

## 多因子指数增强策略及优化

**摘要：**本报告旨在研究和优化多因子指数增强策略在量化投资中的应用。研究选取 HS300 指数成分股作为股票池，通过构建和优化多因子指数增强策略，最终实现了在控制跟踪误差的同时获取超额收益的目标。分层抽样策略和多因子线性优化模型策略均在回测中显示出超越基准指数的潜力，但最大回撤较高。通过对分层抽样策略的优化，本文成功降低了最大回撤风险，提升了策略的风险调整后收益。本报告的研究表明，通过精细化的因子选择和风险管理，可以有效提升指数增强策略的风险调整后收益。

**关键词：**量化投资；多因子策略；指数增强；风险因子

# 目录

一、引言 .....	3
1.1 研究背景 .....	3
1.2 指数增强策略思路 .....	3
1.3 主要内容 .....	3
二、数据处理 .....	2
2.1 数据筛选 .....	2
2.1.1 回测区间与调仓规则 .....	2
2.1.2 股票池筛选 .....	2
2.1.3 因子构造 .....	2
2.2 数据预处理 .....	3
2.2.1 缺失值处理 .....	3
2.2.2 去极值 .....	4
2.2.3 标准化 .....	4
2.2.4 市值和行业中性化 .....	4
2.3 因子多重共线性处理 .....	4
2.4 打分法计算预期收益 .....	6
三、指数增强策略 .....	7
3.1 分层抽样指数增强策略实现 .....	7
3.1.1 策略逻辑 .....	7
3.1.2 策略研究思路 .....	7
3.1.3 细节处理 .....	7
3.1.4 策略展示 .....	8
3.2 多因子线性优化模型指数增强策略实现 .....	9
3.2.1 策略逻辑 .....	9
3.2.2 策略研究思路 .....	9
3.2.3 细节处理 .....	10
3.2.4 策略展示 .....	10
四、优化策略 .....	12
4.1 策略概述 .....	12
4.2 策略步骤 .....	12
4.2.1 因子筛选 .....	12
4.2.2 策略实现 .....	14
4.2.3 策略展示 .....	15
五、总结 .....	17
5.1 策略总结 .....	17
5.2 心得体会 .....	17
参考文献 .....	18
数据及代码说明 .....	18

# 一、引言

## 1.1 研究背景

在现代金融市场中，指数增强策略因其能够在跟踪基准指数的同时获取超额收益而受到广泛关注。本文的研究背景在于，通过量化投资方法，探索 and 实现一种能够在控制跟踪误差的前提下，获取超越市场平均水平收益的投资策略。这种策略不仅要求与基准指数保持高度相关性，而且还要通过主动管理实现超额收益。

## 1.2 指数增强策略思路

指数增强策略旨在跟踪特定指数的同时，通过量化和多因子选股超越指数表现。以沪深 300 指数为例，策略需与指数走势保持高相关性，并力求超越其涨幅。华泰证券研报指出，指数增强策略可分为仓位控制、行业轮动和选股三个层面，实践中可区分为主动和量化两种方式，宏观层面差异小，微观层面差异大，主流公募产品以量化为主。

1. 仓位控制：根据市场状况和风险评估调整股票仓位，以在市场趋势明确时增加仓位获取收益，在风险较高时减少仓位控制风险。
2. 行业轮动：利用不同行业在不同经济周期的表现差异提高组合收益，通过量化模型评估行业基本面、估值和动量，适时调整行业配置，控制行业偏离度以保持跟踪误差合理。
3. 选股：使用多因子模型筛选优质股票组合，关键因子包括估值、盈利、成长、动量和波动率等。选股模型包括多因子、SmartBeta、基本面量化和事件驱动模型，选股范围可以是基准指数成分股、成分股及指定股票，或全部 A 股。

指数增强策略的目标是在对基准指数跟踪的同时获得超额收益，即在控制跟踪误差的前提下获取增强收益，一般认为指数增强型基金的年化跟踪误差需控制在 7.75% 以内。跟踪误差是跟踪偏离度的标准差，其公式如下：

$$TD_{ti} = R_{ti} - R_{tb}$$
$$TE = \frac{1}{n-1} \sum_{t=1}^n \left[ TD_{ti} - \frac{1}{n} \left( \sum_{t=1}^n TD_{ti} \right) \right]^2$$

其中  $TD_{ti}$  为基金跟踪偏离度，是基金净值变化率 ( $R_{ti}$ ) 与基金的业绩基准同期收益率 ( $R_{tb}$ ) 之差。

本文就是基于以上思想，站在微观视角，采用的量化多因子指数增强策略，相比主动方法，该策略的优势在于能科学控制跟踪误差，减少波动。此外，不论采用以上哪种操作思路，投资管理人都可以考虑配合衍生金融工具或一些其它方式进行增强，如打新或借助股指期货、融资融券、期权、可转债等工具。

## 1.3 主要内容

本文介绍了两种指数增强策略：简单分层抽样策略和复杂多因子线性优化模型。前者基于市值和行业风险因子，选择预期收益最高的股票以匹配基准指数；后者通过优化控制更多风险因子，减少跟踪误差。历史分析显示两种策略年化收益率可观，但最大回撤高。对分层抽样策略优化后，通过资产配置和风险管理调整，以及因子筛选，将最大回撤降至 29.88%，提升了风险调整后收益。

所以整体而言，本文不仅展示了多因子指数增强策略的构建和优化过程，还体现了在量化投资领域中，如何通过合理的方法来控制风险和提高收益。

## 二、数据处理

### 2.1 数据筛选

#### 2.1.1 回测区间与调仓规则

本研究选取的回测区间为 **2013 年 01 月 01 日至 2023 年 09 月 30 日**，旨在评估所选因子在该时间段内的表现。调仓日设定为每月最后一个交易日，以该日的收盘价作为买入和卖出的依据。这样的设置旨在模拟实际交易中的操作，并减少交易成本的影响。

#### 2.1.2 股票池筛选

股票池限定为 HS300 指数成分股，沪深 300 指数是 A 股市场中经典的规模指数，刻画了全市场核心龙头上市公司的整体表现，代表了 A 股市场中“核心资产”的整体走势，也是国内外各类资金重点关注的对象。

在选股过程中，本报告遵循以下筛选规则：**1.剔除 ST 股票**：ST 股票通常代表公司经营状况异常或存在退市风险，因此被排除在外，以降低投资风险。**2.剔除上市不满 1 年的股票**：新上市股票由于缺乏足够的历史数据和市场表现，可能影响因子的有效性，因此也被排除。**3.剔除无法买入的股票**：包括因停牌等原因无法交易的股票，确保所选股票在调仓日可进行买卖操作。

