

ABCM：基于神经网络的 alpha 因子和 beta 因子协同挖掘模型

——因子选股系列之一——

报告发布日期

2024 年 12 月 03 日

证券分析师

杨怡玲 yangyiling@orientsec.com.cn
执业证书编号：S0860523040002
陶文启 taowenqi@orientsec.com.cn
执业证书编号：S0860524080003

相关报告

相对定价类基本面因子挖掘：——因子选股系列之一〇九 2024-10-11
KD-Ensemble：基于知识蒸馏的 alpha 因子挖掘模型：——因子选股系列之一〇八 2024-08-19
融合基本面信息的 ASTGNN 因子挖掘模型：——因子选股系列之一〇四 2024-05-27

研究结论

- 基于基本面的 Barra 模型中的风险因子往往考虑的是超长周期的风险，且模型中基本面信息占比相对量价信息更高，因此这种风险模型对于中高频策略可能力不从心。为了克服这些问题，我们提出了一套基于神经网络和量价数据驱动的风险因子生成模型，我们称之为 ABCM (Alpha-Beta Co-mining) 模型。

样本外各风险因子的表现：

- RankIC 基本上都在 3%以内、ICIR 基本上都在 0.2 以内而 RankIC 胜率基本上都在 60%以内，意味着各个风险因子均不具有显著的选股能力；
- Abs(RankIC)均较高，说明各因子对未来收益率的解释能力较强；
- 自相关系数均在 70%以上，说明短期来看各个风险因子衰减速度较弱；而各风险因子两两之间相关性基本上都在 20%以内，信息独立性较强；
- 各个股票池上 ABCM1&2 风险因子滚动 Rsquare 曲线均稳定位于 Barra 风险因子上方，且相对于 Barra 风险因子，ABCM2 风险模型 Rsquare 显著提升，在沪深 300、中证 500 和中证全指上分别提升至 47.02%、34.28%和 20.53%。因此基于神经网络生成的风险模型对未来收益率的解释能力整体强于 Barra 风险模型。

样本外“伴生”的 alpha 因子的表现：

- 有着较强的获取超额收益的能力，2017 年至今年化超额达 36.55%。且该因子往年超额收益回撤仅有 5.86%，且今年该因子最大回撤相对不大，这说明该因子能稳定的获取较高的超额。
- 今年三季度该 alpha 因子的多头表现较为稳定未出现大幅回撤，另外一方面该因子与各个 Barra 风险因子相关性较低，这说明通过 ABCM 模型生成 alpha 因子获取 alpha 信息的来源与常规机器学习生成的 alpha 因子差异性较大，与常规机器学习因子有着较强互补性。
- ABCM2 模型生成的 alpha 因子对我们已有的 AI 量价模型生成 alpha 因子能够起到较大增量作用，通过把前者纳入到我们已有的 AI 量价模型中进行合成，新生成的 alpha 因子在中证全指、中证 500 和中证 1000 上 RankIC 分别提升至 16.39%、11.94%和 14.94%，多头年化超额分别提升至 52.63%、26.46%和 39.61%，在中证全指上 RankICIR 也提升至 1.66，说明新因子获取 alpha 收益的能力更强，稳定性也更好。

风险因子用于组合优化的表现：

- 相对于使用 Barra 风险模型，将 ABCM2 风险模型生成的风险因子直接用于组合优化，新组合具有更高的信息比率，其中中证 500 增强组合由 2.79 提升至 3.30，中证 1000 增强组合则由 3.83 提升至 4.25。并且组合的年化超额以及超额最大回撤均有显著改善。

