



## 衍生品量化择时系列专题（七）： 基于遗传规划的期货因子挖掘

报告日期：2022年5月6日

### ★研究背景：

在机器学习领域，遗传规划能够有效地进行特征生成。一些早期的实证研究已经将遗传规划运用于金融工程的建模当中，然而纵观近些年在遗传规划领域的研究，多数集中在股票因子的挖掘之中。在量化CTA领域中，随着基本面量化的不断发展，本报告尝试结合基本面量化与遗传规划模型，力求挖掘具有强交易信号的期货基本面量化因子。

### ★模型构建：

本报告采用的遗传规划的方式是：将降维之后的基本面因子作为初始种群，以因子的夏普值作为适应度函数进行演化迭代。

**降维：**报告尝试利用不同的降维方式对期货的基本面数据进行降维，包括PCA、KPCA、LLE、LDA、MDS和Isomap，为统一起见，降维维度均设置为12；

**回测：**利用普通最小二乘法进行线性回归，最优拟合曲线使各点到直线的距离的平方和（残差平方和RSS）最小；

### ★模型结果：

铝的回测结果显示，经过遗传规划演化后的因子将夏普值表现从1.43提升至1.55；原油的回测结果显示，经过遗传规划演化后的因子将夏普值表现从1.96提升至2.12；其余品种也表现出不同幅度的提升。

### ★风险提示：

市场风格的切换会造成特征有效性发生变化，导致模型效果下降。

**致谢：**实习生徐皓茜对本报告亦有贡献。

王冬黎 高级分析师(金融工程)  
从业资格号：F3032817  
投资咨询号：Z0014348  
Tel: 8621-63325888-3975  
Email: [dongli.wang@orientfutures.com](mailto:dongli.wang@orientfutures.com)

联系人：谢怡伦（分析师）  
从业资格号：F03091687  
Tel: 8621-63325888-1585  
Email: [yilun.xie@orientfutures.com](mailto:yilun.xie@orientfutures.com)

### 相关报告

《衍生品量化择时系列专题（二）——螺纹钢指标筛选与大类因子研究》

《衍生品量化择时系列专题（三）——PTA指标筛选与大类因子合成研究》

《衍生品量化择时系列专题（五）——基于机器学习的螺纹钢价格周度预测》

## 目录

|                   |    |
|-------------------|----|
| 1、研究背景.....       | 5  |
| 2、模型构建.....       | 6  |
| 2.1、遗传规划构建因子..... | 6  |
| 2.2、降维.....       | 7  |
| 2.2.1、PCA.....    | 8  |
| 2.2.2、KPCA.....   | 8  |
| 2.2.3、LLE.....    | 9  |
| 2.2.4、LDA.....    | 9  |
| 2.2.5、MDS.....    | 9  |
| 2.2.6、ISOMAP..... | 10 |
| 2.3、函数库.....      | 10 |
| 2.4、回测参数.....     | 11 |
| 2.5、数据处理.....     | 11 |
| 3、铝的单因子回测.....    | 12 |
| 4、铝的遗传规划.....     | 12 |
| 4.1、PCA.....      | 14 |
| 4.2、KPCA.....     | 14 |
| 4.3、LDA.....      | 15 |
| 4.4、MDS.....      | 15 |
| 4.5、Isomap.....   | 15 |
| 5、原油的单因子检验.....   | 16 |
| 6、原油的遗传规划.....    | 17 |
| 6.1、PCA.....      | 18 |
| 6.2、KPCA.....     | 19 |
| 6.3、MDS.....      | 19 |
| 6.4、Isomap.....   | 20 |
| 7、其他品种.....       | 20 |
| 7.1、镍.....        | 20 |

|               |    |
|---------------|----|
| 7.2、锡.....    | 21 |
| 7.3、锌.....    | 21 |
| 7.4、LPG.....  | 21 |
| 7.5、沥青.....   | 22 |
| 7.6、PTA.....  | 22 |
| 7.7、PVC.....  | 22 |
| 7.8、硅铁.....   | 23 |
| 7.9、豆粕.....   | 23 |
| 7.10、豆油.....  | 24 |
| 7.11、棕榈油..... | 24 |
| 8、结论.....     | 24 |

## 图表目录

|                           |    |
|---------------------------|----|
| 图表 1: Alpha101 因子展示.....  | 5  |
| 图表 2: 遗传规划流程.....         | 6  |
| 图表 3: 编码方式.....           | 7  |
| 图表 4: 交叉.....             | 7  |
| 图表 5: 流行体.....            | 9  |
| 图表 6: 函数库.....            | 10 |
| 图表 7: 滚动回归.....           | 11 |
| 图表 8: 铝的单因子回测指标直方图.....   | 12 |
| 图表 9: 铝的初始种群.....         | 13 |
| 图表 10: 铝的种群演化.....        | 13 |
| 图表 11: 样本外表现.....         | 14 |
| 图表 12: 铝的 PCA 降维.....     | 14 |
| 图表 13: 铝的 KPCA 降维.....    | 15 |
| 图表 14: 铝的 LDA 降维.....     | 15 |
| 图表 15: 铝的 MDS 降维.....     | 15 |
| 图表 16: 铝的 Isomap 降维.....  | 16 |
| 图表 17: 原油的单因子回测指标直方图..... | 17 |
| 图表 18: 原油的初始种群.....       | 17 |
| 图表 19: 原油的遗传规划.....       | 18 |
| 图表 20: 样本外表现.....         | 18 |
| 图表 21: 原油的 PCA 降维.....    | 19 |
| 图表 22: 原油的 KPCA 降维.....   | 19 |
| 图表 23: 原油的 MDS 降维.....    | 19 |
| 图表 24: 原油的 Isomap 降维..... | 20 |
| 图表 25: 镍的遗传规划.....        | 21 |
| 图表 26: 锡的遗传规划.....        | 21 |
| 图表 27: 锌的遗传规划.....        | 21 |
| 图表 28: LPG 的遗传规划.....     | 22 |
| 图表 29: 沥青的遗传规划.....       | 22 |
| 图表 30: PTA 的遗传规划.....     | 22 |
| 图表 31: PVC 的遗传规划.....     | 23 |
| 图表 32: 硅铁的遗传规划.....       | 23 |
| 图表 33: 豆粕的遗传规划.....       | 23 |
| 图表 34: 豆油的遗传规划.....       | 24 |
| 图表 35: 棕榈油的遗传规划.....      | 24 |

## 1、研究背景

在机器学习领域，遗传规划能够有效地进行特征生成。一些早期的实证研究已经将遗传规划运用于金融工程的建模当中，并得到可用于实盘的交易因子。然而纵观近些年在遗传规划领域的研究，多数研究集中在对股票因子的挖掘之中，一个比较典型的例子就是Worldquant在2015年发布的Alpha 101报告，该报告展示了101个在美股市场上具有一定信号的股票因子，通过分析不难发现，这些因子大多具有较为复杂的表达形式，从主观交易逻辑的层面缺乏一定的可解释性，从其表达式的结构上来看可以推断出其交易因子具有遗传规划模型的痕迹。在此之后，市场上出现了大量遗传规划相关的研究报告，最终目的均大同小异，即找出更强的交易信号。

需要着重强调的是，遗传规划寻找信号的过程中具有一些弊端，其最大的短板就是由于信号的不可解释性，故无法保证其在样本外是否依然有效，所以在遗传规划的过程中，需要检验因子的鲁棒性以考量该因子对参数的敏感程度。

图表 1：Alpha101 因子展示

Alpha#1: (rank(Ts\_ArgMax(SignedPower(((returns < 0) ? stddev(returns, 20) : close), 2.), 5)) - 0.5)

Alpha#2: (-1 \* correlation(rank(delta(log(volume), 2)), rank(((close - open) / open)), 6))

Alpha#3: (-1 \* correlation(rank(open), rank(volume), 10))

Alpha#4: (-1 \* Ts\_Rank(rank(low), 9))

Alpha#5: (rank((open - (sum(vwap, 10) / 10))) \* (-1 \* abs(rank((close - vwap)))))

Alpha#6: (-1 \* correlation(open, volume, 10))

Alpha#7: ((adv20 < volume) ? ((-1 \* ts\_rank(abs(delta(close, 7)), 60)) \* sign(delta(close, 7))) : (-1 \* 1))

资料来源：101 Formulaic Alphas by Worldquant, LLC

目前在量化CTA领域，遗传规划的应用相对而言比较有限。一是因为，量化CTA市场自身的发展程度不及股票量化，故市场的关注度受限；二是因为，量化CTA的数据不及股票市场完备，且品种数量远不及股票数量那么庞大。在目前市场上关于遗传规划在量化CTA中的应用的报告均使用期货的价格量数据，其因子挖掘的本质与股票因子挖掘类似，即通过某品种期货的价格量数据代入遗传规划模型进行因子挖掘。这样的因子挖掘方式能够在一定程度上构建期货的价格量因子，但在期货研究领域，期货的基本面数据也应得到更加广泛的研究及应用。现实的市场环境已经表明，期货的基本面因