#### 2.1.3 因子构造

在因子构造方面，研究面临一些挑战。首先，一致性预期和分红数据在聚宽平台中不可用，因此模型中也相应缺失这些因子。其次，部分因子无法直接从聚宽因子库中获取，需要手动构造。这些因子在模型所需表中的因子计算方式项中标记有“Factor”，表示它们是通过 Factor 类构造的。此外，部分因子在原始模型基础上进行了修改，以适应本研究的特定需求，在因子计算方式项中标记为“修改”。

通过这样的数据筛选和因子构造，本报告旨在构建一个既符合实际交易条件，又能准确反映股票特性的模型，为后续的回测分析提供坚实的数据基础。所需因子清单如下：

类别	因子名称	因子 code	因子计算方式
规模	对数市值	natural_log_of_market_cap	总市值取对数
估值	账面市值比	book_to_price_ratio	净资产/总市值
估值	市盈率倒数 TTM	EPTTM	归母净利润 TTM/总市值 (Factor)
估值	市销率倒数 TTM	SPTTM	营业收入 TTM/总市值 (Factor)
技术指标	一个月反转	ROC20	过去 20 个交易日涨跌幅
技术指标	三个月反转	ROC60	过去 60 个交易日涨跌幅

成长	净利润增长率	net_profit_growth_rate	(今年净利润 (TTM) / 去年净利润 (TTM)) - 1 (修改)
成长	营业收入增长率	operating_revenue_growth_rate	今年营业收入 (TTM) / 去年营业收入 (TTM) - 1 (修改)
成长	利润总额增长率	total_profit_growth_rate	(今年利润总额 (TTM) / 去年利润总额 (TTM)) - 1 (修改)
成长	标准化预期外盈利	SUE	(单季度实际净利润 - 预期净利润) / 预期净利润标准差 (Factor)
成长	标准化预期外收入	SUR	(单季度实际营业收入 - 预期营业收入) / 预期营业收入标准差 (Factor)
盈利	单季度净资产收益率	roe_ttm	单季度归母净利润 2 / (期初归母净资产 + 期末归母净资产) (修改)
盈利	单季度总资产收益率	roa_ttm	单季度息税前利润 2 / (期初总资产 + 期末总资产) (修改)
盈利	单季度净资产收益率同比变化	DELTAROE	单季度净资产收益率 - 去年同期单季度净资产收益率 (Factor)
盈利	单季度总资产收益率同比变化	DELTAROA	单季度总资产收益率 - 去年同期单季度总资产收益率 (Factor)
流动性	非流动性冲击	ILLIQ	过去 20 个交易日涨幅绝对值 / 成交额均值 (Factor)
流动性	一个月日均换手	VOL20	过去 20 个交易日换手率均值
流动性	三个月日均换手	VOL60	过去 60 个交易日换手率均值
波动	特异度	IVR	过去 20 个交易日 Fama-French 三因子回归拟合度 (Factor)
波动	一个月真实波动率	ATR1M	过去 20 个交易日日内真实波幅均值 (Factor)
波动	三个月真实波动率	ATR3M	过去 60 个交易日日内真实波幅均值 (Factor)

## 2.2 数据预处理

### 2.2.1 缺失值处理

在数据预处理阶段，本报告首先处理因子值的缺失问题。对于因子值缺失的股票，本报告采取两种策略：一是用同行业股票的中位数填充，以保持行业内的一致性；二是如果整个行业都缺失，则直接删除这些数据。这种处理方式旨在减少缺失值对模型的影响，同时保持数据的完整性和代表性。

## 2.2.2 去极值

为了减少异常值对模型的影响，本报告采用 MAD 方法去极值，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。具体步骤如下：

1. 计算当期所有股票在因子 $f$ 上的中位数 $m_f$ 。
2. 计算绝对中位数偏差  $MAD = median(|f - m_f|)$ 。
3. 保留在 $[m_f - 3 \times 1.483 \times MAD, m_f + 3 \times 1.483 \times MAD]$ 范围内的股票因子值不变。
4. 对于大于 $m_f + 3 \times 1.483 \times MAD$ 的所有股票的因子取值，按排序均匀压缩到 $[m_f + 3 \times 1.483 \times MAD, m_f + 3.5 \times 1.483 \times MAD]$ 之间。
5. 对于小于 $m_f - 3 \times 1.483 \times MAD$ 的所有股票的因子取值，按排序均匀压缩到 $[m_f - 3.5 \times 1.483 \times MAD, m_f - 3 \times 1.483 \times MAD]$ 之间。

这种方法不仅去除了极值，而且在极值股票之间保持了原有的排序关系。

## 2.2.3 标准化

进行 MAD 去极值后，为了使不同因子在构造复合因子时具有统一的量纲，本报告对每个因子进行 Z-Score 标准化处理，使因子的均值为 0，标准差为 1，即

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

## 2.2.4 市值和行业中性化

由于因子可能受到市值和行业的影响较大，为尽可能提出因子间相关性，本报告对市值和行业进行中性化处理，即对下式做回归取残差：

$$f_i = \beta^{MV} MV_i + \sum \beta_j^{ind} X_{ij} + \epsilon$$

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 $\epsilon$ 继续做去极值、标准化处理得到中性化后的因子取值。

通过上述步骤，本报告能够有效地减少因子受到市值和行业的影响，从而得到更加纯净的因子信号。

## 2.3 因子多重共线性处理

在构建多因子选股模型时，本报告通常根据多个因子的线性加权来为个股进行综合打分，即以下形式：

$$F = v^1 \cdot f^1 + v^2 \cdot f^2 + \dots + v^K \cdot f^K$$

其中  $K$  表示因子的数量， $f^K$  表示股票在第  $K$  个因子上的取值， $v^K$  表示因子的权重。这种打分方式有一个很重要的隐含假设是因子之间的相关性较低。因此，对于选股模型中因子的共线性问题，本报告使用对称正交来处理共线性的问题。

对称正交有几个重要的性质：

- (1) 相对于施密特正交法（逐步回归取残差的方式），对称正交无需提供正交次序，对每个因子平等看待；
- (2) 在所有正交过渡矩阵中，对称正交后的矩阵和原始矩阵的相似性最大，即正交前后矩阵的距离最小，对原始因子矩阵的修改最小；
- (3) 对称正交的计算只需要截面因子数据，计算效率非常高。