风险提示

- 量化模型基于历史数据分析，未来存在失效风险，建议投资者紧密跟踪模型表现。
- 极端市场环境可能对模型效果造成剧烈冲击，导致收益亏损。

目录

一、 引言	4
1.1 多因子选股策略	4
1.2 传统风险模型	4
1.3 机器学习风险模型	5
二、 基于 AI 的风险因子挖掘框架改进方案	7
三、 ABCM 模型生成因子的表现	8
3.1 风险因子表现	9
3.2 生成的 alpha 因子表现	12
3.3 生成的因子相关性分析	13
四、 ABCM 模型对 alpha 端的增量	14
4.1 中证全指上的表现	14
4.2 各宽基指数上的表现	16
4.3 因子风险暴露分析	18
五、 ABCM 风险因子在指增上的应用	19
5.1 增强组合构建说明	19
5.2 沪深 300 指数增强	19
5.3 中证 500 指数增强	21
5.4 中证 1000 指数增强	22
六、 结论	23
风险提示	24

图表目录

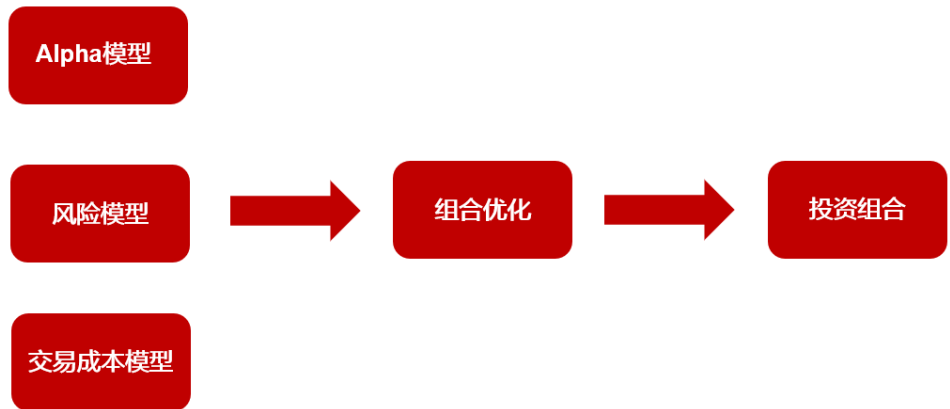
图 1: 多因子选股策略框架.....	4
图 2: DFQ-Risk 风险因子.....	5
图 3: 风险因子生成模型.....	6
图 4: 数据集 1 生成风险因子绩效表现 (20170101~20240930).....	9
图 5: 数据集 2 生成风险因子绩效表现 (20170101~20240930).....	9
图 6: 因子 id0 时序 RankIC 曲线和多空净值曲线 (20170101~20240930).....	10
图 7: 因子 id1 时序 RankIC 曲线和多空净值曲线 (20170101~20240930).....	10
图 8: 全市场风险因子滚动 243 日 Rsquare (20180101~20240930).....	11
图 9: 沪深 300 风险因子滚动 243 日 Rsquare (20180101~20240930).....	11
图 10: 中证 500 风险因子滚动 243 日 Rsquare (20180101~20240930).....	11
图 11: ABCM 模型生成 alpha 因子的绩效表现 (20170101~20240930).....	12
图 12: ABCM 模型生成 alpha 因子今年多头超额净值走势 (20231229~20240930).....	12
图 13: ABCM 模型生成因子相关系数矩阵.....	13
图 14: ABCM 模型数据集 1 上生成因子与 Barra 风险因子相关系数.....	13
图 15: ABCM 模型数据集 2 上生成因子与 Barra 风险因子相关系数.....	14
图 16: 中证全指选股汇总表现 (回溯期 20171229~20240930).....	15
图 17: 中证全指因子各分组超额表现.....	15
图 18: 各模型多头超额净值走势 (20180101~20231231).....	15
图 19: 各模型多头超额净值走势 (20231229~20241031).....	15
图 20: 中证全指各年度多头组合选股表现 (回溯期 20171229~20241031).....	16
图 21: 各宽基指数上选股表现 (回溯期 20180101~20241031).....	16
图 22: 今年沪深 300 多头超额走势 (截至 20241031).....	17
图 23: 今年中证 500 多头超额走势 (截至 20241031).....	17
图 24: 今年中证 1000 多头超额走势 (截至 20241031).....	17
图 25: 因子风险暴露时序曲线.....	18
图 26: 沪深 300 指增组合分年度超额收益率 (回溯期 20171229~20241031).....	20
图 27: 沪深 300 指增组合超额净值走势 (20171229~20241031).....	20
图 28: 不同暴露倍数下各风险模型对应沪深 300 组合的信息比率.....	20
图 29: 中证 500 指增组合分年度超额收益率 (回溯期 20171229~20241031).....	21
图 30: 中证 500 指增组合净值走势 (20171229~20241031).....	21
图 31: 不同暴露倍数下各风险模型对应中证 500 组合的信息比率.....	21
图 32: 中证 1000 指增组合分年度超额收益率 (回溯期 20171229~20241031).....	22
图 33: 中证 1000 指增组合净值走势 (20171229~20241031).....	22
图 34: 不同暴露倍数下各风险模型对应中证 1000 组合的信息比率.....	22