子相较于价量因子能够提供更为全面的期货价格的预测。

另一方面，作者发现，在CTA领域，近些年随着基本面量化的发展，越来越多的模型尝试以期货的基本面数据作为输入，去预测期货价格的未来收益率。本报告基于以上的背景，尝试结合基本面量化与遗传规划模型，力求挖掘具有强交易信号的期货基本面量化因子。

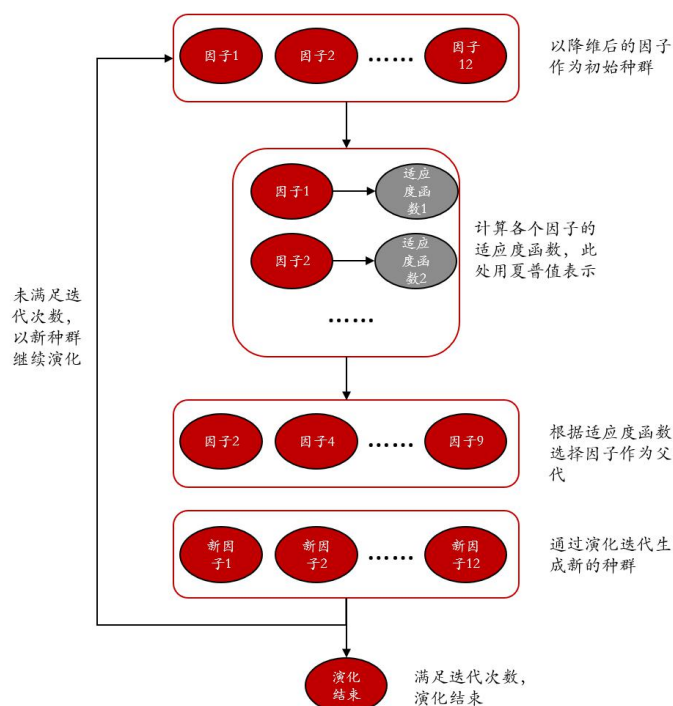
## 2、模型构建

本报告采用的遗传规划的流程是：将降维之后的基本面因子作为初始种群，以因子的夏普值作为适应度函数进行演化迭代。为避免表达式过度复杂，因子迭代次数设置为4次。

### 2.1、遗传规划构建因子

遗传规划是一种启发式的公式演化技术，适用于特征工程。其主要流程为基于初始种群设置相应的适应度函数，根据适应度函数筛选因子，之后通过筛选出来的因子演化生成新的种群，判断是否满足迭代次数，若满足，则演化结束；若未满足，则代入之前流程继续迭代。

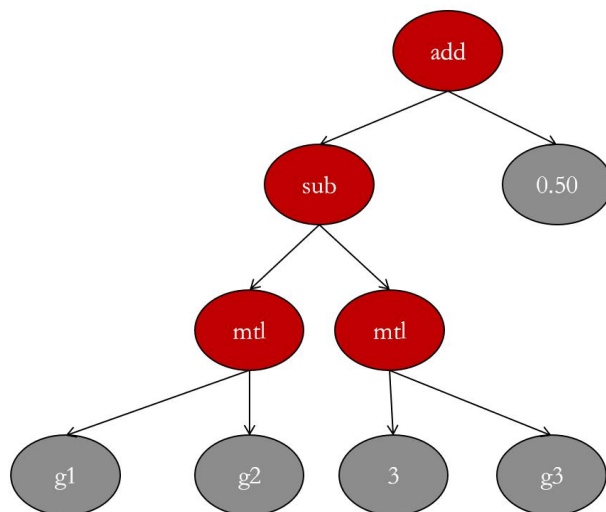
图表2：遗传规划流程



资料来源：东证衍生品研究院

针对公式化的因子，首先需要对其进行编码使其能够适用到遗传规划算法当中，下面以一个简单的因子表达式为例来演示具体的编码方式：

图表 3：编码方式

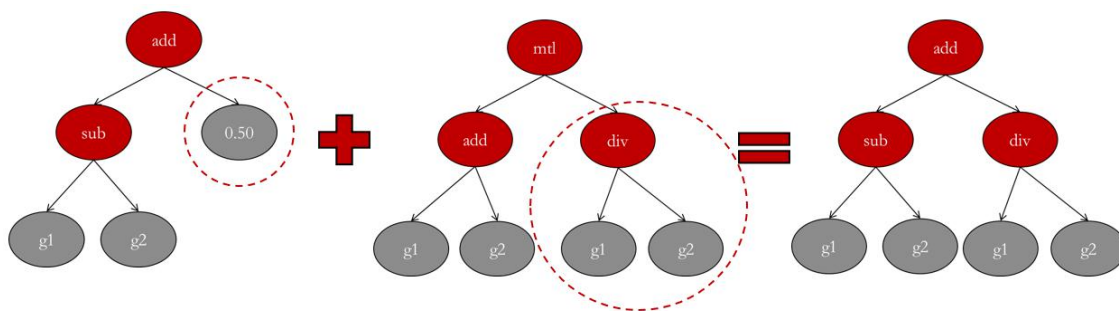


公式：  $y = g1 * g2 - 3g3 + 0.5$   
在遗传规划表达式采用树形结构表示，  
结构为：  $y = (+(-(* g1g2)(* 3g3))0.5)$

资料来源：东证衍生品研究院

在对因子表达式进行树形结构的编码之后，便可以对其进行交叉、变异等操作。所谓交叉，就是结合两个适应度较强因子的基因信息，以生成相应的新因子，以其中一棵树的子树（节点）替代另一颗树的子树（节点）即可以完成此操作。

图表 4：交叉



资料来源：东证衍生品研究院

## 2.2、降维

针对期货的基本面数据，本报告提出首先对数据进行降维的操作。由于基本面数据彼此之间的相关性较高，且单条数据对期货价格的解释强度低于价量数据，基于以上原



因，考虑首先对基本面数据进行降维操作，一方面可以提升单个因子的解释力度，另一方面可以降低模型的复杂程度，以加快模型运行效率。

本报告中涉及多种降维方式，以下针对不同降维方式作简要介绍：

### 2.2.1、PCA

主成分分析算法（PCA）是最常用的线性降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中，并期望在所投影的维度上数据的信息量最大（方差最大），以此使用较少的数据维度，同时保留住较多的原数据点的特性。

PCA 降维的目的，就是为了在尽量保证“信息量不丢失”的情况下，对原始特征进行降维，也就是尽可能将原始特征往具有最大投影信息量的维度上进行投影。将原特征投影到这些维度上，使降维后信息量损失最小。其算法步骤如下：

设有  $m$  条  $n$  维数据：

- 将原始数据按列组成  $n$  行  $m$  列矩阵  $X$ ；
- 将  $X$  的每一行进行零均值化，即减去这一行的均值；
- 求出协方差矩阵  $C = \frac{1}{m}XX^T$ ；
- 求出协方差矩阵的特征值及对应的特征向量；
- 将特征向量按对应特征值大小从上到下按行排列成矩阵，取前  $k$  行组成矩阵  $P$ ；
- $Y = PX$  即为降维到  $k$  维后的数据。

在进行 PCA 降维之后，可以缓解维度灾难，并对数据进行降噪，同时将数据压缩到低维之后，使得降维之后的数据各特征相互独立。但是在另一方面，由于 PCA 保留了主要信息，舍弃了一些看似无用的信息，但这些“无用信息”只是在训练集上没有有效表现，因此产生了过拟合的可能性，这一问题在模型训练时需要注意。

### 2.2.2、KPCA

针对非线性的数据，KPCA（核函数主成分分析法）将非线性可分的数据转换到一个适合对其进行线性分类的新的低维子空间上。利用核 PCA 可以通过非线性映射将数据转换到一个高维空间中，在高维空间中使用 PCA 将其映射到另一个低维空间中，并通过线性分类器对样本对其划分。

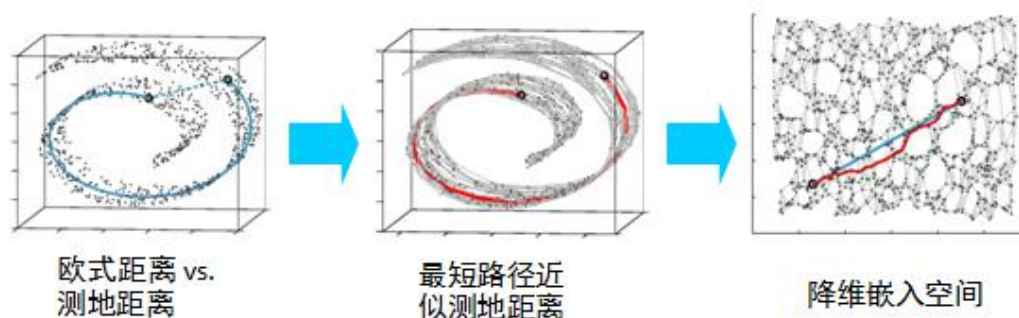


### 2.2.3、LLE

局部线性嵌入(Locally Linear Embedding, 以下简称 LLE)也是非常重要的降维方法。和传统的 PCA, LDA 等关注样本方差的降维方法相比, LLE 关注于降维时保持样本局部的线性特征, 由于 LLE 在降维时保持了样本的局部特征, 它广泛的用于图像识别, 高维数据可视化等领域。LLE 属于流形学习(Manifold Learning)的一种。流形学习是一大类基于流形的框架。数学意义上的流形比较抽象, 不过可以认为 LLE 中的流形是一个不闭合的曲面。这个流形曲面有数据分布比较均匀, 且比较稠密的特征。基于流行的降维算法就是将流形从高维到低维的降维过程, 在降维的过程中我们希望流形在高维的一些特征可以得到保留。在 LLE 方法中, 主要有如下的基本假设:

- 一个流形的局部可以近似于一个欧式空间;
- 每个样本均可以利用其邻居进行线性重构。

图表 5: 流行体



资料来源: Zhang, Zhenyue, and Hongyuan Zha. "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment." SIAM journal on scientific computing 26.1 (2004): 313-338.