从这些性质出发，对称正交后的因子和原始因子有较好的对应关系，因子的经济意义保持能力较好，并且在 Frobenius 范数下保持了最高的相似性。为了方便直观理解，本报告

以两个因子 1,2 分别进行施密特正交和对称正交的示意图：

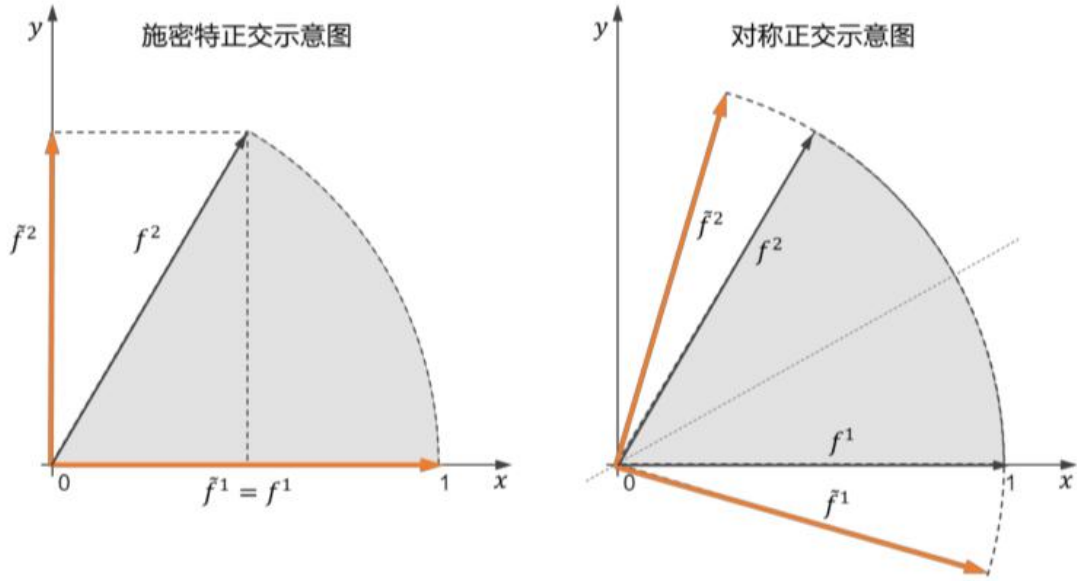


图 1. 施密特正交与对称正交的示意图

从图中可以看出，施密特正交中排序越高的因子旋转角度越大，对因子并没有平等对待；而对称正交中因子都各自旋转了同样很小的角度来得到正交基，对因子平等对待，且正交前后因子的对应关系保持的很好。

具体的过渡矩阵计算过程如下：本报告定义一个从 $F_{N \times K}$ 旋转到 $\widetilde{F}_{N \times K}$ 的过渡矩阵 $S_{K \times K}$ ，公式如下所示：

$$S_{K \times K} = U_{K \times K} D_{K \times K}^{-1/2} U_{K \times K}'$$

其中，记矩阵 $M = F_{N \times K}' F_{N \times K}$ ， $D_{K \times K}$ 为 M 的特征根构成的对角阵， $U_{K \times K}$ 为每一列由 M 的特征向量构成的矩阵。 $S_{K \times K}$ 是一个对角矩阵，所以本报告才称之为对称正交， $F_{N \times K}$ 的旋转公式如下：

$$\widetilde{F}_{N \times K} = F_{N \times K} S_{K \times K}$$

未进行正交处理,各因子的相关性如下： 而进行正交处理后，各因子相关性明显下降：

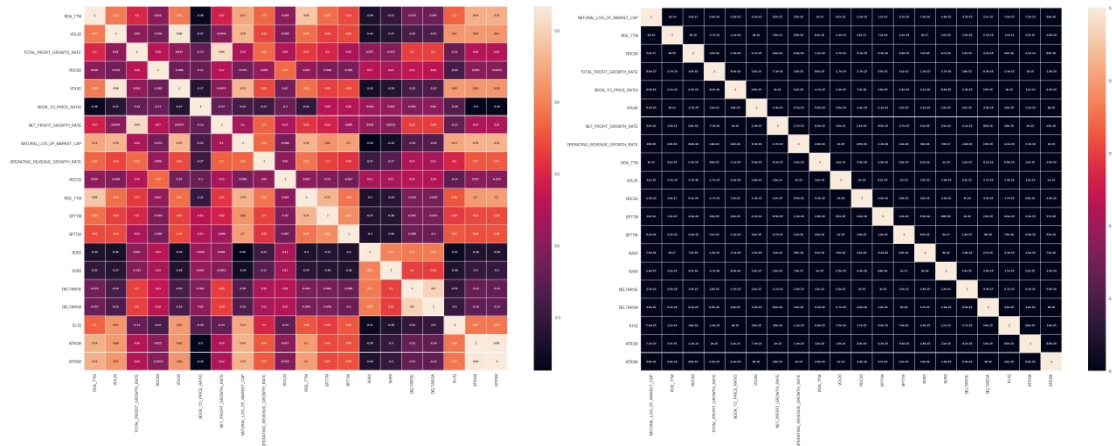


图 2. 未正交处理的各因子相关性热力图 图 3. 正交处理后的各因子相关性热力图

## 2.4 打分法计算预期收益

本报告以 ICIR 作为因子的权重。在利用 IR 加权法计算权重时，各因子权重值为各因子在滚动时间内平均 IR 值在所有因子在滚动时间内平均 IR 值之和中的占比，因子在滚动时间内平均 IR 值越大，说明该因子在选股时作用越大，权重也越大，反之亦然。

在利用 IR 计算权重时需注意的是：在滚动计算 IC 序列标准差时，起始日期得到的是缺失值，所以对于起始日期的 IR 值本报告用该时点的 IC 值来代替，得到的权重也为 IC 值计算的 IC 权重。