一、引言

1.1 多因子选股策略

多因子选股策略是量化投资中的主流 Alpha 策略，它通过结合多个具有预测能力的因子来构建投资组合。多因子选股框架通常包括 alpha 模型、风险模型、交易成本模型和组合优化四个部分，整个多因子选股策略的框架可表示为下图所示的过程。其中 alpha 模型主要是对个股未来的收益率或特质性收益率在整体股票池中的排序进行预测，风险模型则是负责对个股之间的协方差矩阵进行估计，交易成本模型是对个股的交易成本进行预测，而组合优化则是在考虑交易成本的前提下，负责在给定风险（即组合的波动）下极大化预期收益从而构建组合，以期在样本外组合能有较好的风险收益比。

图 1: 多因子选股策略框架



数据来源：东方证券研究所绘制

这四个部分中 alpha 模型更加受到重视，近年来基于人工挖掘和机器学习相关的 alpha 因子挖掘方法层出不穷，前期报告《基于循环神经网络的多频率因子挖掘》、《基于残差网络端到端因子挖掘模型》和《融合基本面信息的 ASTGNN 因子挖掘模型》等报告中我们也对 alpha 因子生成提出了一系列基于机器学习的方案，此处就不再赘述了。但研究人员对于风险模型的研究则少之又少，在今年各类 alpha 模型遭受几轮较大回撤之后，风险模型在多因子选股体系下的作用显得尤为重要。

1.2 传统风险模型

主流的风险模型主要分为两类，基本面主导的 Barra 风险模型和价格数据驱动的统计量风险模型，前者主要通过人工方法构建相应的风险因子从而建立模型，后者主要通过一些降维等数据分析的方法构建风险因子从而建立模型。研究人员通常使用的风险模型是 Barra 体系下的风险模型或者其的改进版本，例如报告《DFQ-Risk: 东方 A 股因子风险模型》中所使用的十个风险因子构建如下图所示：

图 2: DFQ-Risk 风险因子



数据来源: 东方证券研究所绘制

这套模型中的风险因子往往考虑的是超长周期的风险(自相关性过高, 长时间内风险因子取值几乎不发生变化), 且基本面信息占比相对量价信息更高, 因此这种风险模型对于低频策略匹配度相对较高, 但对于中高频策略可能显得力不从心。

与 Barra 风险模型对应的是**统计风险模型**, 这个模型完全通过价格数据进行驱动, 使用各个股票过去 252 个交易日的收益率序列计算相关系数矩阵或度量矩阵, 之后使用 PCA、谱聚类等方法进行降维, 取相关系数矩阵前十大特征值(或主成分)对应特征向量作为风险因子, 以谱聚类算法为例, 假设各个股票过去 252 个交易日的收益率序列为 x_1, x_2, \dots, x_n , 则主特征向量 u_1, u_2, \dots, u_n 为下列优化问题的非常解:

$$\min_u R(x_i, x_j) \|u_i - u_j\|^2$$

其中核函数 $R(x_i, x_j)$ 通常反比于数据点 x_i 和 x_j 之间的距离度量。而上述优化问题的解则可转化为求解矩阵 $D - L$ 的主特征向量, 这里