### 2.2.4、LDA

LDA 是一种监督学习的降维技术, 也就是说它的数据集的每个样本是有类别输出的。这点和 PCA 不同。PCA 是不考虑样本类别输出的无监督降维技术。LDA 的思想可以用一句话概括, 就是“投影后类内方差最小, 类间方差最大”。

### 2.2.5、MDS

多维尺度变换算法解决的问题是: 当  $n$  个对象之间的相似性给定, 确定这些对象在低维空间中的表示, 并使其尽可能与原先的相似性大致匹配。高维空间中每一个点代表一个对象, 因此点与点之间的距离和对象之间的相似度高度相关。可以这么理解, 两个相似的对象在高维空间中由两个距离相近的点所表示, 两个不相似的对象在高维空间中由两个距离比较远的点表示。其基本思想是将高维坐标中的点投影到低维空间中, 保持点彼此之间的相似性尽可能不变。

## 2.2.6、ISOMAP

本质上讲，ISOMAP 与前面讲过的 MDS（降维方法之 MDS）是一模一样的，只不过它考虑使用一个合适的距离度量  $d_{ij}$  使得该降维方法能够适用于流形的数据。其算法流程为根据给定的欧氏距离阵，计算得到调整的距离阵；再将调整的距离阵代入 MDS 算法框架中。

## 2.3、函数库

本报告构建了一系列函数组成的函数库用于种群的演化，包括单元函数和多元函数，除了传统的多个序列之间的算符，还加入了针对单个序列的时序处理函数。具体函数列表如下：

图表 6：函数库

| 函数               | 说明      |
|------------------|---------|
| ts_fill(f)       | 前值填充    |
| ts_ratio(f,n)    | 环比/同比   |
| ts_diff(f,n)     | 差分      |
| ts_corr(f1,f2,n) | 序列相关性   |
| ts_cov(f1,f2,n)  | 序列协方差   |
| ts_rank(f,n)     | 时序排序    |
| add(f1,f2)       | 相加      |
| sub(f1,f2)       | 相减      |
| myl(f1,f2)       | 相乘      |
| div(f1,f2)       | 相除      |
| sqrt(f)          | 开方      |
| abs(f)           | 绝对值     |
| log(f)           | 取对数     |
| square(f)        | 平方      |
| exp(f)           | 以e为底的指数 |
| sign(f)          | 数值正负判断  |
| ceil(f)          | 向下取整    |
| floor(f)         | 向上取整    |
| rint(f)          | 四舍五入    |
| scalc(f)         | 数据等比例压缩 |

资料来源：东证衍生品研究院

以  $ts\_rank$  函数为例，该函数有两个输入，其中  $f$  为数据， $n$  为时间参数，其定义为对某个数据在其时间序列上对过去  $n$  的数据进行排序，若序列过去 5 天的数据值为 [3, 5, 6, 1, 10]，则其经过  $ts\_rank(f,5)$  的操作之后，对应新的序列为 [0.4, 0.6, 0.8, 0.2, 1.0]，以此类推。

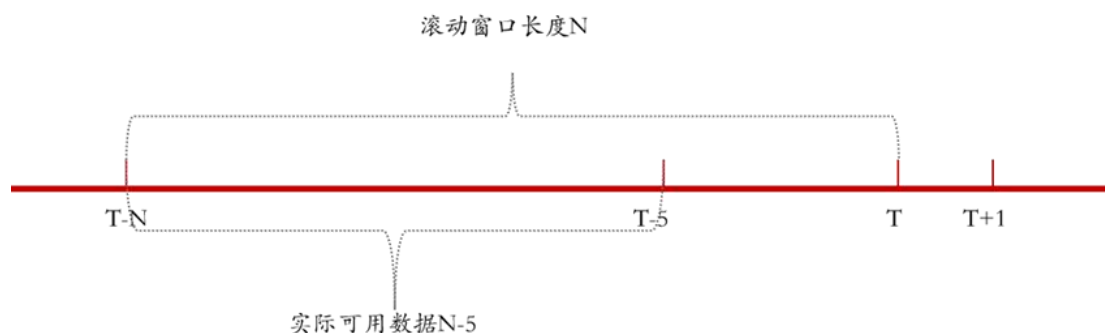
## 2.4、回测参数

本报告依据 OLS 进行滚动回归预测。OLS（普通最小二乘法）多元回归的原理为，最优拟合曲线使得各点到直线的距离的平方和（残差平方和 RSS）最小：

$$RSS = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2$$

本报告采用滚动回归的方式进行回测，以避免使用未来数据。以周度预测为例，首先设置相应的滚动回归窗口长度 N，对每一天 T 都截取 T-N 到 T 时间段的基本面数据，由于为周度预测，需要对基本面数据进行 5 天的移动处理，所以实际可用数据点为 N-5 天内的数据，随后根据训练模型得到一系列预测值，再根据预测值与真实值的比较去构造回测模型。回测过程基于历史窗口长度为 60 个交易日的基本面数据对未来 5 个交易日的价格进行滚动回测。

图表 7：滚动回归



资料来源：东证衍生品研究院

## 2.5、数据处理

**频率调整：**基本面的原始数据多为低频数据（月频或周频），为便于处理，将所有数据前值填充为日频数据；

**标准化：**对所有填充后的数据进行 z-score 标准化处理，提高数据之间的可比性；

**异常值处理：**对所有偏离均值 3 个标准差的数据进行处理；

**移仓换月处理：**为避免期货展期导致的价格影响，本报告以豆粕复权价格进行回测；

**周期性影响：**为剔除数据的周期性影响，对数据分别进行环比，同比等处理；

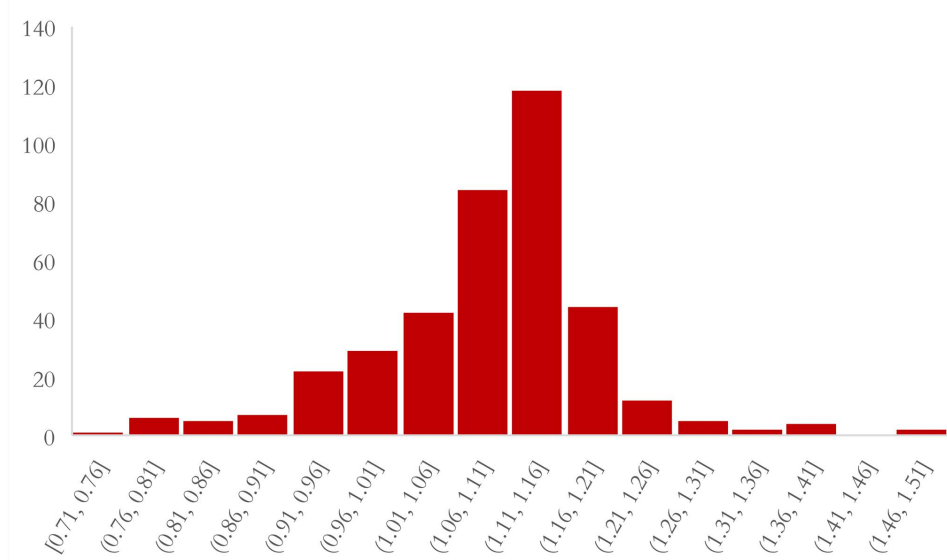
**可得性处理：**按指标具体可得性进行相应滞后处理；

**回测参数：**本报告基于日频数据生成周频级别的多空信号，每周更新仓位，手续费设置为双边万三。

### 3、铝的单因子回测

基于繁微数据平台，本报告共整理铝相关基本面因子共 383 个，这些因子涵盖了铝的上下游产业链相关数据，包括铝原料，铝冶炼，铝回收，铝加工，铝库存等基本面数据，数据来源方包括 Wind，上海钢联，Bloomberg。以经过处理后的 383 个基本面因子作为单因子对铝期货价格进行预测，计算其回测结果的夏普值，据此绘制直方分布图。观察该直方图可以发现，其呈现类左偏的分布。多数因子的夏普值集中在 [1.06, 1.16] 这个范围区间里面，只有极少数因子的夏普值超过 1.31，说明针对单因子而言，其预测能力存在一定上限。

图表 8：铝的单因子回测指标直方图



资料来源：东证衍生品研究院

在这 383 个单因子的夏普值中，表现最好的单因子达到了 1.48，意味着该单因子具有较好的预测能力，然后作为单个基本面数据指标而言，其预测能力缺乏一定的稳定性，因此需要对基本面作降维操作，以使得单个因子具有更多的信息量，稳定性得到增强。

