划类的算法。基于遗传算法来挖掘因子的主要思想是通过因子一代一代地不断变异和进化来找出越来越有效的因子。其能够挖掘因子的空间主要取决于输入的原始特征及支持的因子变换操作。[Ying 2007, Huang 2012] 等文献都是基于这种思路来挖掘因子。

第三种做法是基于树模型、神经网络等机器学习算法来构建因子。以决策树为例，决策树的每一条路径都是对特征空间的一个划分，可以看做是一个新的特征因子。Facebook 在 2014 年发表的点击率预测的论文中就是采用 GBDT 训练后得到的特征作为 LR 模型的输入 [He 2014]，如下图所示。

图 1：Facebook 的 GBDT+LR 模型



资料来源：ADKDD，天风证券研究所

在这篇报告中，我们着重介绍如何基于遗传规划类启发式算法来挖掘有效的价量因子。

2. 遗传规划

首先我们简单介绍一下遗传规划类算法的主要概念，并着重介绍基因表达式规划。

2.1. 遗传规划的原理

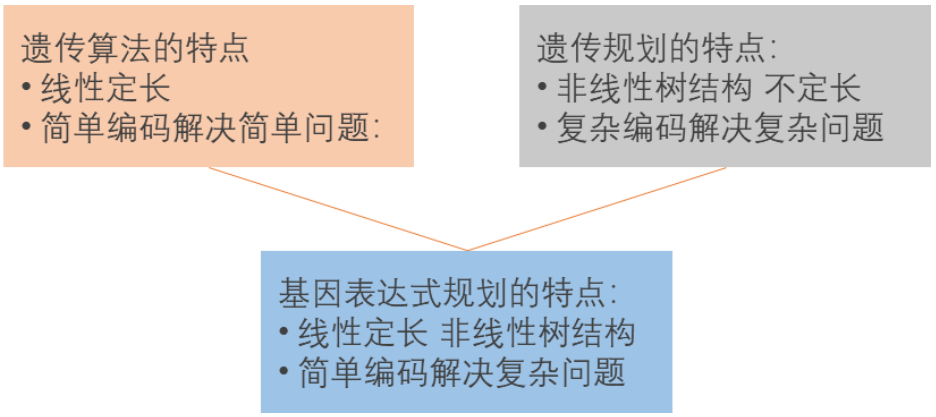
遗传规划的主要思想是通过随机生成种群，在种群中的个体之间进行交叉、变异来实现种群的进化，在目标函数下优胜劣汰，从而实现最优解的获取。其基本流程如下：

1. 随机初始化一系列个体作为一个种群；
2. 计算种群内个体的适应度，如果满足目标值或者达到终止条件则退出；
3. 精英保留，保留最好的个体；
4. 对种群中的个体进行选择、交叉、变异等操作，和保留的精英一起生成新种群，再跳转到第 2 步。

遗传规划的逐步进化的特点能够帮助我们不断发现有效因子以及不断提升因子的有效性。

遗传规划算法有很多衍生算法，其中较为常见和实用的是 Ferreira 博士提出的基因表达式规划（Genetic Expression Programming，简称 GEP）算法 [Ferreira 2001]。GEP 借用了生命科学中的基因、染色体等概念和思路，借鉴遗传进化进行数据挖掘、公式发现，及最优化。

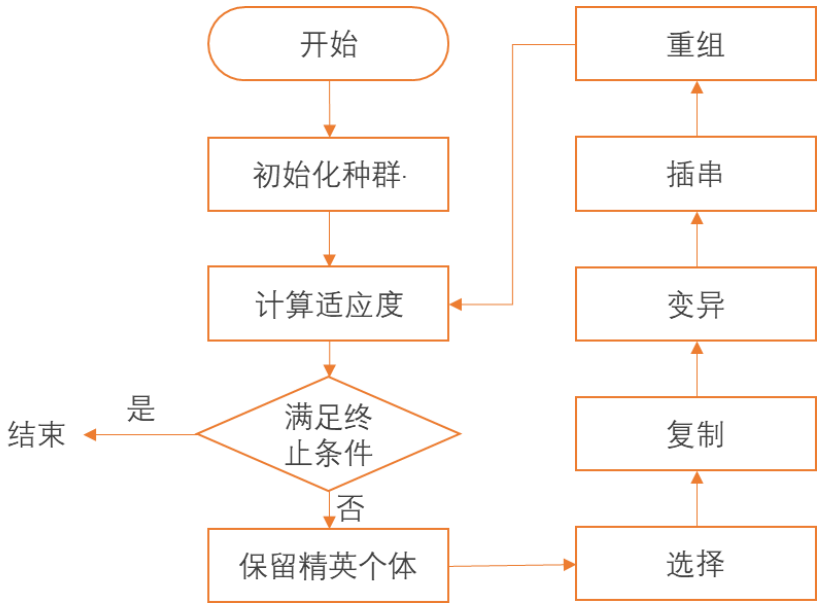
图 2：基因表达式规划与遗传算法和遗传规划的关系



资料来源：天风证券研究所

如上图说明的，GEP 兼备了遗传算法的快速易用和遗传规划的表达能力，因此在解决很多问题上，GEP 的效率远远要高于遗传算法和遗传规划。GEP 的基本流程和遗传规划是一致的，差别主要体现在对于个体的编码和解码上，一个典型的 GEP 的流程如下图。

图 3：GEP 的主要流程



资料来源：天风证券研究所

2.2. 基因表达式规划

2.2.1. 因子的表示

挖掘有效因子的前提是需要有基础的数据和函数操作，在 GEP 的语境下就是需要基础的基因和基因的变异和进化。一个基因有两种类型，一种是终结符，可以表示常数或者是变量，例如收盘价、成交量等就是终结符；另一种是函数符，可以表示运算符或函数操作，例如 $+, log$ 等都属于函数符。

下表是我们采用的终结符集合：

表 1：终结符集合

变量名	变量说明
OPEN	开盘价
HIGH	最高价
LOW	最低价
CLOSE	收盘价
PRECLOSE	前收盘价
VWAP	均价
TURN	换手率
VOLUME	成交量
AMOUNT	成交额
常数	1-20 的整数

资料来源：天风证券研究所

而下表是我们采用的函数符集合：

表 2：函数符集合

函数名	参数个数	函数说明
$add(a,b)$	2	a 加 b
$sub(a,b)$	2	a 减 b
$mul(a,b)$	2	a 乘 b
$div(a,b)$	2	a 除 b
$sqrt(a)$	1	a 的开方
$s\_sqrt(a)$	1	带符号的开方，$sign(a) * sqrt(abs(a))$
$log(a)$	1	a 的对数
$s\_log(a)$	1	带符号的对数，$sign(a) * log(abs(a))$
$abs(a)$	1	a 的绝对值
$cube(a)$	1	a 的立方
$square(a)$	1	a 的平方
$curt(a)$	1	a 的开立方
$delay(a,b)$	2	a 往前 b 天的值
$delta(a,b)$	2	$a - delay(a,b)$
$delta\_perc(a,b)$	2	$delta(a,b)/delay(a,b)$
$demean(a)$	1	a 减截面均值
$ind\_neutral(a)$	1	a 减截面行业均值
$inv(a)$	1	a 的倒数
$max(a,b)$	2	a, b 中的最大值
$min(a,b)$	2	a, b 中的最小值
$mean(a,b)$	2	a 和 b 的均值
$neg(a)$	1	a 取负数
$rank(a)$	1	截面排序
$scale(a)$	1	截面归一化，即 $a/sum(a)$
$ts\_argmax(a,b)$	2	过去 b 天 a 的最大值的下标
$ts\_argmin(a,b)$	2	过去 b 天 a 的最小值的下标
$ts\_argmaxmin(a,b)$	2	过去 b 天 a 最大值的下标-过去 b 天 a 最小值的下标
$ts\_corr(a,b,c)$	3	过去 c 天 a 和 b 的相关系数
$ts\_cov(a,b,c)$	3	过去 c 天 a 和 b 的协方差
$ts\_incv(a,b)$	2	过去 b 天 a 的均值/标准差
$ts\_max(a,b)$	2	过去 b 天 a 的最大值
$ts\_min(a,b)$	2	过去 b 天 a 的最小值
$ts\_maxmin\_norm(a,b)$	2	过去 b 天 a 的 maxmin 标准化，即 $(a - ts\_min(a,b))/(ts\_max(a,b) - ts\_min(a,b))$
$ts\_mean(a,b)$	2	过去 b 天 a 的均值
$ts\_median(a,b)$	2	过去 b 天 a 的中位数
$ts\_product(a,b)$	2	过去 b 天 a 的累乘
$ts\_rank(a,b)$	2	当天 a 取值在过去 b 天内的排序下标
$ts\_regbeta(a,b,c)$	3	过去 c 天 a 对 b 回归的系数
$ts\_std(a,b)$	2	过去 b 天 a 的标准差
$ts\_sum(a,b)$	2	过去 b 天 a 的累加
$ts\_t(a,b)$	2	$(a - ts\_mean(a,b))/ts\_std(a,b)$
$ts\_wmean(a,b)$	2	过去 b 天 a 的衰减均值，衰减系数为 1 到 b