$$L = (R(x_i, x_j))_{i,j}, D = \text{diag}(\sum_j R(x_i, x_j))$$

该算法核心思想是使用数据之间构建的有限维的图 Laplacian 矩阵 $D - L$ 及其特征向量来逼近数据分布的底层流形上的 Laplacian 算子及其特征函数, 因为流形的 Laplacian 算子是其几何内蕴量, 其特征函数空间可以张成整个流形上的非线性函数空间, 因此图 Laplacian 矩阵的主特征向量可视作流形上的最佳逼近量, 相关的实验研究表明这套风险模型相较于 Barra 模型有着更加优异的性能。

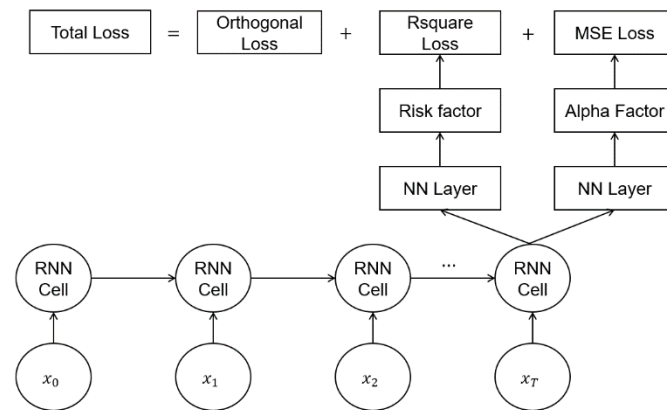
1.3 机器学习风险模型

Alpha 模型更多的是提升所构建组合的收益, 而风险模型则对提升组合业绩稳定性起到重要作用。在有着强大 alpha 模型的基础上, 我们有必要构建一套与之匹配的风险模型。而机器学习系列方法有着强大的拟合能力, 其是否能够和生成 alpha 因子一样生成风险因子。首先, 风险模型的关键在于构建相应的风险因子, 我们认为风险因子应该满足以下性质:

1. 没有显著的选股能力，即 RankIC 相对较低，对未来收益率预测的方向波动十分剧烈，即 ICIR、RankIC 胜率等指标较低。
2. 对未来收益率有较强的解释能力（具有较强的线性相关性），体现在因子 RankIC 绝对值、Rsquare 等指标相对较高。
3. 因子衰减速度较慢，即因子跨期截面自相关性较高。

基于上述性质，在前期报告《融合基本面信息的 ASTGNN 因子挖掘模型》和《KD-Ensemble: 基于知识蒸馏的 alpha 因子挖掘模型》中，我们借助 RNN 模型搭建了一套如下图所示的风险因子生成模型，整个模型结构如下图所示：

图 3：风险因子生成模型



数据来源：东方证券研究所绘制

该模型将个股的量价数据和风险因子构成时序数据，然后再将该时序数据作为 RNN 的输入，取最后一个 RNN-Cell 输出的特征向量通过两个参数不同的全连接层分别生成 alpha 因子和风险因子，该模型训练的损失函数则定义为：

$$Loss = MSE(F_{:K}, y_{neu}) + Rsquare(F_{K:}, y) + \lambda \|corr(F, F)\|$$

其中， F 为所有生成因子，我们设定前 K 个为 alpha 因子， K 个以后为风险因子， y_{neu} 为行业市值中性化之后的收益率标签， y 为原始收益率标签， λ 是人工调节的超参数， $\|corr(F, F)\|$ 表示所有因子的相关系数矩阵的范数。这套框架生成的风险因子样本外对未来收益率有着较好的解释力度。但这套框架仍然存在一些技术问题：