### 4、铝的遗传规划

在对铝的基本面数据进行单因子的考量之后，本部分尝试对铝的基本面数据进行降维，用以提取数据信息量更高的因子，本报告为统一起见，选择的降维维度均设置为 12。

首先，本报告采用 LLE 降维方式对铝的基本面因子进行降维，将降维后的因子作为模型的初始种群。对于所有种群中的个体，本报告关注的是因子的夏普值，夏普值相较于年华收益能够更好地衡量策略的稳定程度，更加贴合本报告对策略优劣的评判标

准，基于以上考虑，采用夏普值而非年华收益等参数作为种群个体的适应度函数。

下图展示了初始种群内各因子的夏普值表现，可以发现经过降维后的因子的整体表现要好于未经降维的单因子，其中表现最好的因子为 g2，其夏普值达到了 1.43，高于未经降维的大部分单因子。

图表 9：铝的初始种群

|      | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群 | 1.20 | 1.43 | 1.12 | 1.22 | 1.02 | 0.90 | 1.07 | 1.31 | 1.31 | 1.22 | 1.11 | 0.99 | 1.43 |

资料来源：东证衍生品研究院

基于初始种群，进行遗传规划模型训练，由于演化次数的增加会导致因子的复杂度不断增加，为避免过拟合情况出现，设置迭代次数为 4 次。每次演化生成新的种群数据均为 12，之后在每一次演化之后基于新生种群选择出表现最好的个体作为每次演化的最佳值。由于在遗传规划模型不断迭代的过程中，我们关注的始终是模型的最佳输出，因此在每一次迭代之后，通过考察当前种群的最佳值，以此作为每次迭代之后新生种群的衡量标准。

图表 10：铝的种群演化

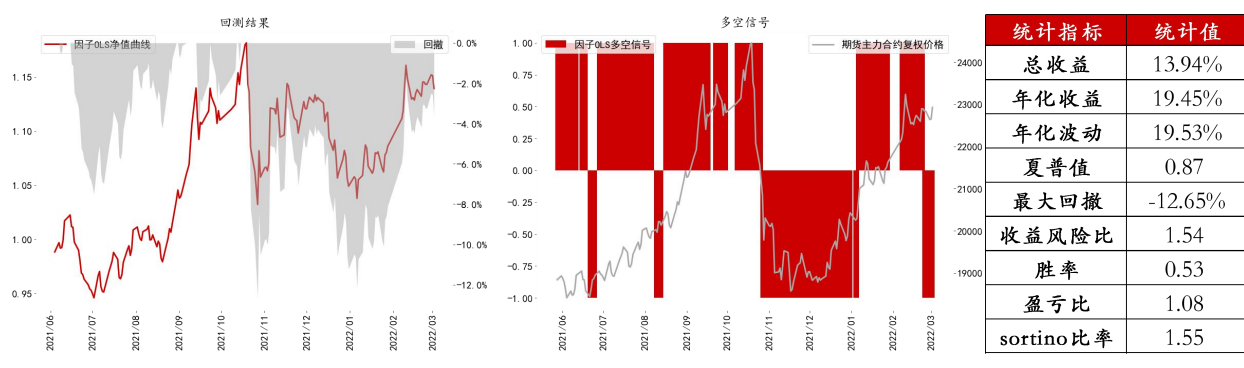
|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.20 | 1.43 | 1.12 | 1.22 | 1.02 | 0.90 | 1.07 | 1.31 | 1.31 | 1.22 | 1.11 | 0.99 | 1.43 |
| 演化次数1 | 1.31 | 0.80 | 0.92 | 1.49 | 1.44 | 1.11 | 0.92 | 1.48 | 0.92 | 1.13 | 1.01 | 1.25 | 1.49 |
| 演化次数2 | 1.11 | 1.19 | 1.26 | 1.07 | 1.48 | 1.44 | 1.08 | 0.84 | 0.80 | 0.76 | 0.80 | 0.80 | 1.48 |
| 演化次数3 | 1.55 | 1.34 | 1.08 | 0.91 | 0.75 | 1.08 | 0.85 | 1.48 | 1.06 | 1.09 | 0.79 | 1.08 | 1.55 |
| 演化次数4 | 1.08 | 1.08 | 1.32 | 1.10 | 1.51 | 0.98 | 0.85 | 0.80 | 0.80 | 0.80 | 1.41 | 1.07 | 1.51 |

资料来源：东证衍生品研究院

通过上表可以发现在经过一次演化之后，种群的最优值从 1.43 提升为 1.55，且种群前 3 表现的因子均有不同幅度的提升，在经过第四次演化之后，新生种群的表现开始下降，种群的最优值从 1.55 下降至 1.51，在此之前的三次演化，模型表现均得到提升，从初始的 1.43 提升至 1.49，再到 1.48，直至 1.55。上述分析可以发现，虽然种群的演化可以生成表现更好的因子，但并不一定随着演化次数的增加整体的表现也随着增加。一个优秀的种群有概率会生成适应度函数更高的种群，但受到多种因素的影响，也有一定概率其后代的表现变差。

上述结果表明，铝的遗传规划模型能够在历史回测阶段找到表现较好的因子，然而对于遗传规划类模型最大的担忧便是其是否存在过拟合的现象，为了验证这一点，选取上述回测过程中表现最好的因子对其进行样本外的回测，样本外回测时间段为 2021 年 3 月至 2022 年 3 月，具体表现如下：

图表 11: 样本外表现



资料来源：东证衍生品研究院

考察该因子最近一年的表现（样本外），发现其夏普值为 0.87，该值相较于 1.55 有了明显下降，但考虑到回测时段只有一年，且 0.87 的夏普值能够表明因子具有一定的预测能力，仍然可以认为该因子有效。

考虑到初始种群因子（即经过降维后的因子）对降维方式的敏感程度较大，在下面部分，分别利用不同的降维方式对铝的基本面因子进行重新降维，降维之后的维度数量依然选择为 12。通过这样的方式来衡量遗传规划算法是否能够在不同降维模式下均对因子表现进行明显提升。

## 4.1、PCA

采用 PCA 进行降维之后的数据整体上相较于 LLE 偏低，然后对 PCA 数据内部进行分析可以发现，初始种群的最佳值为 1.52，在经过 3 次演化之后得到模型的最优输出 1.58。

图表 12: 铝的 PCA 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.22 | 1.45 | 1.43 | 0.8  | 0.83 | 1.52 | 1.15 | 1.24 | 0.94 | 1.06 | 1.51 | 1.25 | 1.52 |
| 演化次数1 | 1.43 | 1.22 | 1.28 | 0.87 | 0.91 | 1.44 | 1.43 | 0.88 | 1.19 | 0.61 | 1.27 | 1.28 | 1.44 |
| 演化次数2 | 1.15 | 0.88 | 1.27 | 1.37 | 0.93 | 1.19 | 1.31 | 1.43 | 1.24 | 0.88 | 1.28 | 1.16 | 1.43 |
| 演化次数3 | 1.28 | 1.58 | 1.24 | 0.99 | 0.88 | 1.21 | 1.43 | 1.19 | 1.37 | 1.06 | 0.88 | 0.89 | 1.58 |
| 演化次数4 | 1.12 | 1.24 | 0.88 | 1.06 | 1.37 | 1.02 | 0.88 | 1.39 | 1.37 | 1.09 | 0.82 | 1.18 | 1.39 |

资料来源：东证衍生品研究院

## 4.2、KPCA

利用 KPCA 进行降维之后（此处使用的核函数为多项式核函数），模型演化的过程并没有对种群的最佳值产生提升，模型的最佳输出始终停留在 1.38。



图表 13: 铝的 KPCA 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.05 | 1.38 | 1.12 | 0.78 | 0.94 | 0.76 | 1.03 | 1.09 | 1.36 | 1.11 | 0.85 | 0.79 | 1.38 |
| 演化次数1 | 1.36 | 1.05 | 0.88 | 1.35 | 0.88 | 1.11 | 1.21 | 1.09 | 1.29 | 1.38 | 0.88 | 0.76 | 1.38 |
| 演化次数2 | 1.36 | 0.88 | 0.88 | 1.17 | 1.38 | 1.36 | 1.03 | 1.12 | 1.2  | 0.91 | 1.03 | 1.24 | 1.38 |
| 演化次数3 | 1.17 | 1.12 | 1.16 | 1.38 | 1.36 | 0.97 | 1.36 | 1.12 | 0.88 | 1.12 | 1.05 | 1.36 | 1.38 |
| 演化次数4 | 1.36 | 1.12 | 1.16 | 1.38 | 1.08 | 1.12 | 0.88 | 1.12 | 1.31 | 1.36 | 1.12 | 1.09 | 1.38 |

资料来源: 东证衍生品研究院

### 4.3、LDA

利用 LDA 降维方式进行降维之后, 模型的演化能够有效提升因子的表现, 首先其初始种群的最优值只有 1.23, 而在其进行了两轮的演化迭代之后, 表现达到了 1.55。