资料来源：天风证券研究所

可以看到，我们列举的部分函数之间是能相互表示的，我们仍然把这些函数加到列表

中主要是为了节省因子查找的时间，以及方便查找出给定结构的因子。

多个基因按一定规则构成的基因串叫一个基因组。一个基因组由头部和尾部组成，头部可以包含函数符和终结符，而尾部只能包含终结符，并且尾部的长度 t 和头部的长度 h 满足如下关系：

$$t = h(n-1)+1$$

其中 n 是函数符中参数个数的最大值，所以给定头部长度和参数个数的最大值，基因组的长度是固定不变的。

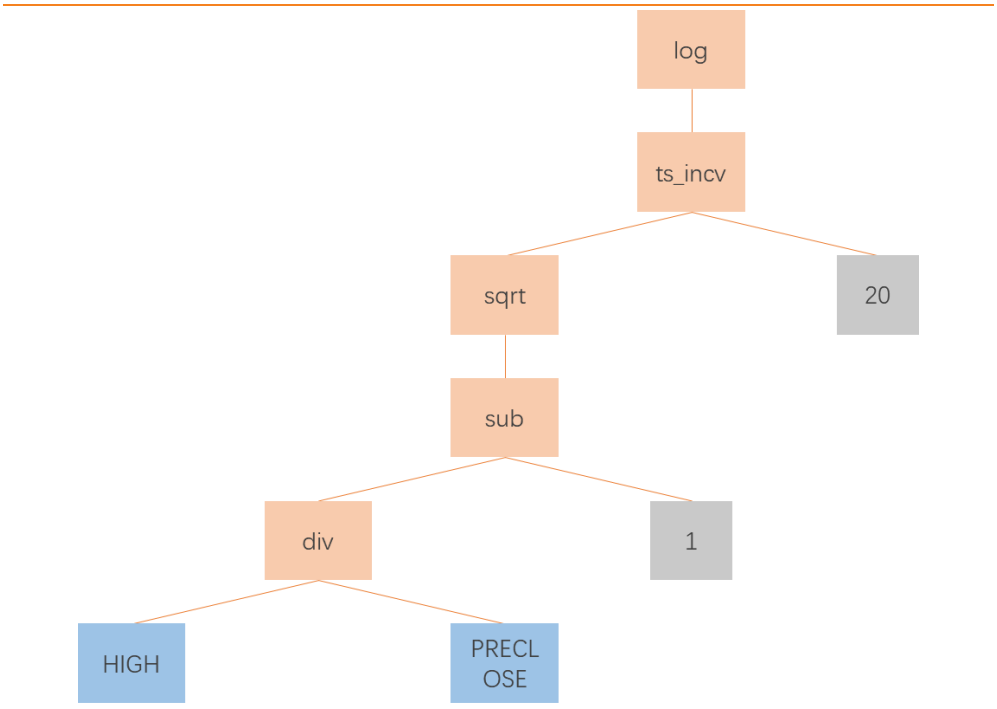
例如当 $n=2$ 时，下面是一个头部 $h=6$ 的基因组 G1，其长度为 13，前 6 个基因为头部，后 7 个基因为尾部，头部中有函数符和终结符，尾部只有终结符。

$$\underbrace{G1:\ log, ts\_incv, sqrt, 20, sub, div}_{\text{基因头部}}, \underbrace{1, HIGH, PRECLOSE, CLOSE, LOW, LOW, CLOSE}_{\text{基因尾部}}$$

一个或多个基因组构成一个染色体，由于这里我们是挖掘因子表达式，所以一个染色体中只包含一个基因组。

上面我们介绍了基因、基因组、染色体，下面我们着重介绍基因组的解析，即将一条基因组解析为一个因子表达式。以上述基因组 G1 为例，其对应的因子表达式树结构如下：

图 4：因子表达式树状结构



资料来源：天风证券研究所

从上图的因子表达式树结构可以看到，对于基因组的解析遵循了从上到下，从左到右的原则。其对应的数学表达式如下：

$$G1:\ log(ts\_incv(sqrt(sub(div(HIGH, PRECLOSE), 1)), 20))$$

可以看到虽然基因组长度为 13，而实际对应的因子表达式长度只有 9，最后的 4 个基因并没有参与解析，我们称这部分的基因为非编码区，它们是支持后续基因演化的关键。在 GEP 中，虽然基因组的长度是固定的，但是它对应的因子表达式的长度却是不定长的。如果第一个基因即为终结符，那么该基因对应的因子表达式即为长度为 1 的该终结符。

2.2.2. 个体适应度

对一个个体基因组解析出其对应的因子表达式后，我们可以根据因子表达式计算出实际的因子取值，然后评价其在目标函数下的有效性，这个目标函数我们叫做个体的适应度。

对于线性多因子模型，我们常见的因子有效性的评价标准有 IC 均值、ICIR、多空收益、多头超额收益、分组检验单调性等，而对于非线性多因子模型，我们通常可以采用最大互信息系数 MIC 等方法来判断因子和未来收益的非线性相关性。

2.2.3. 基因的演化

对于基因组的遗传操作主要包括选择、复制、变异、插串和重组,通常它们顺次执行,但变异、插串、重组之间操作的顺序对最终结果的影响并不十分重要。

选择：常见的选择操作有轮盘赌和锦标赛法。轮盘赌即以个体的适应度在总体的占比作为轮盘上的面积，即为它会被选中的概率，每次转动轮盘来选择个体；锦标赛为每次选取若干个个体，从中选择适应度最高的个体进入下一轮。

变异：对基因组上的每一位基因进行随机测试，如果满足变异概率，则重新生成该位的基因。为了避免非法个体的产生,基因头部的基因元素值可以变异成任何基因元素，而基因尾部的基因元素值只能变异成终结符，因此，GEP 的变异操作产生的新生个体都是合法的个体。这与遗传规划的变异操作会产生非法个体有明显的不同。

插串：插串是 GEP 特有的算子，它的主要做法是随机在基因组中选择一段子串，将其插入到基因头部的随机位置，并将该位置后面的基因向后顺延，超出头部长度的基因截去。

重组：重组包括单点和双点重组。单点重组是在两个父代基因组上随机选择一个位置，然后交换该位置后面的基因串得到两个子代的基因组。双点重组是随机选择两个位置，将之间的基因串交换。

2.3. 遗传规划的开源实现

遗传规划类算法在实际使用时通常不需要我们自行实现，有很多开源的框架可以采用。下表列举了不同编程语言下的开源实现。我们采用 geppy 包来实现基因表达式规划。Geppy 基于较为成熟的 deap 包来实现基因表达式规划，因此能够支持向量、矩阵、面板等各种数据形式，并且方便支持自定义的函数算子。

表 3：遗传规划开源包

名称	开发语言	支持特性	地址
Jenetics	Java/Scala/.Net	遗传规划、并行计算	https://github.com/jenetics/jenetics
Gplearn	Python	遗传规划、并行计算	https://github.com/trevorstephens/gplearn
Deap	Python	遗传算法，遗传规划，并行计算	https://github.com/DEAP/deap
Tpot	Python	遗传规划	https://github.com/EpistasisLab/tpot
Geppy	Python	基因表达式规划，并行计算	https://github.com/ShuhuaGao/geppy
Gplab	Matlab	遗传规划	http://gplab.sourceforge.net/
Open BEAGLE	C++	遗传算法，遗传规划，	https://github.com/chgagne/beagle

资料来源：天风证券研究所

3. 基因表达式规划下的价量因子挖掘

前面我们介绍了基因表达式规划的主要构成和运作流程，下面我们主要讨论如何基于基因表达式规划进行因子挖掘。基于开源的 geppy 包，我们需要做的就是：

- 定义终结符和函数符；
- 指定筛选有效因子的指标，即因子的适应度；
- 在遗传规划的流程上按需自定义操作流程。

本报告中我们以周频价量因子挖掘为例，采用以下区间的数据来挖掘有效因子：

- 股票池：全 A 股票池，剔除没有行业分类及过去 20 个交易日日均成交额低于 1000 万的股票；
- 挖掘数据区间：2017-2018 年的日度数据；
- 预测目标：未来 5 个交易日的收益率。

3.1. 因子适应度指标

在选择适应度评价指标时，首先我们要明确使用的收益预测模型。本报告中我们仍然沿用线性模型来进行收益预测，即

$$r = r_m + \Sigma X_i f_i + \Sigma X_s f_s + \Sigma X_a f_a + \varepsilon$$

其中，r 为股票收益率，r_m 为市场收益率，X_i，X_s，X_a 分别为股票对行业、风格、alpha 因子的暴露，f_i，f_s，f_a 分别为行业、风格、alpha 因子的收益，ε 为特质收益率。我们的目标是寻找到显著有效的 alpha 因子。由于我们挖掘的是价量类因子，因此我们在生成因子表达式并计算出原始因子取值后，将因子对行业以及常见的价量类风格因子进行了剥离，即对中信一级行业、市值、过去一个月收益率、过去一个月波动率、过去一个月日均换手回归取残差，以残差作为中性化后的因子取值。