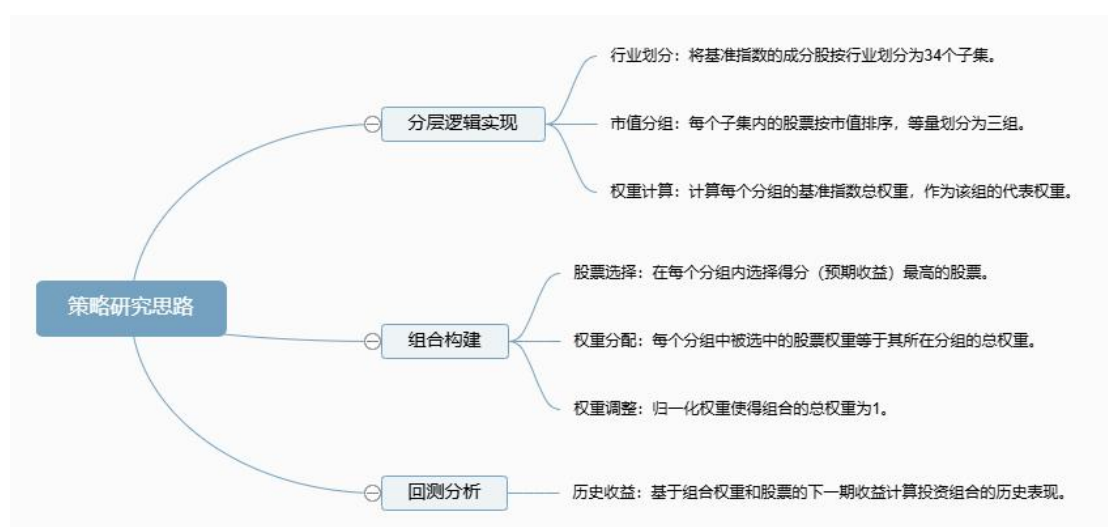
## 三、指数增强策略

### 3.1 分层抽样指数增强策略实现

#### 3.1.1 策略逻辑

分层抽样策略通过分层处理风险因子，构建与基准指数风险暴露一致的组合，并在每层中选优股票以获取超额收益。以申万 34 个一级行业为基准，将股票分至子集，再按市值分为三组，确保市值分布一致。在每组中，根据预期收益得分选股票，权重与基准指数中该组权重相同，归一化后构建组合。

#### 3.1.2 策略研究思路



#### 3.1.3 细节处理

##### 1. 在每个子集中用市值因子将股票划分为数目相等的三组

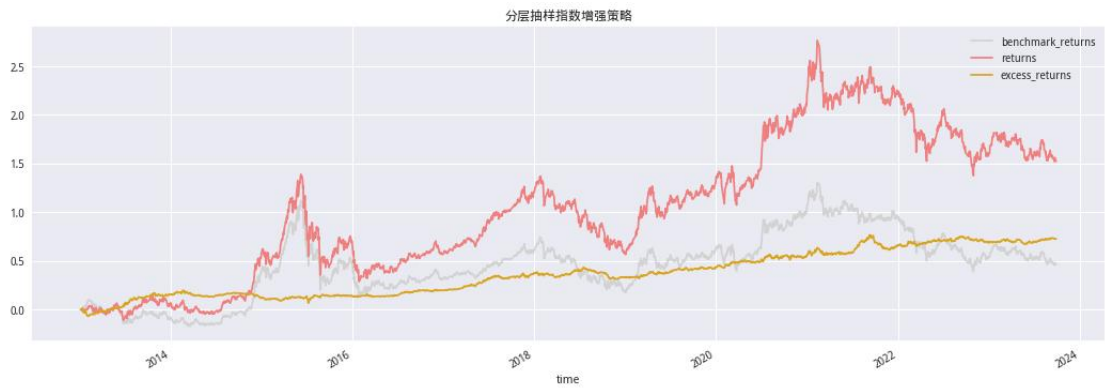
市值因子是量化投资中常用的因子之一，它基于“规模效应”理论，即市值较小的股票往往能够获得较高的收益率。通过市值分层，可以确保投资组合在不同市值区间的公司之间实现风险分散，避免对单一市值区间的过度依赖，策略也可以更有效地捕捉小市值股票的潜在超额收益，允许策略根据不同市值区间的市场表现进行调整，从而更好地适应市场环境的变化。同时，市值分层有助于实现市值中性，即投资组合的市值分布与基准指数相似，从而控制市值风险。

##### 2. 通过打分法选择股票

通过选择每个小组内得分最高的股票，策略能够集中投资于预期收益最高的个股，从而提高整个投资组合的效率。相对于在每个小组内选择多只股票，只选择得分最高的股票可以减少交易次数，从而降低交易成本和潜在的冲击成本。通过选择每个小组内得分最高的股票，策略的决策过程更加直观和易于理解，便于投资者评估和信任策略。打分法相对稳健，不容易受到极端值的影响，有助于在不同市场环境下保持策略的稳定性。同时，这种策略允许在不同的市场条件下快速调整投资组合，因为只需重新评估每个小组内的股票得分并选择新的领头羊。

3.1.4 策略展示

分层抽样指数增强策略回测图如下：



分层抽样指数增强策略收益情况如下：

指标名称	数值
策略收益	152.16%
策略年化收益	9.26%
超额收益	72.43%
基准收益	46.24%
阿尔法	0.056
贝塔	1.032
夏普比率	0.223
胜率	0.549
盈亏比	1.423
最大回撤率	46.10%
索提诺比率	0.284
信息比率	0.912
盈利次数	3078
亏损次数	2529
日胜率	0.530
日均超额收益	0.02%
超额收益最大回撤	10.94%
超额收益夏普比率	0.223
策略波动率	0.236
基准波动率	0.221
最大回撤区间	2015/06/08, 2016/01/28

策略在考察期内实现了 152.16%的总收益，显著超越了基准收益的 46.24%，显示出策略的卓越表现。年化收益达到 9.26%，这是一个相对健康的回报率，但考虑到 46.10%的最大回撤，投资者需要具备较高的风险承受能力。阿尔法值为 0.056，表明策略在调整风险后，相对于市场基准有正的超额回报。贝塔值为 1.032，意味着策略的波动性略高于市场，但风

险并未过度放大。夏普比率为 0.223，虽然低于 1，但在考虑市场无风险利率和市场情况后，仍可接受。胜率为 0.549，盈亏比为 1.423，这表明策略在大多数交易中是盈利的，并且盈利交易的收益能够覆盖亏损。日均超额收益为 0.02%，日胜率为 0.530，这表明策略追求的是稳定而非高波动的收益。信息比率为 0.912，显示策略在控制跟踪误差的同时，实现了较高的超额收益。策略波动率为 0.236，与基准波动率 0.221 相近，表明策略的波动性与市场基准相似。最大回撤区间为 2015/06/08 至 2016/01/28，投资者应考虑这个时间段内市场的整体表现以及策略的特定风险因素。