图表 14: 铝的 LDA 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.04 | 1.15 | 1.22 | 0.9  | 0.87 | 0.62 | 1.12 | 0.76 | 1    | 1.2  | 1.23 | 1.05 | 1.23 |
| 演化次数1 | 1.09 | 1.48 | 1.19 | 1.06 | 0.88 | 1.05 | 0.89 | 1.09 | 1.13 | 0.89 | 1.11 | 1.19 | 1.48 |
| 演化次数2 | 1.49 | 0.94 | 1.19 | 1.12 | 1.13 | 1.03 | 0.92 | 0.88 | 1.02 | 1.09 | 1.19 | 1.55 | 1.55 |
| 演化次数3 | 1.13 | 1.49 | 1.19 | 1.12 | 1.19 | 1.13 | 1.09 | 1.12 | 1.24 | 1.19 | 1.49 | 0.88 | 1.49 |
| 演化次数4 | 1.17 | 1.13 | 1.13 | 1.24 | 1.49 | 1    | 1.02 | 1.07 | 1.15 | 1.24 | 0.91 | 1.13 | 1.49 |

资料来源: 东证衍生品研究院

### 4.4、MDS

利用 MDS 进行降维后, 初始种群的最佳输出为 1.53, 在经过 3 次演化之后, 其最佳输出达到了 1.56。

图表 15: 铝的 MDS 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.16 | 1.32 | 1.15 | 1.31 | 0.97 | 0.95 | 0.91 | 1.53 | 1.15 | 1    | 1.02 | 1.33 | 1.53 |
| 演化次数1 | 0.91 | 1.01 | 1.15 | 1.1  | 1.22 | 0.88 | 1.53 | 1.16 | 1.48 | 0.88 | 1.31 | 1.06 | 1.53 |
| 演化次数2 | 0.89 | 0.82 | 0.96 | 1.53 | 1.03 | 1.04 | 0.88 | 1.54 | 1.31 | 1.15 | 1.04 | 0.88 | 1.54 |
| 演化次数3 | 1.5  | 1.03 | 0.94 | 1.04 | 0.9  | 1.53 | 1.11 | 1.31 | 1.19 | 0.96 | 1.56 | 0.88 | 1.56 |
| 演化次数4 | 0.71 | 1.15 | 1.49 | 0.88 | 1.31 | 1.31 | 0.81 | 0.9  | 0.88 | 1.5  | 1.45 | 1.47 | 1.50 |

资料来源: 东证衍生品研究院

### 4.5、Isomap

Isomap 的降维方式延续了 MDS 的特点, 然而其效果却略显逊色, 初始种群的最佳输出



为 1.30，在经过一次演化迭代之后，最佳值达到了 1.32。

图表 16：铝的 Isomap 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.14 | 1.08 | 1.15 | 1.19 | 1.14 | 1.3  | 1.29 | 1.26 | 1.15 | 1.05 | 0.95 | 0.95 | 1.30 |
| 演化次数1 | 1.13 | 0.88 | 1.25 | 1.01 | 1.19 | 0.88 | 0.88 | 1.21 | 1    | 1.21 | 1.15 | 1.32 | 1.32 |
| 演化次数2 | 1.28 | 0.82 | 1.19 | 1.11 | 1.23 | 1.27 | 1.12 | 0.91 | 1.2  | 1.15 | 1.03 | 1.12 | 1.28 |
| 演化次数3 | 1.22 | 1.07 | 1.23 | 1.11 | 1.27 | 0.88 | 1.21 | 1.13 | 1.15 | 1.2  | 1.27 | 1.19 | 1.27 |
| 演化次数4 | 1.18 | 1.06 | 0.9  | 0.95 | 1.17 | 0.88 | 1.25 | 1.29 | 1.22 | 1.19 | 1.08 | 1.14 | 1.29 |

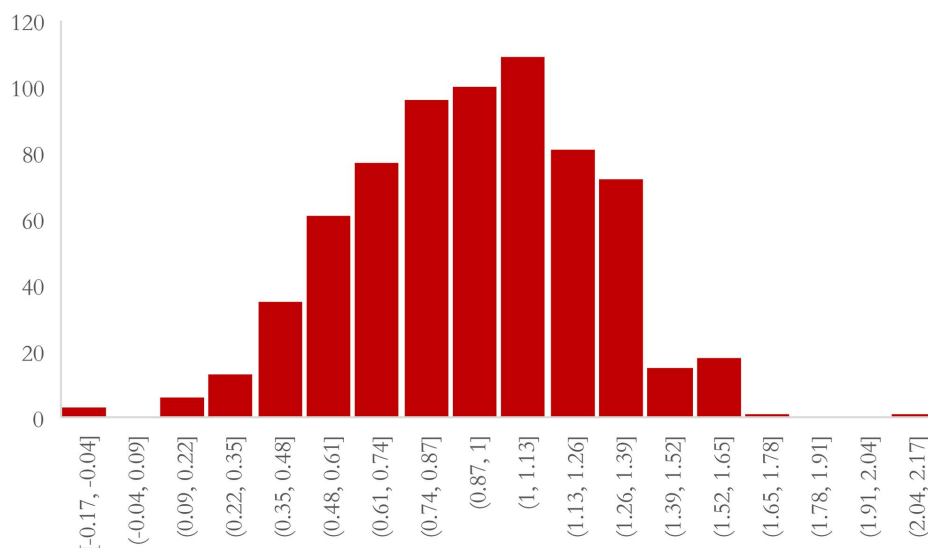
资料来源：东证衍生品研究院

在此针对铝期货的遗传规划结果进行总结性讨论：铝单因子的预测能力较为有限，多数因子的夏普值不超过 1，极个别表现较好的因子其夏普值达到 1.40，然而考虑到数据的单一性，该单因子的预测能力并不稳定。在对铝基本面因子进行 LLE 降维之后，得到的各因子的预测能力有了显著的提升，12 个因子对应的夏普值范围为[0.90,1.43]。再经过遗传规划更新迭代之后，最佳值由 1.43 上升至 1.55，然后随着进一步的迭代，后续种群的表现却没有得到提升。最后，考察基于不同降维方式的遗传规划模型效果，多数表现不及 LLE，但都通过演化迭代得到了表现更好的交易因子。

## 5、原油的单因子检验

按照之前的做法，首先对原油相关基本面因子作单因子回测。原油类基本面因子数量较多，共有 688 个，涵盖了期货行情数据，各地现货价格数据，全球原油及成品油库存，全球炼厂运行等多个维度的数据。通过直方图分布可以发现大多数因子的夏普值集中在 1.0 上下。说明对于原油的基本面单因子而言，整体具有较强的预测能力。

图表 17: 原油的单因子回测指标直方图



资料来源：东证衍生品研究院

## 6、原油的遗传规划

本部分继续针对原油进行同样的考察，观测遗传规划算法是否能够提升其回测结果。依然采用 LLE 算法对原油的基本面数据进行降维处理，将维度降至 12 维，将这 12 个经过降维后的因子作为初始种群。在初始种群中，最佳值为 1.96。

图表 18: 原油的初始种群

|      | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群 | 1.31 | 0.60 | 1.26 | 1.75 | 1.96 | 0.46 | 0.06 | 0.08 | 0.37 | 0.03 | 1.54 | 0.57 | 1.96 |

资料来源：东证衍生品研究院

接下来按照之前的步骤对原油期货的初始种群进行迭代次数为 4 次的演化，经过第一次演化之后，新生种群的最佳值达到了 2.00，表现有一定的提升，进一步演化之后，最佳值的表现升至 2.12，在经过全部 4 次演化之后，最终的表现值为 1.89。整体来说，种群演化对于降维后的因子产生了明显的提升，全部衍生种群中的最佳值为第二次演化之后的 2.12。

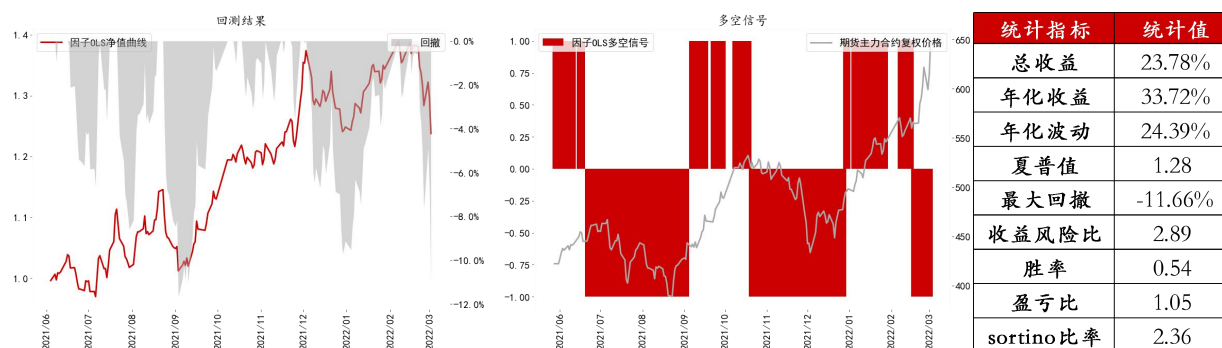
图表 19：原油的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.31 | 0.60 | 1.26 | 1.75 | 1.96 | 0.46 | 0.06 | 0.08 | 0.37 | 0.03 | 1.54 | 0.57 | 1.96 |
| 演化次数1 | 1.05 | 0.25 | 0.60 | 0.87 | 1.42 | 2.00 | 1.26 | 1.54 | 1.26 | 0.58 | 0.82 | 1.16 | 2.00 |
| 演化次数2 | 0.98 | 1.58 | 2.12 | 0.82 | 0.82 | 0.77 | 0.98 | 0.73 | 0.63 | 1.89 | 1.28 | 1.25 | 2.12 |
| 演化次数3 | 0.82 | 1.58 | 0.82 | 1.28 | 1.54 | 0.18 | 1.89 | 1.25 | 1.42 | 1.46 | 1.06 | 2.12 | 2.12 |
| 演化次数4 | 1.15 | 0.82 | 1.89 | 0.64 | 0.82 | 1.22 | 1.22 | 0.70 | 0.82 | 0.90 | 0.82 | 1.19 | 1.89 |