在适应度指标的选择上，我们通常采用因子的 ICIR 绝对值来筛选有效因子。因子的 rank ic 定义为当期因子取值 f_a 与未来 d 天收益率 r_d 的秩相关系数，即

$$ic = corr(rank(f_a), rank(r_d))$$

而 ICIR 定义为

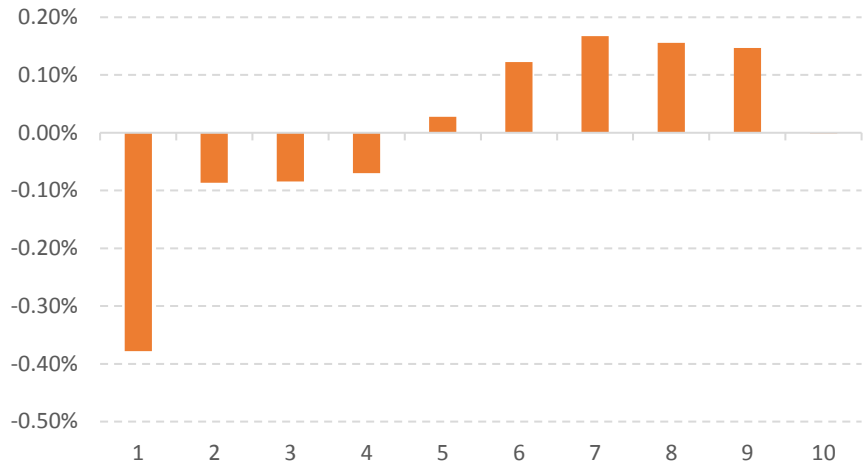
$$ic\_ir = \frac{mean(ic)}{std(ic)} \cdot \sqrt{\frac{244}{d}}$$

其能够衡量因子对于未来收益预测能力的稳定性，通常 ICIR 绝对值越高，因子的预测能力越显著。然而，在价量因子中，如果我们单纯以因子的 ICIR 绝对值来作为适应度指标时，很容易发现挖掘出来的绝大部分因子其分组收益并不单调，以挖掘出来的因子

$$G2:\ ts\_std(sub(div(HIGH, PRECLOSE), 1), 10)$$

为例，其因子 IC 均值为-0.041，ICIR 为-7.03，从 ICIR 看是非常显著的，而其十组分档周度超额收益如下图所示。可以看到，因子的收益总体是单调的，空头很强但是多头并没有超额收益。由于我们采用的是线性预测模型，这样的因子加入到收益预测模型中并不一定能带来超额收益的提升。

图 5：G2 因子的十组分档超额收益表现



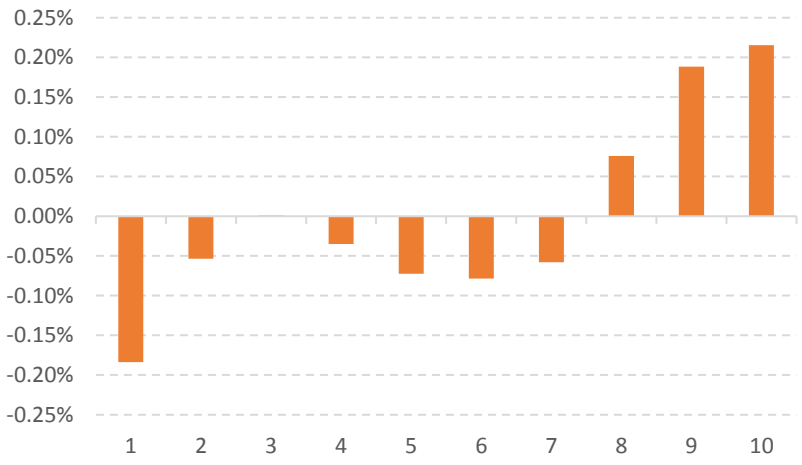
资料来源：Wind，天风证券研究所

而如果我们将因子的适应度定义为因子多头组合超额收益的信息比，即

$$long\_ir = \frac{mean(r_{long})}{std(r_{long})} \cdot \sqrt{\frac{244}{d}}$$

其中 r_{long} 为十组分档后多头组合的周度超额收益。这样我们挖掘出来的因子多头超额收益的稳定性会明显改善，但是仍然会挖掘出一些十组分档表现如下图的因子。其 ICIR 为 -5.43，多头超额信息比在 4 以上，然而其十组分档仍然不单调。

图 6：G3 因子的十组分档超额收益表现



资料来源：天风证券研究所

所以只看因子的 ICIR 或者是多头超额信息比对于筛选出线性有效的因子仍然是不够的，还需要关注分组收益的单调性。我们在分组的基础上，以分组收益和分组排序的加权秩相关系数定义为因子分组收益的单调性指标，假设分组数量为 n，空头到多头各组的权重为 1 到 n 的整数序列，即多头组的权重最高，空头组的权重最低。假设空头到多头分组超额收益分别为 $r_1, r_2, \dots, r_n$，w 为权重向量，满足

$$w_i = i / \Sigma_{j=1}^{n} j$$

分组收益单调性指标定义如下：

$$mon = max\ (0, weight\_corr(rank([r_1, r_2, \dots, r_n]), [1,2, \dots, n], w))$$

取值范围 $[0, 1]$。其中 $weight\_corr(x, y, w)$ 为加权相关系数，计算方式如下：

$$weighted\_corr(x,y,w) = \frac{\sum wxy - \sum wx \sum wy}{\sqrt{w(x-\sum wx)^2} \cdot \sqrt{w(y-\sum wy)^2}}$$

十组分档收益如果完全单调，则单调性指标取值为 1，前文中因子 G2 的单调性取值为 0.48，G3 的单调性取值为 0.7。

对于线性多因子预测模型，我们要求因子的 ICIR 显著有效、多头超额收益显著，并且分组收益的单调性要好，因此，综合以上的指标我们可以定义一个综合评价指标：

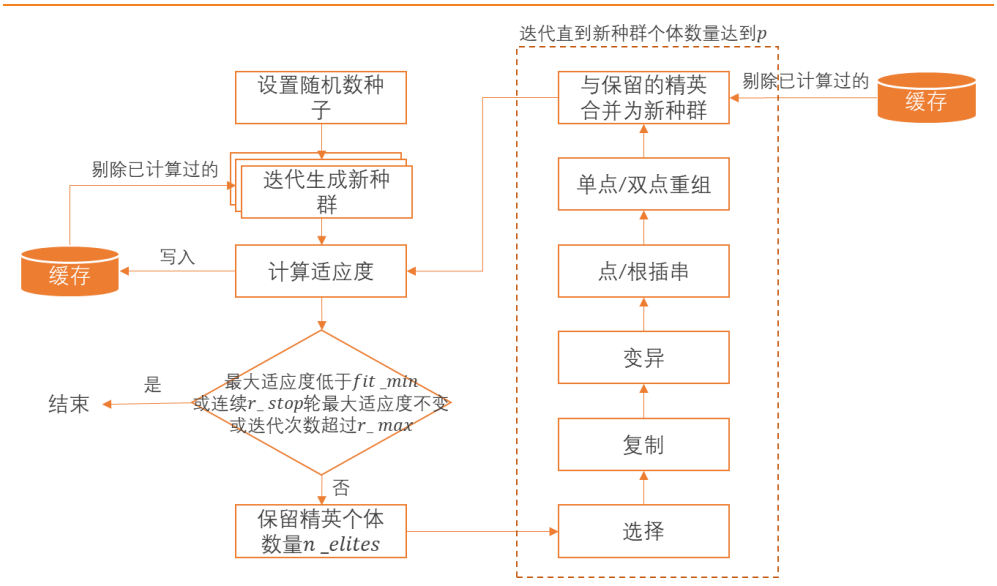
$$fitness = \sqrt{\frac{abs(ic\_ir)}{5}} \cdot long\_ir \cdot mon$$

我们如此定义适应度指标的原则是多头超额收益越显著越好，分组单调性越高越好，而因子的 ICIR 达到一定程度后，对于因子收益的边际改善较为有限。

3.2. 因子挖掘流程

定义好因子适应度指标后，我们需要自定义挖掘流程来满足我们的实际需求。由于遗传规划的主要思路是一轮一轮地迭代进化来找出适应度更高的因子，但是并不是每一轮都会产生比上一轮更好的因子，所以在挖掘时我们可以设置一些提早结束的条件来节省时间。当第一轮中因子的适应度都很差时，我们直接跳过该次挖掘。当连续若干轮都没有发现比上一轮更好的因子时，我们就提前结束该次挖掘，重新随机因子继续下一次挖掘。另外，因子在变异进化后可能会产生曾经计算过的因子，为了节省计算时间，我们在外部添加了缓存，记录每个计算过的因子的适应度取值，在进化的过程中如果发现计算过该因子则剔除。最终，我们采用的挖掘流程如下图：