综合来看，这个策略在选定时间段内表现出色，具有较高的超额收益和合理的风险调整后收益，但同时也伴随着较大的回撤风险，需要投资者有相应的风险管理措施来应对可能的市场波动。

## 3.2 多因子线性优化模型指数增强策略实现

### 3.2.1 策略逻辑

多因子线性优化模型指数增强策略旨在使组合在多风险因子上的暴露与基准指数相匹配，减少跟踪误差。风险控制包括风险、行业暴露、跟踪误差和个股权重约束等。策略通常通过解二次或非线性优化问题来构建最优组合。

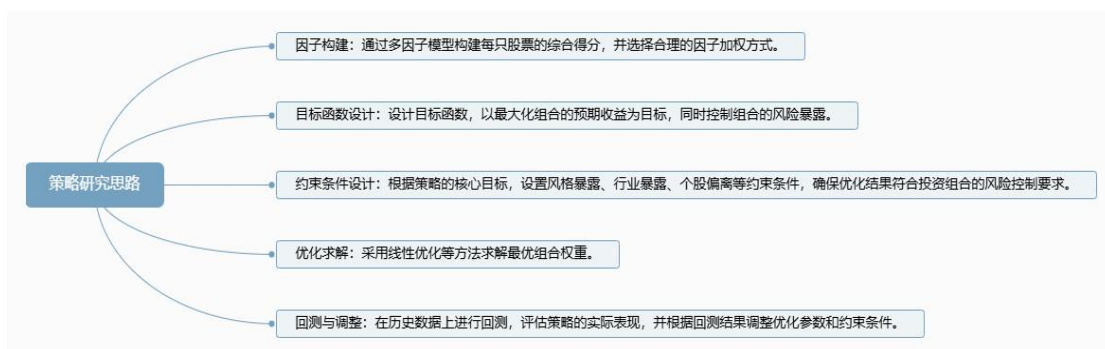
策略的组合优化模型形式如下所示：

$$\begin{aligned}
 \text{目标函数:} \quad & \max \mathbf{r}^T \mathbf{w} \\
 \text{约束条件:} \quad & \mathbf{s}_l \leq \mathbf{X}(\mathbf{w} - \mathbf{w}_b) \leq \mathbf{s}_h \\
 & \mathbf{h}_l \leq \mathbf{H}(\mathbf{w} - \mathbf{w}_b) \leq \mathbf{h}_h \\
 & \mathbf{w}_l \leq \mathbf{w} - \mathbf{w}_b \leq \mathbf{w}_h \\
 & \mathbf{b}_l \leq \mathbf{B}_b \mathbf{w} \leq \mathbf{b}_h \\
 & 0 \leq \mathbf{w} \leq \mathbf{l} \\
 & \mathbf{I}^T \mathbf{w} = 1
 \end{aligned}$$

该优化问题的目标函数为最大化组合收益，其中  $\mathbf{r}^T \mathbf{w}$  为组合预期收益， $\mathbf{w}$  为待求解的股票权重向量， $\mathbf{r}$  为前面计算得到的每股综合得分。模型约束条件包括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，具体解释如下：

- 约束条件 1 控制了组合相对于基准指数的风格暴露， $\mathbf{X}$  为股票对风格因子的因子暴露矩阵， $\mathbf{w}_b$  为基准指数成分股的权重向量， $\mathbf{s}_l$  和  $\mathbf{s}_h$  分别为风格因子相对暴露的下限和上限；
- 约束条件 2 限制了组合相对于基准指数的行业偏离， $\mathbf{H}$  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， $\mathbf{h}_l$  和  $\mathbf{h}_h$  分别为组合行业偏离的下限和上限，此处行业采用前文的 32 个行业分类方法；
- 约束条件 3 限制了个股相对于基准指数成分股的偏离， $\mathbf{w}_l$ ， $\mathbf{w}_h$  分别为个股偏离的下限和上限；
- 约束条件 4 限制了组合在成分股内权重的占比下限及上限， $\mathbf{B}_b$  为个股是否属于基准指数成分股的 0-1 向量， $\mathbf{b}_l$ ， $\mathbf{b}_h$  分别为成分股内权重占比的下限及上限；
- 约束条件 5 限制了卖空，并且限制了个股权限的上限 1（即满仓）；
- 约束条件 6 要求权重和为 1，即组合始终满仓运作。

### 3.2.2 策略研究思路



### 3.2.3 细节处理

#### 1. 用个股相对于基准指数成分股的偏离度来控制跟踪误差

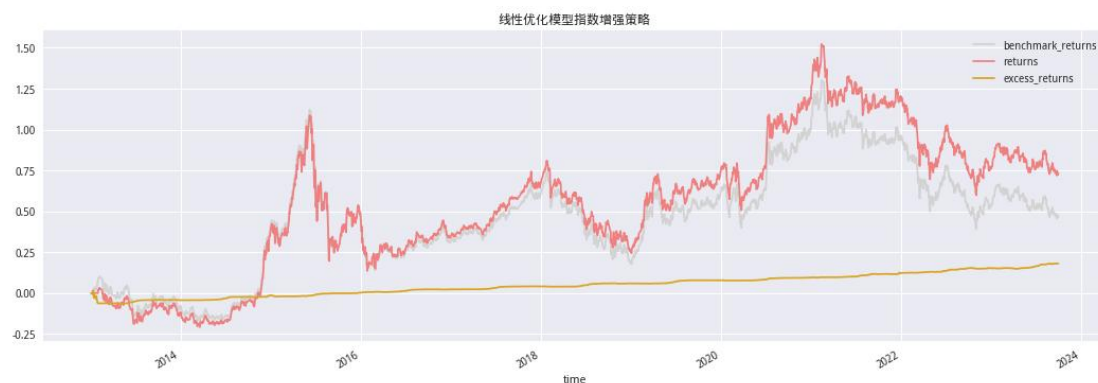
该组合优化模型没有采用二次项的跟踪误差约束  $(\mathbf{w} - \mathbf{w}_b)^T \Sigma (\mathbf{w} - \mathbf{w}_b) \leq TE^2$  为组合的预期年化跟踪误差的上限来控制组合对基准的偏离，而是用的约束条件 3 中的个股相对于基准指数成分股的偏离度，原因：①直接用跟踪误差作为约束条件进行风险控制需要估计协方差矩阵，跟踪误差是否控制成功依赖于协方差矩阵估计的准确性，而协方差矩阵的估计本身受很多假设条件的限制，估计较为困难，不稳定性较大，而用个股相对于基准指数成分股的偏离度来控制跟踪误差更为直接，个股偏离度越小，对基准指数的跟踪误差也就越小；②跟踪误差约束是二次约束，而偏离度约束是线性约束，后者比前者的求解更高效。