1. Alpha 因子对应标签设置为了中性化之后的收益率标签，但这个标签中性化方法如果“误差”较大的话，可能既没有完全剥离风险信息也损失了一部分 alpha 信息，最终导致 alpha 因子样本外表现不理想。
2. 损失函数中只考虑了风险因子对未来收益率的解释力度（即 Rsquare），但事实上 alpha 因子对未来收益率仍然能贡献一部分的 Rsquare，因此风险因子可能混入较多的 alpha 信息，导致生成的风险因子 RankIC、多头超额等指标较高，使得模型目标无法达成。
3. 对风险因子自相关性没有进行考量，可能使得部分风险因子自相关性较低。
4. 风险因子产生的原因很大程度上由于某种维度上聚类得到的相似股票在不同时期内表现出的“同涨同跌”效应，因此风险因子生成应该考虑到个股之间的交互关系，而 alpha

因子应该是个股特质性收益率，不应该受到其他股票影响。基于此，对应这套框架的 RNN 模型结构也有必要做出调整。

二、基于 AI 的风险因子挖掘框架改进方案

正如前文所述，传统基于 Barra 体系构建的风险因子更适合描述超长期风险，而对于中短期的风险描述可能有所欠缺。而我们前期报告基于 RNN 构建的风险因子挖掘模型在结构和损失函数的设计存在较多不合理之处，导致样本外风险因子仍不能满足我们所希望达到的目标。为了克服上述问题，我们对收益率成分进行如下分解：

$$Ret_t = \alpha_t + \beta_t + \varepsilon_t$$

其中 α_t 表示 t 时刻个股收益率中股票特质性收益率成分（该成分方向稳定较容易预测）， β_t 表示 t 时刻个股收益率中受市场、风格等宏观因素带来的影响的成分（该成分方向波动剧烈较难预测，但其绝对值排序信息较容易预测）， ε_t 表示 t 时刻个股收益率中噪声成分。

模型结构中 alpha 因子部分主要是对成分 α_t 的相对大小和方向同时进行拟合，其对应的 NN-Layer 为简单的全连接层，而其对应的损失函数为与原始收益率的 MSE 损失（MSE 损失考虑了标签与预测结果的同方向）。而风险因子主要是对成分 β_t 的相对大小进行拟合，其对应的损失函数为 Rsquare（Rsquare 只考虑了线性相关性而不考虑方向），其对应的 NN-Layer 为具有非对称图结构的 ASTGNN 结构，该非对称图结构具有如下数学表达式：

$$\text{adj_martrix} = \text{ReLU}(\text{fill_diagonal}(\mathbf{M}_1 \mathbf{M}_2^T))$$

$$\mathbf{H} = \mathbf{X} + \text{softmax}(\text{adj_martrix}) \mathbf{M}_3$$

$$\mathbf{F}_K = \mathbf{W}_2 \mathbf{H}$$

这里矩阵 \mathbf{X} 表示最后一个 RNN-Cell 的输出的风险因子部分，矩阵 \mathbf{M}_1 、 \mathbf{M}_2 和 \mathbf{M}_3 分别为矩阵 \mathbf{X} 经过不同参数的全连接变换得到的，矩阵 \mathbf{M}_1 、 \mathbf{M}_2 和 \mathbf{M}_3 的规模均为 $N \times M$ ， N 表示截面股票个数， M 表示生成风险因子的个数， \mathbf{W}_2 为全连接层的权重参数，fill_diagonal 表示邻接矩阵对角元置 0 操作。上述结构设计主要基于以下两个原因：

1. 基于统计风险模型的想法，我们希望图结构主要是捕捉其他股票对个股本身的影响，因而剔除自身影响的部分（对应股票邻接矩阵的对角元）。
2. 考虑到个股之间影响可能并不对等，如概念、风格和板块的领涨龙头往往对成分内其他个股的影响较大，而反过来其他个股对龙头股影响可能微乎其微，因此我们设计了具有非对称形式的邻接矩阵。

根据上述过程，该模型训练的损失函数则可定义为：

$$\text{Loss} = \text{MSE}(\mathbf{F}_{:K}, y_1) + \text{Rsquare}(\mathbf{F}, y_2) + \lambda \|\text{corr}(\mathbf{F}, \mathbf{F})\| + \text{TO}(\mathbf{F}_K)$$

$$\text{Rsquare}(\mathbf{F}, y_2) = 1 - \frac{\|y_2 - \mathbf{F}(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T y_2\|^2}{\|y_2\|^2}$$