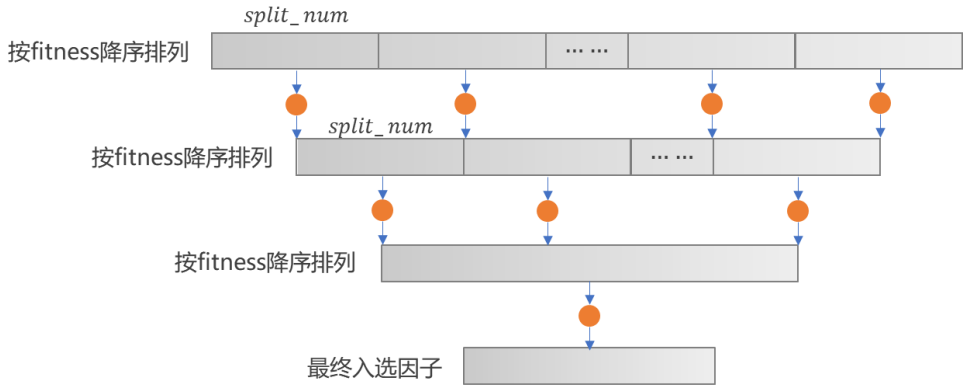
图 7：基于 GEP 挖掘因子的主要流程



资料来源：天风证券研究所

由于挖掘出的因子可能有较高的共线性，因此我们在挖掘出因子后需要剔除相关性较高的因子。由于因子数量较大，一次性计算所有因子的相关系数不太实际也没有必要，因此我们对因子进行分块剔除，主要步骤如下图所示。

图 8：剔除相关性高因子的步骤



资料来源：天风证券研究所

我们的主要操作思路是，首先将所有计算过的因子按 $fitness$ 降序排列，每 $split\_num$ 个因子为一块，在每一块中，优先入选 $fitness$ 较高的因子，将和其相关系数绝对值超过 $corr\_t$ 的因子剔除。然后将每一块中所有筛选后的因子重新汇合后再按 $fitness$ 降序排列，重新分块并迭代多次该过程，直到获得一个较小的因子集合。

在因子挖掘过程中，我们采用的主要参数如下表：

表 4：因子挖掘主要参数表

参数名称	参数说明	取值
p	种群数量	500
fit_min	种群最大适应度的提早结束阈值	0.8
r_stop	未发现更好因子时提前结束的轮数	2
r_max	最大迭代轮数	100
n_elites	保留精英个体数量	5
is_ts	点插串概率	0.1
ris_ts	根插串概率	0.1
cx_1p	单点重组概率	0.4
cx_2p	双点重组概率	0.2
$split\_num$	剔除因子时的分块大小	200
$corr\_t$	因子相关系数的阈值	0.7

资料来源：天风证券研究所

3.3. 因子示例

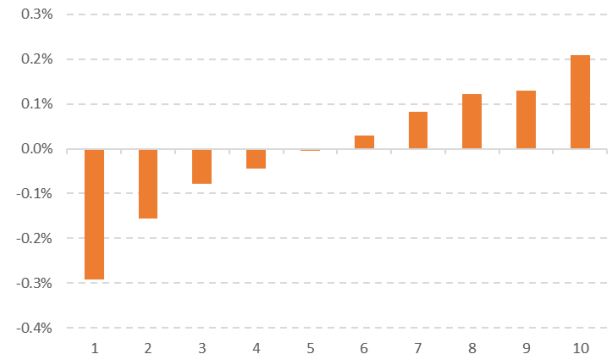
我们计算了 2w 多个因子，最终保留了 15 个相关性较低的有效因子。下面我们展示一下挖掘出的部分因子的选股效果。挖掘的因子我们保持其原始表达式，没有对其进行简化。虽然我们挖掘因子只采用了 2017-2018 的日度数据，下面我们在展示因子选股效果时，将回测区间扩展到 2007-2019 年的数据，以此可以观察因子在样本外数据的表现。

3.3.1. Alpha1

$$Alpha1:\ log(ts\_incv(sqrt(sub(div(HIGH, PRECLOSE), 1)), 20))$$

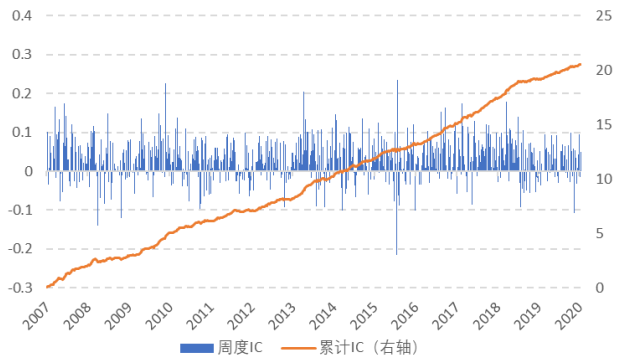
可以看到，alpha1 就是前文中的因子 G1。其因子十组分档收益如下左图所示。因子十组分档超额收益单调性非常好，多头组合周度超额收益 0.21%。

图 9：Alpha1 十组分档超额收益表现



资料来源：Wind，天风证券研究所

图 10：Alpha1 周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定。因子 IC 表现统计值如下表，因子周度 IC 均值 0.0328，年化 ICIR 在 4 以上。

表 5：Alpha1 因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
0.0328	0.0528	4.336	74.4%

资料来源：Wind，天风证券研究所

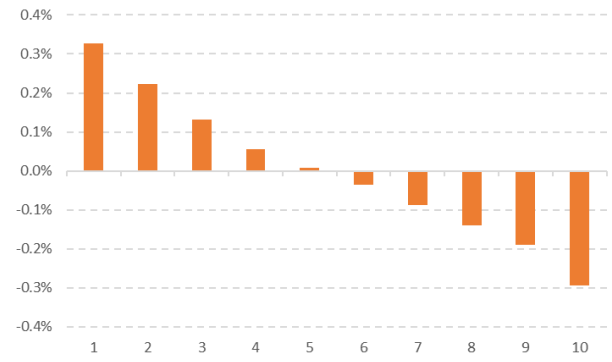
3.3.2. Alpha2

$$Alpha2: ts\_regbeta(neg(s\_log(sub(div(VWAP, PRECLOSE), 1))),$$

$$min(sub(div(HIGH, PRECLOSE), 1), AMOUNT), 20)$$

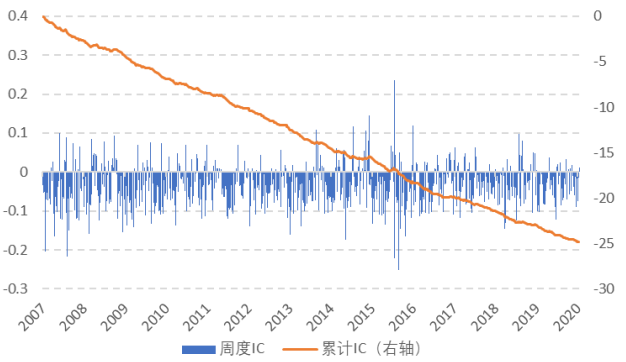
其因子十组分档收益如下左图所示。因子十组分档超额收益单调性非常好，多头组合周度超额收益 0.33%。

图 11：Alpha2 十组分档超额收益表现



资料来源：Wind，天风证券研究所

图 12：Alpha2 周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定。因子 IC 表现统计值如下表，因子周度 IC 均值-0.0394，年化 ICIR 为-4.96。

表 6：Alpha2 因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
-0.0394	0.0554	-4.96	76.7%

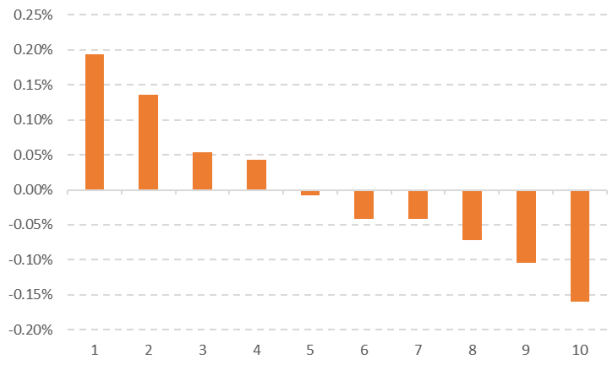
资料来源：Wind，天风证券研究所

3.3.3. Alpha3

$$Alpha3: ts\_corr(ts\_rank(TURN,5), ts\_maxmin\_norm(CLOSE,7), 15)$$

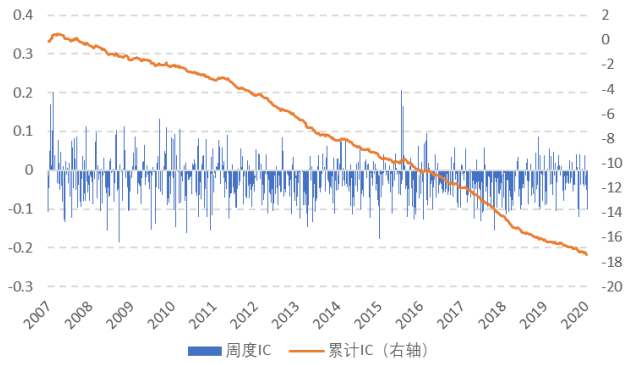
因子十组分档收益如下左图所示。因子十组分档超额收益单调性非常好，多头组合周度超额收益 0.19%。