#### 2. 约束条件的细节

- 风格暴露：确保组合的风格因子暴露与基准指数一致。通过计算风格因子矩阵 (f) 并对组合的权重调整，控制偏离度在 [0.95, 1.05]。
- 行业暴露：对行业分类进行 one-hot 编码，限制行业暴露的权重偏离范围，确保组合与基准在行业上的分布接近。
- 个股权重偏离：限制每只股票权重与基准权重的偏离范围在 [0, 5%]，避免单只股票过度集中或权重过小。
- 权重上下限：设置单个股票的权重范围为 [0, 1.5]。下限为 0 避免负持仓，上限为 1.5 避免超额集中持仓。

### 3.2.4 策略展示

多因子线性优化模型指数增强策略回测图如下：



分层抽样指数增强策略收益情况如下：

指标名称	数值
策略收益	72.53%
策略年化收益	5.36%
超额收益	17.98%
基准收益	46.24%
阿尔法	0.017
贝塔	0.986
胜率	0.625
最大回撤	45.44%
夏普比率	0.062
盈亏比	2.124
索提诺比率	0.078
日均超额收益	0.01%
超额收益最大回撤	7.87%
超额收益夏普比率	-1.260
盈利次数	3551
亏损次数	2134
信息比率	0.850
策略波动率	0.219
基准波动率	0.221
最大回撤区间	2015/06/08, 2016/01/28

可观察到与之前分析的策略相比，当前策略的整体表现有所下降。策略收益为 62.23%，虽然仍然显著高于基准收益的 51.63%，但相比之前的 152.16%有明显减少。年化收益为 7.98%，略低于之前的 9.26%，显示出策略的年化回报有所降低。超额收益为 6.99%，远低于之前的 72.43%，这可能表明策略在超越基准方面的表现有所减弱。

策略波动率为 0.238，基准波动率为 0.240，两者相近，表明策略的波动性与市场基准相似。最大回撤区间与之前相同，为 2015/06/08 至 2016/01/28，这可能表明在这一特定时间段内，策略面临了较大的市场压力。

尽管当前策略在选定时间段内的表现有所减弱，尤其是在超额收益、夏普比率、索提诺比率和信息比率方面，但它仍然实现了正的超额收益，并且最大回撤有所减少，显示出一定的风险控制能力。投资者在考虑这一策略时应权衡其风险调整后的回报与市场波动性。

## 四、优化策略

### 4.1 策略概述

在对分层抽样指数增强策略和多因子线性优化模型指数增强策略进行历史数据回测分析后，本报告观察到两种策略均实现了较为可观的年化收益率。然而，两种策略均面临一个共同的挑战：最大回撤较高，均接近 45%。这一现象可能与市场波动性相关，但也表明在风险控制方面存在进一步优化的潜力。

为了降低最大回撤，本报告基于整体效果较好的分层抽样指数增强策略进行了深入优化。优化的核心在于调整策略的资产配置和风险管理机制，以减少极端市场条件下的潜在损失。本报告认为提升因子质量是优化策略的关键。因此，本报告对聚宽平台提供的因子库进行了全面的统计分析和实证检验，最终筛选出了六个预测表现优异的因子，并自定义了前期检测有效的部分因子将它们纳入模型中，替换原有的因子组合，以期提升策略的整体表现。

### 4.2 策略步骤

#### 4.2.1 因子筛选

在构建优化模型的过程中，因子的有效性是策略成功的关键。本节将详细阐述在优化策略中，本报告如何通过分析筛选出因子库中具有显著预测能力的因子。因子筛选过程的核心为以下内容：

##### 1.IC（信息系数）和 ICIR（信息比率）：

IC 衡量因子与未来收益相关性，用 Spearman 秩相关系数计算。ICIR 是 IC 绝对值与其标准差的比值，反映 IC 稳定性。高 IC 和 ICIR 值显示因子强预测力。股票按因子值排序分 10 组，计算日和调仓周期收益率，与指数收益率比对得出超额收益。

##### 2. 因子分组与收益分析：

将因子值排序，分 10 个组合（每组 10%股票）计算日和调仓期收益率，为组合表现分析提供数据。同时，计算基准指数收益率，比较得出超额收益。

##### 3.因子有效性统计测试：

测试一：评估因子分组的年化累计收益序列与线性序列（0 到 9）的相关性。若相关系数的绝对值大于 0.5，则认为该因子在统计上具有显著的预测能力。

测试二：检验高因子值组（赢家组）是否显著超越基准指数，同时低因子值组（输家组）是否显著低于基准指数。若赢家组超额收益超过 5%且输家组超额收益低于-5%，则该因子在实际投资中具有区分能力。

测试三：评估赢家组和输家组相对于基准指数的超额收益概率。若赢家组超额收益概率超过 50%且输家组超额收益概率低于 50%，则该因子在概率上具有显著的预测优势。

##### 4.因子筛选标准：

首先做基本筛选：仅选择 IC 值大于 0.07 且 ICIR 绝对值大于 0.4 的因子，确保因子具有较高的预测能力和稳定性。其次是有效性测试：因子必须通过测试一、测试二（胜者组）和测试三（胜者组），确保其不同市场条件下的稳健性。最后进行额外测试：至少通过测试二（败者组）或测试三（败者组）中任一，以进一步验证因子的预测能力。因子测试结果如下：