资料来源：东证衍生品研究院

上述结果表明，原油的遗传规划模型能够在历史回测阶段找到表现较好的因子，然而对于遗传规划类模型最大的担忧便是其是否存在过拟合的现象，为了验证这一点，选取上述回测过程中表现最好的因子对其进行样本外的回测，样本外回测时间段为2021年3月至2022年3月，具体表现如下：

图表 20：样本外表现



资料来源：东证衍生品研究院

考察该因子最近一年的表现，夏普值为 1.28，该值相较于 2.12 有明显的下降，但考虑到回测时段只有一年，且 1.28 的夏普值能够表明因子具有一定的预测能力，仍然可以认为该因子有效。

考虑到初始种群因子（即经过降维后的因子）对降维方式的敏感程度较大，在接下来的部分，利用不同的降维方式对原油的基本面因子进行重新降维，降维之后的维度数量依然选择为 12。通过这样的方式来衡量遗传规划算法是否能够在不同降维模式下均对因子表现进行明显提升。

## 6.1、PCA

在 PCA 降维模式下，遗传规划模型的效果得到更进一步的显现。PCA 降维的初始种群夏普值只有 1.60，而在经过两次演化迭代之后，其表现达到了 2.24，显著提升，然而令

人意外的是，在第二次演化之后，模型的表现连续下降，最后在经过四次演化后，其表现为 1.59。

图表 21：原油的 PCA 降维

|       | g1   | g2    | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|-------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1.34 | -0.07 | 0.91 | 1.33 | 1.6  | 0.35 | 0.76 | 0.38 | 0.76 | 1.36 | 0.54 | 0.85 | 1.60 |
| 演化次数1 | 1.39 | 0.7   | 1.47 | 1.56 | 1.39 | 1.13 | 1.33 | 1.05 | 0.76 | 0.75 | 0.66 | 0.58 | 1.56 |
| 演化次数2 | 1.39 | 1.33  | 1.39 | 2.24 | 0.9  | 0.92 | 1.39 | 1.37 | 0.83 | 0.9  | 1.56 | 1.47 | 2.24 |
| 演化次数3 | 0.58 | 0.41  | 1.39 | 0.61 | 0.99 | 0.67 | 1.59 | 1.39 | 1.37 | 1.39 | 1.77 | 0.61 | 1.77 |
| 演化次数4 | 1.39 | 1.39  | 1.19 | 1.39 | 0.99 | 1.39 | 1.59 | 1.27 | 0.54 | 1.39 | 0.66 | 1.39 | 1.59 |

资料来源：东证衍生品研究院

## 6.2、KPCA

KPCA 降维模式下，遗传规划整体表现先降后升，最终四次迭代之后，其表现为 1.27。整体而言，KPCA 降维效果不及 PCA。

图表 22：原油的 KPCA 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7    | g8   | g9   | g10   | g11  | g12  | max  |
|-------|------|------|------|------|------|------|-------|------|------|-------|------|------|------|
| 初始种群  | 0.29 | 0.51 | 0.77 | 0.28 | 0.22 | 0.26 | -0.24 | 0.1  | 1.19 | -0.12 | 0.9  | 0.61 | 1.19 |
| 演化次数1 | 0.22 | 0.27 | 0.73 | 0.73 | 0.22 | 0.66 | 0.54  | 0.78 | 0.73 | 0.51  | 0.57 | 0.66 | 0.78 |
| 演化次数2 | 0.66 | 1.01 | 0.66 | 0.66 | 0.73 | 1.25 | 0.73  | 0.66 | 0.73 | 0.73  | 0.15 | 0.73 | 1.25 |
| 演化次数3 | 0.73 | 0.73 | 0.44 | 1.08 | 0.73 | 1.2  | 0.73  | 0.73 | 0.86 | 1.17  | 0.73 | 1.01 | 1.20 |
| 演化次数4 | 0.66 | 0.73 | 0.74 | 0.5  | 1.2  | 1.2  | 0.73  | 1.01 | 1.27 | 1.17  | 0.73 | 1.08 | 1.27 |

资料来源：东证衍生品研究院

## 6.3、MDS

MDS 降维模式下，演化得到的提升并不显著，在第一次演化后，最佳输出从初始的 1.50 提升至 1.54，之后表现并没有增长。

图表 23：原油的 MDS 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1    | 0.28 | 1.29 | 0.61 | 0.75 | 0.8  | 0.8  | 0.59 | 1.5  | 0.85 | 0.97 | 0.86 | 1.50 |
| 演化次数1 | 0.86 | 1.5  | 1.05 | 0.8  | 0.67 | 0.82 | 1.29 | 0.78 | 0.97 | 0.85 | 1.54 | 0.8  | 1.54 |
| 演化次数2 | 1.1  | 1.54 | 1.54 | 0.66 | 1.05 | 0.85 | 0.59 | 0.66 | 0.84 | 0.63 | 1.54 | 0.6  | 1.54 |
| 演化次数3 | 1    | 1.54 | 0.25 | 0.77 | 1.27 | 0.66 | 0.77 | 1.54 | 0.66 | 1.54 | 1.54 | 1.54 | 1.54 |
| 演化次数4 | 1.54 | 1    | 1.54 | 1.54 | 1.54 | 1.44 | 1.54 | 1.23 | 1.54 | 1.54 | 0.66 | 1.27 | 1.54 |

资料来源：东证衍生品研究院

## 6.4、Isomap

Isomap 降维模式下，遗传规划对模型表现的提升十分明显，通过下表可以发现，几乎每经过一次迭代，种群的最优值均会得到提升，从初始种群的 1.42 一步步上升至最后的 1.78。

图表 24：原油的 Isomap 降维

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.43 | 0.92 | 1.02 | 0.58 | 1.27 | 0.96 | 0.48 | 1.06 | 0.97 | 0.88 | 1.42 | 1.32 | 1.42 |
| 演化次数1 | 0.96 | 1.29 | 1.26 | 1.19 | 1.21 | 1.29 | 0.96 | 0.88 | 1.38 | 0.96 | 1.48 | 1.16 | 1.48 |
| 演化次数2 | 1.27 | 1.33 | 1.38 | 1.36 | 1.5  | 1.32 | 0.66 | 0.9  | 1.24 | 1.26 | 0.92 | 1.67 | 1.67 |
| 演化次数3 | 1.54 | 1.04 | 0.83 | 0.38 | 1.37 | 1.38 | 1.36 | 1.27 | 0.74 | 1.4  | 1.27 | 1.67 | 1.67 |
| 演化次数4 | 1.78 | 1.27 | 1.27 | 1.37 | 1.27 | 1.67 | 1.54 | 1.27 | 1.67 | 1.36 | 1.36 | 1.54 | 1.78 |

资料来源：东证衍生品研究院

在此针对原油期货的遗传规划结果进行总结性讨论：原油单因子的预测能力较为有限，多数因子的夏普值分布在 1.0 左右，极个别表现较好的因子其夏普值达到 1.5 以上，然而考虑到数据的单一性，该单因子的预测能力并不稳定。在对原油基本面因子进行 LLE 降维之后，12 个因子对应的夏普值范围为[0.03,1.96]。在经过遗传规划更新迭代之后，最佳值由 1.96 上升至 2.12，然后随着进一步的迭代，后续种群的表现却没有得到提升。最后，考察基于不同降维方式的遗传规划模型效果，均通过演化迭代得到了表现更好的交易因子。其中 PCA 降维模式下，表现尤为明显，初始表现从 1.60 上升至 2.24。

## 7、其他品种

为了验证遗传算法在其他品种上是否有效，在这一部分，选取部分热门品种对遗传规划模型结果作进一步展示：

### 7.1、镍

镍期货的单因子预测能力并不强，初始种群的最优值为 0.67，整体最佳输出出现在第三次演化之后的 0.75。

图表 25: 镍的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.54 | 0.55 | 0.67 | 0.59 | 0.1  | 0.62 | 0.46 | 0.64 | 0.54 | 0.63 | 0.51 | 0.4  | 0.67 |
| 演化次数1 | 0.65 | 0.51 | 0.58 | 0.33 | 0.48 | 0.68 | 0.67 | 0.62 | 0.54 | 0.54 | 0.44 | 0.33 | 0.68 |
| 演化次数2 | 0.33 | 0.62 | 0.33 | 0.73 | 0.17 | 0.65 | 0.65 | 0.52 | 0.33 | 0.67 | 0.65 | 0.73 |      |
| 演化次数3 | 0.33 | 0.61 | 0.24 | 0.3  | 0.37 | 0.33 | 0.73 | 0.3  | 0.62 | 0.33 | 0.75 | 0.28 | 0.75 |
| 演化次数4 | 0.37 | 0.33 | 0.37 | 0.61 | 0.52 | 0.32 | 0.41 | 0.33 | 0.33 | 0.33 | 0.57 | 0.33 | 0.61 |