图 13：Alpha3 十组分档超额收益表现



资料来源：Wind，天风证券研究所

图 14：Alpha3 周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定。因子 IC 表现统计值如下表，因子周度 IC 均值-0.0267，年化 ICIR 为-3.357。

表 7：Alpha3 因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
-0.0267	0.0555	-3.357	70.6%

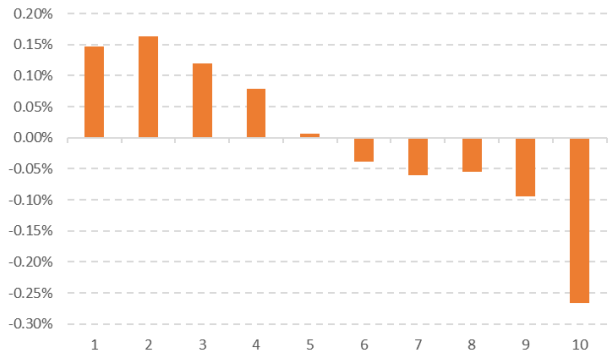
资料来源：Wind，天风证券研究所

3.3.4. Alpha4

$$Alpha4: rank(log(ts\_maxmin(TURN,15))),$$

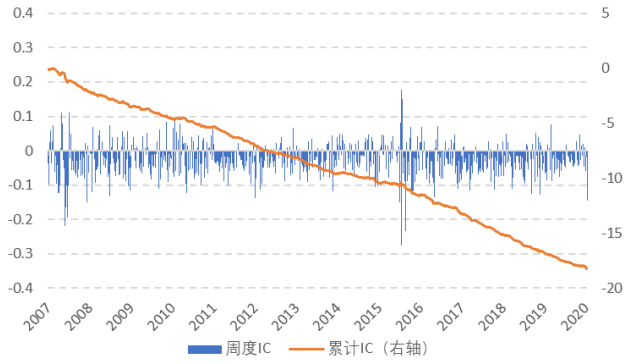
其因子十组分档收益如下左图所示。因子十组分档超额收益单调性总体较好，多头组合周度超额收益 0.15%。

图 15：Alpha4 十组分档超额收益表现



资料来源：Wind，天风证券研究所

图 16：Alpha4 周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定，并且 2016 年以后因子 IC 更显著。因子 IC 表现统计值如下表，因子周度 IC 均值-0.0279，年化 ICIR 为-4.249。

表 8：Alpha4 因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
-0.0279	0.0459	-4.249	73.8%

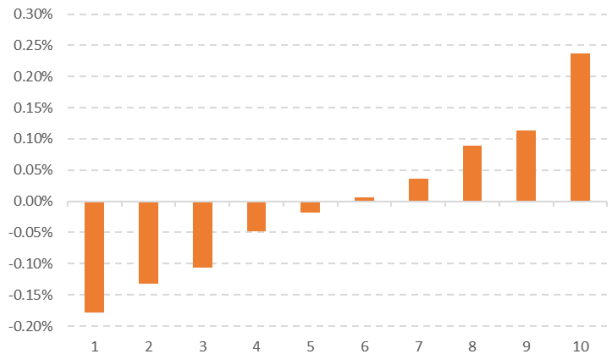
资料来源：Wind，天风证券研究所

3.3.5. Alpha5

$$Alpha5: ts\_incv(scale(mul(sub(div(HIGH, PRECLOSE), 1), ts\_argmax(AMOUNT, 5))), 15)$$

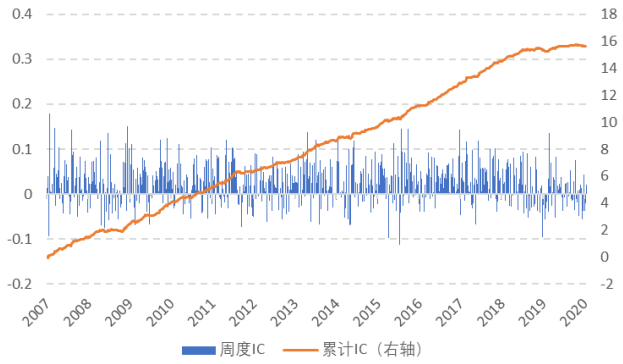
因子十组分档收益如下左图所示。因子十组分档超额收益单调性非常好，多头组合周度超额收益 0.24%。

图 17：Alpha5 十组分档超额收益表现



资料来源：Wind，天风证券研究所

图 18：Alpha5 周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定，但是 2019 年以来表现较差。因子 IC 表现统计值如下表，因子周度 IC 均值 0.0256，年化 ICIR 为 3.912。

表 9：Alpha5 因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
0.0256	0.0457	3.912	71%

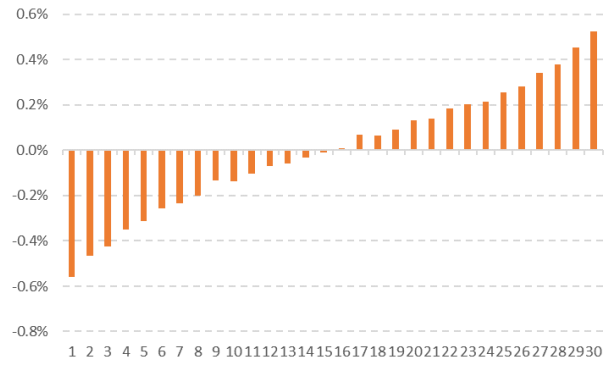
资料来源：Wind，天风证券研究所

3.4. 复合因子表现

我们将入选的 15 个因子复合起来，查看复合因子的表现。虽然入选的因子间的相关系数都低于 0.7，但是互相之间仍然存在共线性，因此在因子复合时，需要将因子间的共线性剔除。对于因子的共线性问题，我们的解决方法是对称正交。因子正交化本质上是对原始因子（通过一系列线性变换）进行旋转，旋转后得到一组两两正交的新因子，它们之间的相关性为零并且对于收益的解释度保持不变。这里我们简单对对称正交的原理和优势进行简单介绍。对称正交是一种无监督无参数的处理因子多重共线性的方法，它的主要思想是尽可能减少对原始因子的修改而得到一组正交的新因子，这样能够最大程度地保持正交后因子和原因子的相似性。并且，我们希望对每个因子平等对待，避免像施密特正交法中偏向正交顺序中靠前的因子。具体的介绍可以参见我们前期的报告《因子正交全攻略——理论、框架与实践》（20171030）。

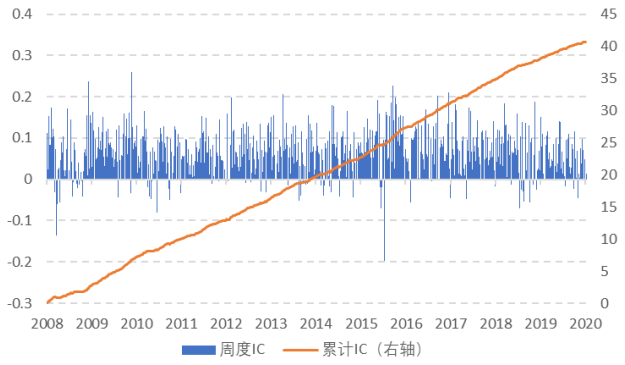
我们对入选的因子进行对称正交，以正交后因子滚动 244 天的 ICIR 加权得到复合因子。复合因子的选股表现如下图。我们对复合因子分 30 组来查看其细粒度的排序效果。可以看到，30 组分档超额收益单调性非常好，多头组合周度超额收益 0.53%。

图 19：复合因子分组超额收益表现



资料来源：Wind，天风证券研究所

图 20：复合因子周度 IC 表现



资料来源：Wind，天风证券研究所

右图展示了因子周度 IC 以及 IC 的累计值，可以看到因子长期以来的表现一直较为稳定。因子 IC 表现统计值如下表，周度 IC 均值 0.0668，年化 ICIR 为 8.463。