	IC	ICIR	测试一	测试二-胜者组	测试二-败者组	测试三-胜者组	测试三-败者组
ACCA	0.009434	0.140539	False	False	False	False	False
AR	0.013775	-0.079030	False	False	True	True	False
ARBR	0.013745	-0.127697	False	False	False	True	True
ATR14	0.021435	0.135814	True	True	False	True	True
ATR6	0.011800	0.077498	True	True	False	True	False
BR	0.016868	-0.109941	True	False	True	False	True
DAVOL10	0.001810	-0.012166	False	False	True	False	False
DAVOL20	0.004833	0.030740	False	False	True	True	True
DAVOL5	0.003056	-0.022502	False	False	True	False	False
DEGM	0.012167	0.171128	False	False	False	True	False
EBIT	0.063906	0.359028	True	True	True	True	False
EBITDA	0.082254	0.470848	True	True	True	True	False
Kurtosis120	0.053184	-0.462869	True	True	False	True	False
Kurtosis20	0.018262	-0.194504	True	False	False	False	False
Kurtosis60	0.042086	-0.364922	True	False	True	False	True
MAWVAD	0.006173	0.030136	True	False	False	False	False
MLEV	0.008405	0.076771	False	False	False	True	False
OperateNetIncome	0.081857	0.413002	True	True	False	True	True
OperatingCycle	0.021718	-0.247652	False	False	False	True	False
ROAEBITTM	0.046948	0.311065	True	True	False	True	False
Skewness120	0.005280	-0.043063	False	False	True	False	True
Skewness20	0.061780	-0.532087	True	False	False	False	True
Skewness60	0.043920	-0.375978	True	False	True	False	True
TVMA20	0.006136	0.038496	True	False	True	False	False
TVMA6	0.002182	-0.013818	True	False	True	False	True
TVSTD20	0.016287	-0.108613	False	False	True	True	True
TVSTD6	0.011039	-0.078070	True	False	True	False	True
VDEA	0.035815	-0.247722	False	False	False	True	True
VDIFF	0.030939	-0.196465	False	False	True	True	True
VEMA10	0.037485	-0.221295	True	False	False	True	False
...	...	...	...	...	...	...	...
operating_profit_ttm	0.087531	0.420166	True	True	True	True	False
operating_revenue_growth_rate	0.026585	0.253185	True	False	True	True	False
operating_revenue_per_share	0.054036	0.436217	True	True	False	True	False
operating_revenue_per_share_ttm	0.055671	0.475950	True	True	True	True	True
operating_revenue_ttm	0.068567	0.406699	True	False	False	True	False
operating_tax_to_operating_revenue_ratio_ttm	0.013732	0.146427	False	False	False	True	False
quick_ratio	0.016013	0.320133	True	False	False	True	False
retained_earnings	0.098521	0.514923	True	True	True	True	True
retained_earnings_per_share	0.084935	0.603128	True	True	False	True	False
retained_profit_per_share	0.085434	0.604663	True	True	False	True	False
roa_ttm	0.036267	0.265224	True	True	False	True	False
roe_ttm	0.061415	0.362024	True	True	True	True	False
sale_expense_to_operating_revenue	0.023433	0.231924	True	False	False	True	False
sale_expense_ttm	0.051880	0.346276	True	True	False	True	True
sharpe_ratio_120	0.047781	0.261124	True	False	True	True	False
sharpe_ratio_20	0.049673	0.277254	True	False	True	True	False
sharpe_ratio_60	0.097450	0.609436	True	True	True	True	False
super_quick_ratio	0.015269	0.305149	True	False	False	True	False
surplus_reserve_fund_per_share	0.060964	0.575908	True	True	False	True	False
total_asset_growth_rate	0.022474	0.227699	True	False	True	True	False
total_asset_turnover_rate	0.025554	0.299802	False	False	True	True	False
total_operating_cost_ttm	0.060871	0.384687	True	True	False	True	False
total_operating_revenue_per_share	0.053978	0.436106	True	True	True	True	False
total_operating_revenue_per_share_ttm	0.055355	0.475209	True	True	True	True	True
total_operating_revenue_ttm	0.068452	0.406537	True	False	False	True	False
total_profit_growth_rate	0.019130	0.221757	True	False	False	True	False
total_profit_to_cost_ratio	0.003631	-0.045856	False	False	False	True	False
total_profit_ttm	0.086535	0.425337	True	True	True	True	True
turnover_volatility	0.103649	-0.604130	True	False	True	True	True
value_change_profit_ttm	0.011379	0.115352	False	True	False	True	False



5. 筛选有效因子：

根据因子测试，筛选出如下因子：

	IC	ICIR	测试一	测试二-胜者组	测试二-败者组	测试三-胜者组	测试三-败者组
EBITDA	0.082254	0.470848	True	True	True	True	False
VOL120	0.104381	-0.571012	True	True	True	True	True
VOL240	0.114044	-0.592253	True	True	True	True	True
net_profit_ttm	0.084305	0.414241	True	True	True	True	False
np_parent_company_owners_ttm	0.083724	0.409508	True	True	True	True	False
operating_profit_per_share_ttm	0.083361	0.477216	True	True	True	True	False
operating_profit_ttm	0.087531	0.420166	True	True	True	True	False
retained_earnings	0.098521	0.514923	True	True	True	True	True
sharpe_ratio_60	0.097450	0.609436	True	True	True	True	False
total_profit_ttm	0.086535	0.425337	True	True	True	True	True

6. 计算因子相关系数：

通过计算因子相关系数，再次筛选有效因子，根据相关系数结果，去除因子 'operating\_profit\_ttm','net\_profit\_ttm','np\_parent\_company\_owners\_ttm','VOL120'

	VOL240	retained_earnings	operating_profit_per_share_ttm	operating_profit_ttm	net_profit_ttm	VOL120	EBITDA	np_parent_company_owners_ttm	total_profit_ttm	sharpe_ratio_60
VOL240	1.00	0.31	0.08	0.22	0.24	0.93	0.20	0.23	0.24	0.13
retained_earnings	0.31	1.00	0.48	0.76	0.76	0.25	0.58	0.76	0.77	0.02
operating_profit_per_share_ttm	0.08	0.48	1.00	0.69	0.64	0.03	0.32	0.65	0.64	0.01
operating_profit_ttm	0.22	0.76	0.69	1.00	0.96	0.17	0.59	0.95	0.96	0.02
net_profit_ttm	0.24	0.76	0.64	0.96	1.00	0.18	0.59	0.98	0.99	0.02
VOL120	0.93	0.25	0.03	0.17	0.18	1.00	0.15	0.18	0.18	0.16
EBITDA	0.20	0.58	0.32	0.59	0.59	0.15	1.00	0.57	0.61	0.00
np_parent_company_owners_ttm	0.23	0.76	0.65	0.95	0.98	0.18	0.57	1.00	0.98	0.03
total_profit_ttm	0.24	0.77	0.64	0.96	0.99	0.18	0.61	0.98	1.00	0.02
sharpe_ratio_60	0.13	0.02	0.01	0.02	0.02	0.16	0.00	0.03	0.02	1.00

经过上述严格的筛选流程，本报告最终确定了以下六个有效因子，它们在统计和实际投资中均展现出了卓越的预测能力：'operating\_profit\_per\_share\_ttm'、'retained\_earnings'、'total\_profit\_ttm'、'EBITDA'、'VOL240'、'sharpe\_ratio\_60'