资料来源: 东证衍生品研究院

## 7.2、锡

锡期货的单因子具有一定的预测能力, 初始种群的最优值为 1.05, 整体最佳输出出现在第四次演化之后的 1.05, 针对锡而言, 遗传规划模型并未对因子表现进行提升。

图表 26: 锡的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.7  | 0.72 | 1.05 | 0.84 | 0.8  | 0.72 | 0.83 | 0.87 | 0.98 | 0.86 | 0.67 | 0.6  | 1.05 |
| 演化次数1 | 0.83 | 0.79 | 0.9  | 0.99 | 0.85 | 0.85 | 0.96 | 0.8  | 0.85 | 0.98 | 0.85 | 0.97 | 0.99 |
| 演化次数2 | 0.97 | 0.75 | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.98 | 0.9  | 0.95 | 0.95 | 0.97 | 0.98 |
| 演化次数3 | 0.75 | 0.85 | 0.85 | 0.85 | 0.97 | 0.97 | 0.85 | 0.9  | 0.97 | 0.9  | 0.98 | 0.97 | 0.98 |
| 演化次数4 | 0.84 | 0.97 | 0.84 | 1.05 | 0.85 | 0.85 | 0.92 | 0.97 | 0.97 | 1.05 | 0.82 | 0.97 | 1.05 |

资料来源: 东证衍生品研究院

## 7.3、锌

锌期货的单因子预测能力并不强, 初始种群的最优值为 0.77, 整体最佳输出出现在第一次演化之后的 0.77。

图表 27: 锌的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12   | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|-------|------|
| 初始种群  | 0.64 | 0.62 | 0.41 | 0.67 | 0.3  | 0.15 | 0.53 | 0.32 | 0.62 | 0.4  | 0.12 | -0.05 | 0.67 |
| 演化次数1 | 0.65 | 0.77 | 0.53 | 0.62 | 0.51 | 0.67 | 0.64 | 0.65 | 0.64 | 0.65 | 0.68 | 0.61  | 0.77 |
| 演化次数2 | 0.62 | 0.64 | 0.65 | 0.77 | 0.67 | 0.67 | 0.65 | 0.65 | 0.64 | 0.64 | 0.64 | 0.65  | 0.77 |
| 演化次数3 | 0.77 | 0.77 | 0.77 | 0.65 | 0.65 | 0.65 | 0.67 | 0.65 | 0.65 | 0.65 | 0.65 | 0.65  | 0.77 |
| 演化次数4 | 0.77 | 0.65 | 0.65 | 0.65 | 0.65 | 0.77 | 0.65 | 0.67 | 0.65 | 0.6  | 0.65 | 0.65  | 0.77 |

资料来源: 东证衍生品研究院

## 7.4、LPG

LPG 期货的单因子具有一定的预测能力, 初始种群的最优值为 1.39, 整体最佳输出出现在第一次演化之后的 1.69。



图表 28: LPG 的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9    | g10  | g11   | g12   | max  |
|-------|------|------|------|------|------|------|------|------|-------|------|-------|-------|------|
| 初始种群  | 0.69 | 0.61 | 0.4  | 1.39 | 0.16 | 0.31 | 0.53 | 0.18 | -0.45 | -0.1 | -0.09 | -0.16 | 1.39 |
| 演化次数1 | 0.31 | 0.36 | 0.23 | 0.23 | 0.18 | 0.41 | 0.18 | 0.4  | 1.69  | 0.23 | 0.23  | 0.61  | 1.69 |
| 演化次数2 | 0.23 | 0.4  | 0.41 | 0.41 | 0.31 | 0.08 | 0.61 | 0.27 | -0.14 | 0.61 | 0.23  | 1.06  | 1.06 |
| 演化次数3 | 0.27 | 0.23 | 0.23 | 0.41 | 0.27 | 0.23 | 0.4  | 0.23 | 0.1   | 0.4  | 1.06  | 0.23  | 1.06 |
| 演化次数4 | 0.23 | 0.22 | 0.27 | 1.06 | 1.01 | 0.41 | 0.23 | 0.23 | 0.23  | 0.4  | 0.43  | 0.43  | 1.06 |

资料来源：东证衍生品研究院

## 7.5、沥青

沥青期货的单因子预测能力并不强，初始种群的最优值为 0.48，整体最佳输出出现在第二次演化之后的 0.57。

图表 29: 沥青的遗传规划

|       | g1   | g2    | g3   | g4   | g5   | g6    | g7   | g8   | g9   | g10   | g11  | g12  | max  |
|-------|------|-------|------|------|------|-------|------|------|------|-------|------|------|------|
| 初始种群  | 0.19 | -0.13 | 0.03 | 0.48 | 0.24 | 0.32  | 0.39 | 0.37 | 0.23 | 0.16  | 0.14 | 0.19 | 0.48 |
| 演化次数1 | 0.41 | 0.09  | 0.21 | 0.35 | 0.48 | 0.24  | 0.39 | 0.34 | 0.21 | -0.13 | 0.09 | 0.27 | 0.48 |
| 演化次数2 | 0.49 | 0.21  | 0.35 | 0.26 | 0.19 | 0.44  | 0.44 | 0.56 | 0.25 | 0.09  | 0.41 | 0.57 | 0.57 |
| 演化次数3 | 0.44 | 0.43  | 0.13 | 0.53 | 0.46 | -0.13 | 0.32 | 0.57 | 0.28 | 0.1   | 0.49 | 0.41 | 0.57 |
| 演化次数4 | 0.26 | 0.26  | 0.24 | 0.15 | 0.53 | 0.12  | 0.21 | 0.21 | 0.53 | 0.44  | 0.22 | 0.42 | 0.53 |

资料来源：东证衍生品研究院

## 7.6、PTA

PTA 期货的单因子预测能力并不强，初始种群的最优值为 0.53，整体最佳输出出现在第一次演化之后的 0.81。

图表 30: PTA 的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.15 | 0.09 | 0.09 | 0.46 | 0.08 | 0.28 | 0.43 | 0.26 | 0.44 | 0.3  | 0.53 | 0.05 | 0.53 |
| 演化次数1 | 0.17 | 0.3  | 0.56 | 0.33 | 0.43 | 0.33 | 0.09 | 0.81 | 0.21 | 0.09 | 0.3  | 0.31 | 0.81 |
| 演化次数2 | 0.31 | 0.25 | 0.3  | 0.27 | 0.56 | 0.47 | 0.3  | 0.56 | 0.56 | 0.43 | 0.4  | 0.26 | 0.56 |
| 演化次数3 | 0.56 | 0.3  | 0.56 | 0.56 | 0.44 | 0.45 | 0.56 | 0.21 | 0.43 | 0.56 | 0.56 | 0.5  | 0.56 |
| 演化次数4 | 0.07 | 0.56 | 0.35 | 0.56 | 0.46 | 0.56 | 0.56 | 0.09 | 0.37 | 0.56 | 0.09 | 0.5  | 0.56 |

资料来源：东证衍生品研究院

## 7.7、PVC

PVC 期货的单因子预测能力并不强，初始种群的最优值为 0.50，整体最佳输出出现在



第三次演化之后的0.62。

图表 31: PVC 的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.5  | 0.41 | 0.38 | 0.44 | 0.38 | 0.44 | 0.38 | 0.38 | 0.44 | 0.38 | 0.43 | 0.42 | 0.50 |
| 演化次数1 | 0.44 | 0.42 | 0.41 | 0.33 | 0.44 | 0.51 | 0.5  | 0.37 | 0.42 | 0.33 | 0.46 | 0.44 | 0.51 |
| 演化次数2 | 0.53 | 0.44 | 0.43 | 0.53 | 0.44 | 0.42 | 0.44 | 0.44 | 0.5  | 0.53 | 0.47 | 0.42 | 0.53 |
| 演化次数3 | 0.33 | 0.44 | 0.51 | 0.44 | 0.51 | 0.5  | 0.53 | 0.44 | 0.33 | 0.47 | 0.47 | 0.62 | 0.62 |
| 演化次数4 | 0.53 | 0.46 | 0.47 | 0.62 | 0.47 | 0.47 | 0.53 | 0.53 | 0.6  | 0.47 | 0.47 | 0.5  | 0.62 |