表 10：复合因子 IC 统计值

IC 均值	IC 标准差	年化 ICIR	IC 胜率
0.0668	0.0552	8.463	88.6%

资料来源：Wind，天风证券研究所

4. 周频因子选股实证

在本节中，我们基于挖掘出的周频因子以及一些基本面财务因子一起构建组合。因子列表如下表所示。

表 11：因子库

类别	因子简称	因子名称	因子计算方式
挖掘	MINING_WEEK	挖掘复合因子	15 个周频因子对称正交后复合
规模	LNCAP	总市值对数	总市值取对数
估值	BP	账面市值比	净资产/总市值
	EP	单季度市盈率倒数	单季度归母净利润/总市值
	SP	单季度市销率倒数	单季度营业收入/总市值
	EPTTM	市盈率倒数 TTM	归母净利润 TTM/总市值
	SPTTM	市销率倒数 TTM	营业收入 TTM/总市值
技术	REVERSE1M	一个月反转	过去 20 个交易日涨跌幅
	REVERSE3M	三个月反转	过去 60 个交易日涨跌幅
成长	NETPROFITINCYOY	单季度净利润同比增速	单季度净利润同比增长率
	OPERREVINCYOY	单季度营业收入同比增速	单季度营业收入同比增长率
	OPERPROFITINCYOY	单季度营业利润同比增速	单季度营业利润同比增长率
	SUE	标准化预期外盈利	（单季度实际净利润-预期净利润）/预期净利润标准差
	SUR	标准化预期外收入	（单季度实际营业收入-预期营业收入）/预期营业收入标准差
盈利	ROE	单季度净资产收益率	单季度归母净利润*2/（期初归母净资产+期末归母净资产）
	ROA	单季度总资产收益率	单季度归母净利润*2/（期初归母总资产+期末归母总资产）
	DELTAROE	单季度净资产收益率同比变化	单季度净资产收益率-去年同期单季度净资产收益率
	DELTAROA	单季度总资产收益率同比变化	单季度总资产收益率-去年同期单季度中资产收益率
一致预期	FEPTTM	一致预期市盈率倒数 TTM	一致预期 PETTM 倒数
	CGBP	一致预期滚动 BP	一致预期滚动 PB 倒数
	CGPEG	一致预期 PEG	一致预期 PEG
流动性	TURNOVER1M	一个月日均换手	过去 20 个交易日换手率均值
	TURNOVER3M	三个月日均换手	过去 60 个交易日换手率均值
波动	IVR	特异度	1-过去 20 个交易日 Fama-French 三因子回归的拟合度
	ATR1M	一个月真实波动率	过去 20 个交易日日内真实波幅均值
	ATR3M	三个月真实波动率	过去 60 个交易日日内真实波幅均值
分红	DIVIDENDRATE	股息率	最近四个季度预案分红金额/总市值

资料来源：Wind，朝阳永续，天风证券研究所

下面介绍一下我们对于因子的标准化处理流程，主要包括缺失值处理、去极值、标准化、市值和行业中性化。

缺失值处理：对因子值有缺失的股票视情况补其因子值为行业中位数或 0。

去极值：我们采用 MAD（Median Absolute Deviation 绝对中位数法）去极值，对于极值部分将其均匀插值到 3-3.5 倍绝对中位数范围内。具体操作如下，首先计算当期所有股票在因子 f 上的中位数 m_f，然后计算绝对中位数

$$MAD = median(|f - m_f|)$$

采用与 3σ 法等价的方法，保留 $[m_f - 3 \cdot 1.483 \cdot MAD,\ m_f + 3 \cdot 1.483 \cdot MAD]$ 之间股票的因子值不变，取值大于 $m_f + 3 \cdot 1.483 \cdot MAD$ 的所有股票的因子取值按排序均匀压缩到 $[m_f + 3 \cdot 1.483 \cdot MAD, m_f + 3.5 \cdot 1.483 \cdot MAD]$ 之间，取值低于 $m_f - 3 \cdot 1.483 \cdot MAD$ 的所有股票的因子取值按排序均匀压缩到 $[m_f - 3.5 \cdot 1.483 \cdot MAD, m_f - 3 \cdot 1.483 \cdot MAD]$ 之间，这样去除了极值同时也在极值的股票之间保序。

标准化：为了使得构造复合因子时各因子间量纲统一，我们对每个因子进行标准化处理，我们采用 Z-Score 方法来对因子取值标准化，使得因子的均值为 0，标准差为 1，即

$$f' = \frac{f - mean(f)}{std(f)}$$

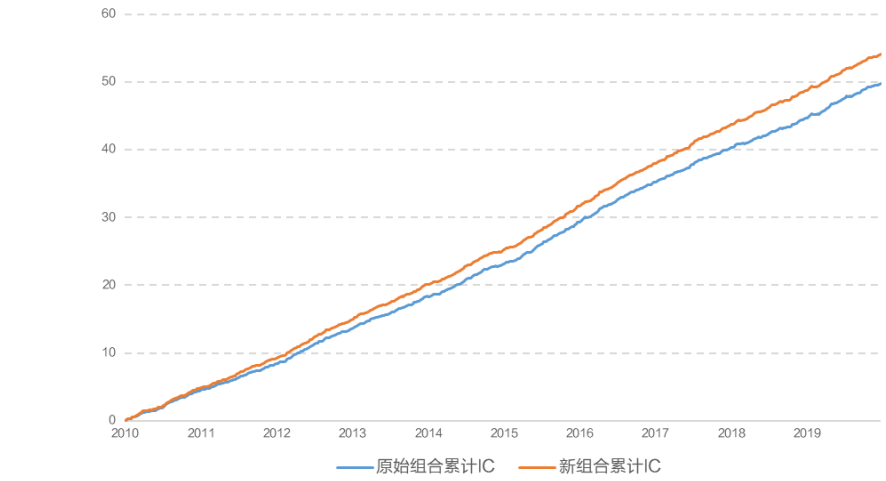
市值和行业中性化：由于因子可能受到市值以及行业的影响较大，因此需要对市值和行业进行中性化处理，即对下式做回归取残差：

$$f_i = \beta^{MV} MV_i + \Sigma_j \beta_j^{ind} X_{ij} + \varepsilon$$

其中 MV_i 为股票 i 的对数总市值，也进行了去极值、标准化的处理，X_{ij} 为股票 i 对于行业 j 的 0-1 哑变量，对回归后得到的残差 ε 继续做去极值、标准化处理得到中性化后的因子取值。我们对市值因子和挖掘因子以外的因子，都进行了行业和市值中性化处理。

因子复合：我们以对称正交后因子滚动 240 期的 ICIR 加权来构建复合因子。这里我们对比一下带/不带挖掘因子的复合因子选股效果。从下图可以看到，添加挖掘因子后复合因子的选股效果确实带来了明显的提升。

图 21：带/不带挖掘因子的复合因子累计周度 IC



资料来源：Wind，天风证券研究所

带/不带挖掘因子的复合因子 IC 统计值如下表。可以看到，复合因子周度 IC 均值从 0.102 提升到 0.111，ICIR 从 11.26 提升到 12.04。

表 12：带/不带挖掘因子的复合因子 IC 统计值

	IC 均值	IC 标准差	年化 ICIR	IC 胜率
不带挖掘因子	0.102	0.064	11.264	94.2%
带挖掘因子	0.111	0.065	12.044	95.7%

资料来源：Wind，天风证券研究所

4.1. 周频多头组合

下面我们首先查看一下周度调仓下纯多头组合的表现情况。模型构建的参数如下：

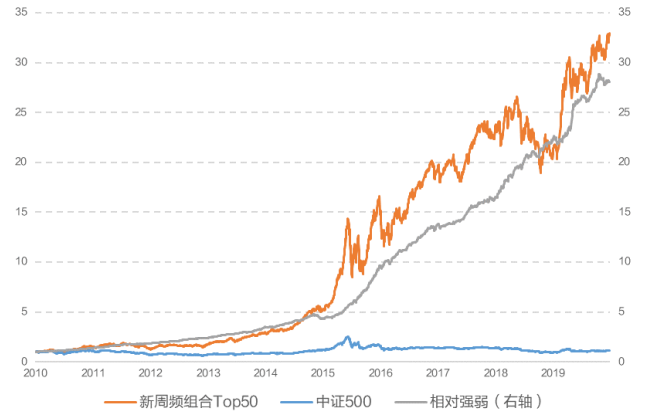
- 回测时间：2010 -2019 年；
- 基准指数：中证 500；
- 交易成本：买入 0.1%，卖出 0.2%；
- 调仓频率：周频；

- 调仓价格：下周第一个交易日的 VWAP；
- 股票池：剔除上市半年以内的新股、ST 股票、ST 摘帽不满 3 个月、退市前 1 个月的股票，调仓时非停牌、涨跌停的股票；过去 20 个交易日日均成交额高于 1000 万；
- 行业及风格约束：无约束；
- 换手率约束：无约束；
- 持仓数量：50 只；
- 个股权重：等权。