4.2.2 策略实现

完成因子筛选后，本报告构建了多因子模型组合，并实施分层抽样指数增强策略，包括数据预处理（清洗、标准化，确保数据质量）和多重共线性处理（通过对称正交方法降低因子间相关性，提高模型稳定性和预测能力）。

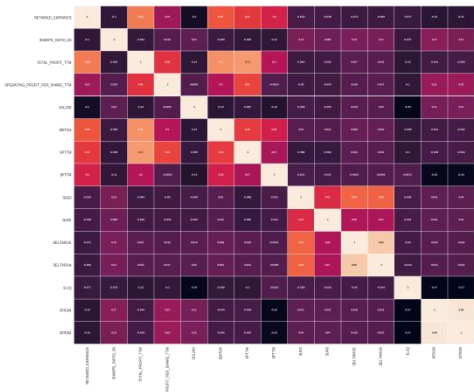


图 4. 因子正交前的相关性的热力图

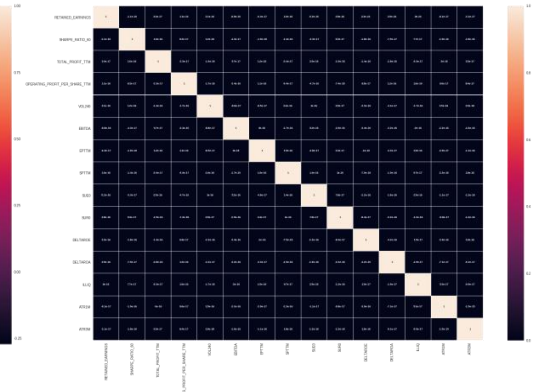


图 5. 因子正交后的相关性热力图

打分法预期收益：用打分法评估因子，结合权重计算股票预期收益，指导投资决策。



策略实施：基于预期收益构建组合，在模拟环境中执行交易。回测分析：通过历史数据评估策略表现。回测显示，优化后策略最大回撤降至 29.88%，保持高收益，验证了优化有效性，显示精细化因子选择和风险管理能提升风险调整后收益。

4.2.3 策略展示

优化策略回测图如下：



优化策略收益情况如下：

指标名称	数值
策略收益	107.24%
策略年化收益	7.22%
超额收益	41.71%
基准收益	46.24%
阿尔法	0.034
贝塔	0.624
胜率	0.493
最大回撤	29.88%
夏普比率	0.218
盈亏比	1.819
索提诺比率	0.278
日均超额收益	0.02%
超额收益最大回撤	25.49%
超额收益夏普比率	-0.061
盈利次数	641
亏损次数	659
信息比率	0.357
策略波动率	0.148
基准波动率	0.221
最大回撤区间	2015/06/08,2015/08/25

优化后的策略在多个关键方面显著提升了风险控制能力。最大回撤率从 46.10%大幅降低至 29.88%，显著减少了投资组合在极端市场情况下的潜在亏损，增强了资金的安全性。策略波动率由 0.236 降至 0.148，表明策略整体的波动性显著降低，投资表现更加稳定。贝塔值从 1.032 降至 0.624，表明策略对市场波动的敏感度降低，减少了系统性风险的暴露。盈亏比也从 1.423 提升至 1.819，显示每单位亏损对应的收益增加，进一步优化了风险回报结构。超额收益最大回撤从 10.94%增至 25.49%，虽然数值有所上升，但整体策略在控制主要回撤方面表现出色。

总体而言，优化后的策略显著降低了风险，提高了稳定性，增强了资金的安全保障。

## 五、总结

### 5.1 策略总结

本文通过构建收益预测模型，开发了两种量化多因子指数增强策略：分层抽样策略和多因子线性优化模型策略。这两种策略均旨在控制跟踪误差的同时获取超越基准指数的收益。研究发现，尽管两种策略均实现了可观的年化收益率，但面临最大回撤较高的共同挑战。

为此，本报告对分层抽样指数增强策略进行了深入优化，特别关注了资产配置和风险管理机制的调整，并采用了一套严格的统计方法来识别更具预测能力的因子。优化后的策略在保持较高整体收益水平的同时，将最大回撤风险降低至 29.88%，显著提升了策略的风险调整后收益。

### 5.2 心得体会

在本次多因子指数增强策略的研究与优化过程中，我们深化了对量化投资多因子模型的理解，进一步感受到了量化分析的独特魅力。从最初课堂上接触的基础多因子模型，如 Fama-French 三因子模型，到在实际项目中构建与优化多因子指数增强模型，我们经历了从理论学习到实践应用的深刻转变。在这一过程中我们认识到，成功的量化投资不仅仅依赖于理论框架的构建，更离不开因子选择与优化的精细化操作。在因子选择上，我们深入探讨了有效因子的识别方法，并结合分层抽样与多因子优化模型进行了系统对比分析。通过这一过程，我们发现初步策略虽能带来一定的超额收益，但也伴随较高的最大回撤风险，这促使我们思考如何优化策略以减少潜在的市场风险。反复的实验和调优使我们意识到，精确的因子选择与严格的风险管理是提升策略长期表现的核心要素。

此次研究不仅显著提升了我们的量化分析能力，也加深了我们对市场机制的理解。从中我们深刻感悟到，在量化投资的世界里，每一个数据点都可能潜藏着市场的规律，每一个模型都可能是解读市场信号的钥匙。然而，量化投资的道路上并没有绝对完美的策略，只有不断试探、不断调整的过程。只有通过持续的策略评估与优化，才能应对市场的复杂性与动态变化，保持策略的竞争力。

此外，我们也意识到，量化投资是一个迅速发展的领域，新的技术与方法不断涌现。机器学习、深度学习等前沿技术的应用也为因子筛选与策略优化提供了新的视角和可能性。为了在激烈的行业竞争中保持领先地位，行业从业者需要不断学习并掌握这些新兴技术，持续提升自身的专业素养，只有不断探索与创新，才能应对未来市场的挑战，以便为投资者创造更大的价值。

## 参考文献

- [1] 李函逊.基于多因子模型的量化选股策略[J].中国外资,2024,(第 18 期).
- [2] 邸艺琳.量化交易策略之多因子指数增强策略[J].电脑乐园,2022,(第 5 期).
- [3] 冼彤.多因子打分和多因子回归模型对比分析——以上证 180 指数为例[J].现代商业,2023,(第 8 期).
- [4] 陈东华.基于中证 500 指数增强的多策略研究[J].华南理工大学, 2022.