资料来源：东证衍生品研究院

## 7.8、硅铁

硅铁期货的单因子预测能力并不强，初始种群的最优值为0.31，整体最佳输出出现在第三次演化之后的0.57。

图表 32: 硅铁的遗传规划

|       | g1    | g2    | g3    | g4    | g5   | g6    | g7   | g8   | g9   | g10   | g11   | g12   | max  |
|-------|-------|-------|-------|-------|------|-------|------|------|------|-------|-------|-------|------|
| 初始种群  | -0.04 | -0.01 | -0.01 | 0.16  | 0.22 | 0.31  | 0.02 | 0    | 0.23 | -0.01 | 0.12  | 0.1   | 0.31 |
| 演化次数1 | 0.16  | 0.23  | 0.16  | -0.01 | 0.49 | -0.01 | 0.02 | 0.28 | 0.09 | 0.05  | 0.29  | 0.22  | 0.49 |
| 演化次数2 | -0.01 | 0.19  | 0.16  | 0.49  | 0.23 | 0.49  | 0.23 | 0.2  | 0.49 | 0.16  | -0.01 | 0.28  | 0.49 |
| 演化次数3 | -0.01 | 0.23  | 0.23  | 0.25  | 0.31 | 0.23  | 0.57 | 0.23 | 0.23 | 0.49  | -0.01 | -0.01 | 0.57 |
| 演化次数4 | 0.49  | 0.23  | 0.57  | 0.05  | 0.23 | 0.18  | 0.23 | 0.23 | 0.25 | 0.25  | 0.23  | 0.25  | 0.57 |

资料来源：东证衍生品研究院

## 7.9、豆粕

豆粕期货的单因子预测能力并不强，初始种群的最优值为0.36，整体最佳输出出现在第三次演化之后的0.69。

图表 33: 豆粕的遗传规划

|       | g1    | g2    | g3    | g4   | g5    | g6    | g7    | g8   | g9    | g10  | g11   | g12   | max  |
|-------|-------|-------|-------|------|-------|-------|-------|------|-------|------|-------|-------|------|
| 初始种群  | -0.15 | 0.03  | -0.23 | 0.03 | -0.14 | -0.16 | 0.08  | 0.14 | 0.31  | 0.36 | 0.27  | -0.14 | 0.36 |
| 演化次数1 | 0.03  | 0.07  | 0.14  | 0.03 | 0.15  | 0.24  | 0.03  | 0.24 | -0.02 | 0.06 | 0.31  | -0.09 | 0.31 |
| 演化次数2 | 0.24  | -0.09 | 0.24  | 0.24 | 0.58  | -0.09 | 0.07  | 0.14 | 0.24  | 0.24 | 0.24  | 0.06  | 0.58 |
| 演化次数3 | 0.04  | 0.16  | 0.13  | 0.53 | 0.24  | 0.69  | -0.09 | 0.24 | 0.24  | 0.24 | -0.01 | 0.52  | 0.69 |
| 演化次数4 | 0.27  | 0.47  | 0.21  | 0.53 | 0.35  | 0.24  | 0.24  | 0.52 | 0.24  | 0.24 | 0.53  | 0.53  | 0.53 |

资料来源：东证衍生品研究院

## 7.10、豆油

豆油期货的单因子预测能力并不强，初始种群的最优值为 0.54，整体最佳输出出现在第二次演化之后的 0.64。

图表 34：豆油的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 0.54 | 0.44 | 0.44 | 0.44 | 0.48 | 0.35 | 0.3  | 0.27 | 0.03 | 0.35 | 0.49 | 0.41 | 0.54 |
| 演化次数1 | 0.5  | 0.55 | 0.32 | 0.44 | 0.49 | 0.02 | 0.58 | 0.46 | 0.39 | 0.39 | 0.55 | 0.45 | 0.58 |
| 演化次数2 | 0.64 | 0.55 | 0.48 | 0.45 | 0.54 | 0.36 | 0.32 | 0.52 | 0.55 | 0.53 | 0.36 | 0.45 | 0.64 |
| 演化次数3 | 0.55 | 0.55 | 0.55 | 0.55 | 0.49 | 0.55 | 0.52 | 0.45 | 0.55 | 0.55 | 0.39 | 0.55 | 0.55 |
| 演化次数4 | 0.55 | 0.55 | 0.54 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 |

资料来源：东证衍生品研究院

## 7.11、棕榈油

棕榈油期货的单因子具有一定的预测能力，初始种群的最优值为 1.35，整体最佳输出出现在第三次演化之后的 1.42。

图表 35：棕榈油的遗传规划

|       | g1   | g2   | g3   | g4   | g5   | g6   | g7   | g8   | g9   | g10  | g11  | g12  | max  |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 初始种群  | 1    | 0.91 | 0.94 | 1.05 | 1.2  | 1.04 | 1.17 | 1.19 | 1.23 | 0.91 | 1.35 | 1.04 | 1.35 |
| 演化次数1 | 1.36 | 0.97 | 0.3  | 1.13 | 1.18 | 1.06 | 0.97 | 0.8  | 1.09 | 1.14 | 0.85 | 1.11 | 1.36 |
| 演化次数2 | 0.8  | 1.01 | 1.13 | 0.97 | 0.97 | 1.36 | 1.06 | 0.96 | 1.18 | 0.99 | 0.75 | 1.29 | 1.36 |
| 演化次数3 | 1.42 | 1.16 | 0.74 | 1.42 | 1.01 | 1.13 | 1.13 | 1.16 | 0.94 | 0.97 | 0.94 | 1.13 | 1.42 |
| 演化次数4 | 1.13 | 1    | 1.42 | 1.42 | 1    | 1.31 | 1.03 | 1.13 | 0.98 | 0.97 | 1.42 | 0.98 | 1.42 |

资料来源：东证衍生品研究院

## 8、结论

相较于现有的遗传规划在 CTA 策略中的应用，本报告尝试使用降维后的因子作为遗传算法的初始种群，以期获得稳定性更强的期货交易因子。在对铝的研究过程中，可以发现经过降维后的因子在不断迭代演化过程中表现相较于初始种群存在一定的提升，但并不是随着迭代次数的增加而不断提升的，说明模型的适应度函数在经过一定次数的演化之后即达到了最佳值，而随着演化次数的增加，因子的复杂程度也在不断增加，与之一同增加的是因子的不可解释性，随之而来的，则是因子的交易逻辑逐渐失效，也是基于以上事实，在本报告中将最长迭代次数设置为 4 次，以此来限制因子的复杂度。

之后继续对原油期货进行相同步骤的研究，可以发现针对原油期货，整体模型的最佳

输出是出现在第二次演化之后，在此之后，随着模型的不断演化，模型的效果逐渐下降。

除此之外，报告针对铝和原油均作了不同降维方式的尝试，一个统一的结论是：经过演化之后的因子表现均好于初始种群，区别在于最佳值出现的时间略有不同，有的经过一次演化便达到峰值，有的则经过所有的四次演化才达到最佳值。

最后，报告罗列其余多个品种的实证数据，结果显示，对于大部分品种而言，该遗传规划模型均可以提升初始种群的表现，无论其初始种群本身的预测能力强弱。当然，针对那些初始种群预测能力较弱的期货品种，不排除演化后的因子由于随机波动的原因而产生效果较好的种群的可能，这一问题有待后续研究。此外，对于个别品种（如锡），其初始种群自身已经具有较强的预测能力，在经过四次演化之后，最终因子的表现并没有得到提升。故在运用此类模型的过程中，需要注意模型的适用性。

### 期货走势评级体系（以收盘价的变动幅度为判断标准）

| 走势评级 | 短期（1-3个月） | 中期（3-6个月） | 长期（6-12个月） |
|------|-----------|-----------|------------|
| 强烈看涨 | 上涨 15%以上  | 上涨 15%以上  | 上涨 15%以上   |
| 看涨   | 上涨 5-15%  | 上涨 5-15%  | 上涨 5-15%   |
| 震荡   | 振幅-5%-+5% | 振幅-5%-+5% | 振幅-5%-+5%  |
| 看跌   | 下跌 5-15%  | 下跌 5-15%  | 下跌 5-15%   |
| 强烈看跌 | 下跌 15%以上  | 下跌 15%以上  | 下跌 15%以上   |

### 上海东证期货有限公司

上海东证期货有限公司成立于2008年，是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司，注册资本金23亿元人民币，员工近600人。公司主要从事商品期货经纪、金融期货经纪、期货投资咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所和上海国际能源交易中心会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

东证期货以上海为总部所在地，在大连、长沙、北京、上海、郑州、太原、常州、广州、青岛、宁波、深圳、杭州、西安、厦门、成都、东营、天津、哈尔滨、南宁、重庆、苏州、南通、泉州、汕头、沈阳、无锡、济南等地共设有33家营业部，并在北京、上海、广州、深圳多个经济发达地区拥有134个证券IB分支机构，未来东证期货将形成立足上海、辐射全国的经营网络。

自2008年成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持市场化、国际化、集团化的发展道路，打造以衍生品风险管理为核心，具有研究和技术两大核心竞争力，为客户提供综合财富管理平台的一流衍生品服务商。

## 分析师承诺

王冬黎、谢怡伦

本人具有中国期货业协会授予的期货执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

## 免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

## 东证衍生品研究院

地址：上海市中山南路318号东方国际金融广场2号楼21楼

联系人：梁爽

电话：8621-63325888-1592

传真：8621-33315862

网址：[www.orientfutures.com](http://www.orientfutures.com)

Email：[research@orientfutures.com](mailto:research@orientfutures.com)