由于 A 股停牌、涨跌停经常出现，考虑调仓时股票的可交易性，如调仓遇到上期持仓中的股票停牌、涨跌停时，我们继续持有该股票，即保持该股票本期权重不变。

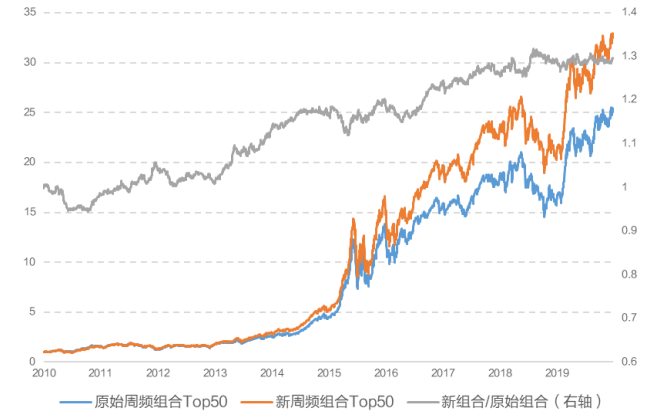
下图展示了多头 Top50 组合的选股效果。可以看到，多头组合能够持续稳定地跑赢中证 500 指数。右图展示了带/不带挖掘因子的多头组合的相对强弱表现，可以看到，带挖掘因子的组合长期能够跑赢不带挖掘因子的原始组合。

图 22：多头 Top50 组合净值走势



资料来源：Wind，天风证券研究所

图 23：带/不带挖掘因子的多头组合净值对比



资料来源：Wind，天风证券研究所

下表展示了多头组合的历年收益情况，组合年化收益 43.26%，年化超额中证 500 数 41.59%，每年超额收益都在 18%以上，相对于中证 500 指数的历史最大回撤-10.4%，信息比 4.94。多头组合周均单边换手 47%。

表 13：周频多头组合分年度表现

年份	绝对收益	指数收益	超额收益	相对最大回撤	收益回撤比	信息比	跟踪误差	回撤高点	回撤低点	月度胜率
2010	54.38%	10.07%	44.31%	-2.63%	16.86	5.16	6.79%	20100416	20100507	91.67%
2011	-12.70%	-33.83%	21.13%	-1.19%	17.81	5.51	5.20%	20110804	20110808	91.67%
2012	31.26%	0.28%	30.98%	-1.58%	19.59	5.67	4.96%	20120620	20120702	100%
2013	67.04%	16.89%	50.15%	-3.68%	13.63	6.18	6.11%	20130823	20130912	100%
2014	74.04%	39.01%	35.03%	-10.38%	3.38	3.31	7.10%	20141106	20141222	83.33%
2015	219.45%	43.12%	176.33%	-3.95%	44.66	8.31	10.11%	20150727	20150803	91.67%
2016	17.15%	-17.78%	34.92%	-3.67%	9.5	5.35	6.97%	20160223	20160304	91.67%
2017	18.23%	-0.20%	18.44%	-3.39%	5.44	3.03	5.85%	20170103	20170119	83.33%
2018	-10.29%	-33.32%	23.03%	-2.95%	7.81	4.17	7.48%	20180831	20181008	91.67%
2019	61.53%	26.38%	35.15%	-3.70%	9.49	3.24	7.86%	20191028	20191127	83.33%
全样本期	43.26%	1.67%	41.59%	-10.40%	4	4.94	7.06%	20141106	20150105	90.83%

资料来源：Wind，天风证券研究所

4.2. 周频指数增强组合

下面我们构建中证 500 指数增强组合并查看添加挖掘因子后增强组合的实际表现。我们采用如下形式的组合优化模型来构建指数增强组合：

$$
\begin{aligned}
max \quad & r^T w \\
s.t. \quad & s_l \le X(w - w_b) \le s_h \\
& h_l \le H(w - w_b) \le h_h \\
& w_l \le w - w_b \le w_h \\
& b_l \le B_b w \le b_h \\
& \mathbf{0} \le w \le l \\
& \mathbf{1}^T w = 1 \\
& \Sigma |w - w_0| \le to_h
\end{aligned}
$$

该优化问题的目标函数为最大化组合收益，其中 $r^T w$ 为组合预期收益，w 为待求解的股票权重向量，r 为收益预测模型计算得到的每只股票的得分。模型的约束条件包括组合在风格因子上的偏离度、行业偏离度、个股偏离度、成分股权重占比控制、个股权重上限控制、换手率约束等。

- 第一个约束条件限制了组合相对于基准指数的风格暴露，X 为股票对风格因子的因子暴露矩阵，w_b 为基准指数成分股的权重向量，s_l, s_h 分别为风格因子相对暴露的下限及上限；
- 第二个约束条件限制了组合相对于基准指数的行业偏离，H 为股票的行业暴露矩阵，当股票 i 属于行业 j 时，H_{ji} 为 1，否则为 0；h_l, h_h 分别为组合行业偏离的下限以及上限；
- 第三个约束条件限制了个股相对于基准指数成分股的偏离，w_l, w_h 分别为个股偏离的下限以及上限；
- 第四个约束条件限制了组合在成分股内权重的占比下限及上限，B_b 为个股是否属于基准指数成分股的 0-1 向量，b_l, b_h 分别为成分股内权重的下限以及上限；
- 第五个约束条件限制了卖空，并且限制了个股权重上限 l；
- 第六个约束条件要求权重和为 1，即组合始终满仓运作；
- 第七个约束条件约束了组合的换手率，w_0 为上一期的持仓权重，to_h 为换手率上限。

上述模型中目标函数、风格偏离约束、个股权重偏离约束、成分股权重占比约束、换手率约束都可以转化成线性约束，因此可以通过线性规划来求解。

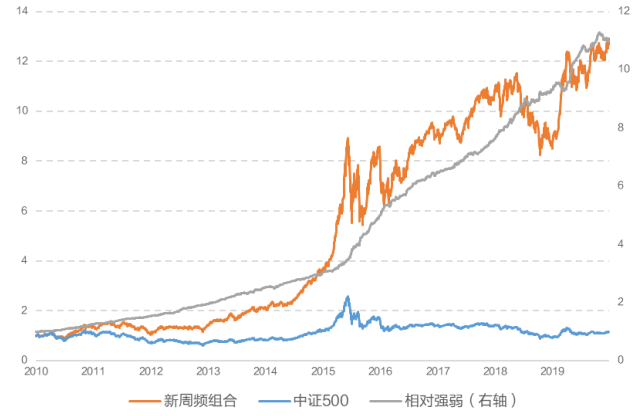
中证 500 指数增强组合构建的参数如下：

- 回测时间：2010 -2019 年；
- 基准指数：中证 500；
- 交易成本：买入 0.1%，卖出 0.2%；
- 调仓频率：周频；
- 调仓价格：下周第一个交易日的 VWAP；
- 股票池：剔除上市半年以内的新股、ST 股票、ST 摘帽不满 3 个月、退市前 1 个月的股票，调仓时非停牌、涨跌停的股票；过去 20 个交易日日均成交额高于 1000 万；

- 行业及风格约束：中信一级行业零暴露，规模因子零暴露；
- 个股权重偏离上限：1%；
- 换手率约束：周度单边 25%。

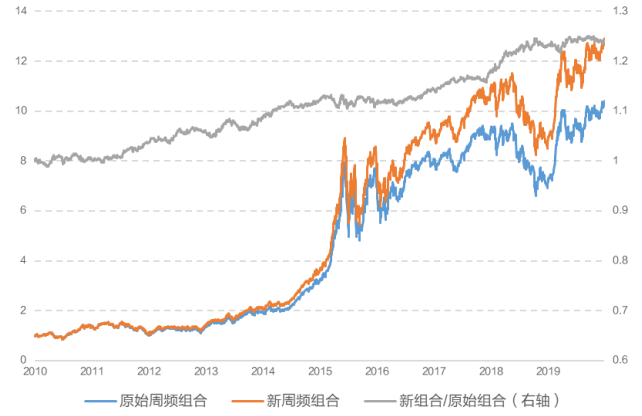
下图展示了指数增强组合的选股效果。可以看到，增强组合能够持续稳定地跑赢中证 500 指数。右图展示了带/不带挖掘因子的指数增强组合的相对强弱表现，可以看到，带挖掘因子的组合长期能够跑赢不带挖掘因子的原始组合。

图 24：中证 500 指数增强组合净值走势



资料来源：Wind，天风证券研究所

图 25：带/不带挖掘因子的中证 500 指数增强组合净值对比



资料来源：Wind，天风证券研究所

下表展示了增强组合的历年收益情况，组合年化超额中证 500 指数 28.43%，每年超额收益都在 14%以上，相对历史最大回撤-4.17%，信息比 5.21，月度胜率 93.33%。

表 14：周频中证 500 指数增强组合分年度表现

年份	绝对收益	指数收益	超额收益	相对最大回撤	收益回撤比	信息比	跟踪误差	回撤高点	回撤低点	月度胜率
2010	37.99%	10.07%	27.93%	-2.56%	10.9	4.62	5.03%	20100430	20100507	91.67%
2011	-19.53%	-33.83%	14.30%	-1.15%	12.42	5.79	3.47%	20110222	20110309	100%
2012	28.59%	0.28%	28.31%	-1.17%	24.27	7.11	3.60%	20120731	20120808	100%
2013	50.17%	16.89%	33.28%	-2.77%	12	5.88	4.46%	20130823	20130911	100%
2014	66.14%	39.01%	27.13%	-2.46%	11.03	4.08	4.46%	20140207	20140227	91.67%
2015	137.69%	43.12%	94.57%	-1.52%	62.21	7.45	7.02%	20150707	20150708	100%
2016	6.93%	-17.78%	24.70%	-1.63%	15.13	6.21	4.36%	20161206	20161219	100%
2017	16.87%	-0.20%	17.07%	-1.02%	16.75	4.51	3.56%	20170804	20170815	100%
2018	-18.34%	-33.32%	14.98%	-2.28%	6.57	4.29	4.88%	20180713	20180808	91.67%
2019	49.43%	26.38%	23.04%	-4.17%	5.53	3.29	5.12%	20190130	20190321	58.33%
全样本期	30.09%	1.67%	28.43%	-4.17%	6.82	5.21	4.74%	20190130	20190321	93.33%

资料来源：Wind，天风证券研究所

4.3. 日频指数增强组合

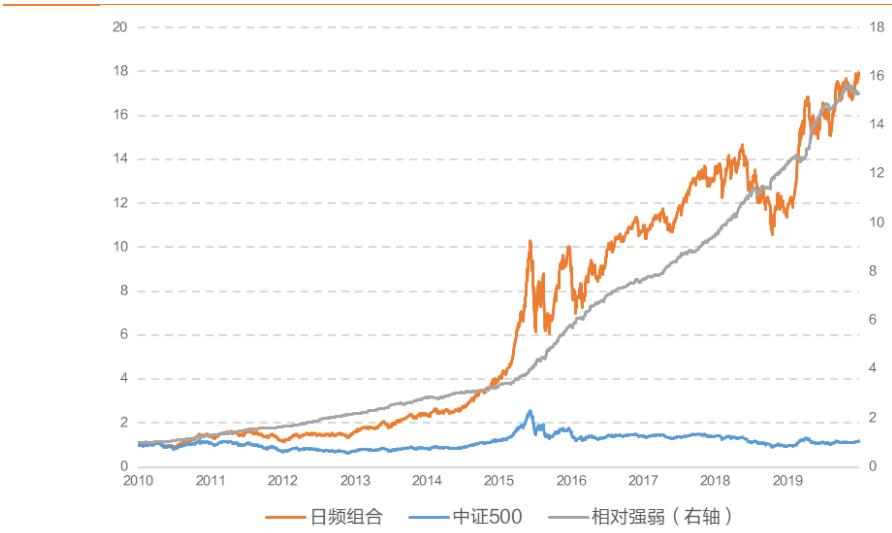
前面我们对比了周频调仓的指数增强组合表现，由于因子取值每天都在更新，我们在周度因子打分的基础上，进一步构建日频调仓的中证 500 指数增强组合，组合构建参数如下：

- 回测时间：2010 -2019 年；
- 基准指数：中证 500；

- 交易成本：买入 0.1%，卖出 0.2%；
- 调仓频率：日频；
- 调仓价格：下一个交易日的 VWAP；
- 股票池：剔除上市半年以内的新股、ST 股票、ST 摘帽不满 3 个月、退市前 1 个月的股票，调仓时非停牌、涨跌停的股票；过去 20 个交易日日均成交额高于 1000 万；
- 行业及风格约束：中信一级行业零暴露，规模因子零暴露；
- 个股权重偏离上限：1%；
- 换手率约束：日度单边 10%。

下图展示了日频调仓中证 500 指数增强组合的选股效果。可以看到，增强组合能够持续稳定地跑赢中证 500 指数。

图 26：日频中证 500 指数增强组合净值走势



资料来源：天风证券研究所

下表展示了日频增强组合的历年收益情况，组合年化超额中证 500 数 32.91%，每年超额收益都在 18%以上，相对于中证 500 指数的历史最大回撤-3.09%，信息比 5.72，月度胜率 94.17%。

表 15：日频中证 500 指数增强组合分年度表现

年份	绝对收益	指数收益	超额收益	相对最大回撤	收益回撤比	信息比	跟踪误差	回撤高点	回撤低点	月度胜率
2010	41.87%	10.07%	31.80%	-2.74%	11.61	4.57	5.73%	20100430	20100507	83.33%
2011	-15.30%	-33.83%	18.53%	-0.76%	24.42	6.75	3.74%	20110428	20110504	100%
2012	32.28%	0.28%	32%	-0.66%	48.19	8.04	3.54%	20120731	20120808	100%
2013	52.74%	16.89%	35.85%	-3.09%	11.59	6.33	4.44%	20130823	20130910	100%
2014	62.28%	39.01%	23.27%	-2.40%	9.69	3.54	4.45%	20140108	20140304	91.67%
2015	151.09%	43.12%	107.97%	-1.93%	55.9	7.77	7.58%	20150723	20150803	100%
2016	8.56%	-17.78%	26.34%	-1.84%	14.29	6.41	4.49%	20161206	20161215	100%
2017	22.73%	-0.20%	22.93%	-0.86%	26.77	6.19	3.41%	20170825	20170913	100%
2018	-12.62%	-33.32%	20.70%	-2.26%	9.17	5.76	4.87%	20180719	20180803	91.67%
2019	55.85%	26.38%	29.46%	-2.70%	10.91	4.09	5.19%	20191104	20191231	75%
全样本期	34.57%	1.67%	32.91%	-3.09%	10.65	5.72	4.93%	20130823	20130910	94.17%

资料来源：Wind，天风证券研究所

5. 总结

基因表达式规划下的价量因子挖掘

多因子模型能够持续改进的核心是持续有效地发现有显著选股能力的因子。基因表达式规划是一种启发式算法，其借鉴生物基因进化的思想，能够通过不断变异与进化来发现更好的解。因此，本报告中我们基于基因表达式规划来挖掘有效的价量因子。

在短周期价量数据构造的因子中，大部分因子选股效果的单调性并不显著，还有很多因子的多头没有超额收益，因此在设置因子有效性筛选指标时，我们结合了因子的 ICIR、多头超额收益以及分组收益的单调性，综合考察因子的选股效果，实际挖掘出的因也具有较为单调和显著的选股效果。

基于挖掘因子构建指数增强组合

我们将挖掘得到的周频价量因子与传统基于基本面的因子结合来构建多头组合与中证 500 指数增强组合，相比于不带挖掘因子的传统基本面因子组合，组合的收益提升显著。

- 周频调仓的多头 Top50 等权组合，年化收益 43.3%，相对于中证 500 指数年化超额 41.6%，每年都能跑赢中证 500 指数 18%以上。
- 周频调仓的中证 500 指数增强组合，年化超额收益 28.4%，信息比 5.2，，每年都能跑赢中证 500 指数 14%以上。
- 日频调仓的中证 500 指数增强组合，年化超额收益 32.9%，信息比 5.7，每年都能跑赢中证 500 指数 18%以上。

6. 参考文献

- **[Ferreira 2001]** Ferreira, Candida. "Gene expression programming: a new adaptive algorithm for solving problems." arXiv preprint cs/0102027 (2001).
- **[He 2014]** He, Xinran, et al. "Practical lessons from predicting clicks on ads at facebook." Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. 2014.
- **[Huang 2012]** Huang, Chien-Feng, et al. "Feature Selection and Parameter Optimization of a Fuzzy-based Stock Selection Model Using Genetic Algorithms." International Journal of Fuzzy Systems 14.1 (2012).
- **[Ying 2007]** Becker, Ying L., Peng Fei, and Anna M. Lester. "Stock selection: An innovative application of genetic programming methodology." Genetic Programming Theory and Practice IV. Springer, Boston, MA, 2007. 315-334.

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

一般声明

特别声明

投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下

天风证券研究

北京	武汉	上海	深圳
北京市西城区佟麟阁路 36 号	湖北武汉市武昌区中南路 99 号保利广场 A 座 37 楼	上海市浦东新区兰花路 333 号 333 世纪大厦 20 楼	深圳市福田区益田路 5033 号平安金融中心 71 楼
邮编：100031	邮编：430071	邮编：201204	邮编：518000
邮箱：research@tfzq.com	电话：(8627)-87618889	电话：(8621)-68815388	电话：(86755)-23915663
	传真：(8627)-87618863	传真：(8621)-68812910	传真：(86755)-82571995
	邮箱：research@tfzq.com	邮箱：research@tfzq.com	邮箱：research@tfzq.com