其中，我们设定生成的因子中前 K 个为 alpha 因子， K 以后 M 个为风险因子， \mathbf{F} 为所有股票对应因子的矩阵，该矩阵的规模为 $N \times (K + M)$ ， y_1 和 y_2 为不同频率的原始收益率标签， λ 是人工调节的超参数， $\text{TO}(\mathbf{F}_K)$ 表示 K 个以后所有风险因子对应的换手惩罚，该项惩罚主要保证生成风

险因子自相关系数不至于过低。上述损失函数的设计理论上可以有效的克服第一章所提到的技术性问题:

1. 通过让 alpha 因子和风险因子共同来对完整收益率进行“解释”，使得 alpha 因子和风险因子能够最大化逼近 $\alpha_t + \beta_t$ 部分，而损失函数 MSE 和 Rsquare 则分别保证 alpha 因子和风险因子分别逼近 α_t 和 β_t ，最后通过正交惩罚保证 alpha 因子和风险因子彼此正交（低相关）来实现模型自助将 alpha 因子中的风险信息进行剥离。这种做法可以一定程度上避免了因人工中性化带来的信息损失。
2. 我们将 alpha 因子和风险因子共同计算 Rsquare 作为损失函数第二项。这样可以考虑到 alpha 因子所贡献的 Rsquare 部分，并且如果 alpha 信息进入风险因子中，将导致 Rsquare 部分损失不变而 MSE 损失部分上升，因此这种损失函数的构建可以有效的抑制风险因子具有较强“选股能力”的问题。

注意到上述损失函数中，MSE 损失和 Rsquare 损失量纲有所差异，此处为了避免调参导致的过拟合风险，我们选取这两项损失比重为一比一，但适当引入超参数调节这两项损失的比重可能能进一步提高风险因子和 alpha 因子在样本外的表现性能。

而模型的输入端主要是一些基础的字段如高开低收、vwap、换手率和风险因子等经过必要的去极值、标准化和补充缺失值等预处理步骤之后形成时序数据，预处理的具体细节可参考报告《基于循环神经网络的多频率因子挖掘模型》再输入神经网络模型进行训练。我们把这一套同时生成 alpha 因子和风险因子的模型称之为 ABCM (Alpha-Beta Co-mining) 模型。特别的我们考虑两个不同频率的数据集分别产生风险因子，并用这两组风险因子来分别刻画中短期风险和长周期量价风险:

数据集 1: 日频 k 线数据包括个股每日高开低收、vwap、换手率和一些风险因子等;

数据集 2: 月频 k 线数据，加上一些长周期基本面因子等。

模型 ABCM1: 仅用数据集 1 通过本章构建的神经网络生成风险因子和 alpha 因子 (风险因子个数为 12 个);

模型 ABCM2: 使用数据集 1 和 2 分别通过神经网络模型生成风险因子和 alpha 因子然后进行合并 (数据集 1 和 2 分别生成 12 和 8 个风险因子)。

三、ABCM 模型生成因子的表现

本章我们将对 ABCM1 和 2 模型生成的风险因子和 alpha 因子在样本外的表现进行分析，各项指标定义如下:

1. RankIC 表示回测区间内双周频 (与 T+1~T+11 收盘之间收益率计算得到的) RankIC 指标的各期均值，Abs(RankIC)则表示回测区间内双周频各期 RankIC 绝对值的均值;
2. ICIR 为双周频各期 RankIC 均值比上标准差 (此项指标未年化)，RankIC 胜率为回测区间内各期 RankIC>0 数量占比;
3. 自相关系数为风险因子在各个截面上取值与滞后五个交易日截面取值的相关系数，Rsquare 则是根据 T 日收盘因子取值和 T+1~T+21 收盘之间收益率计算得到的;

4. Top 组超额收益率, 中证全指为分二十组第一组相对基准的超额, 而沪深 300、中证 500 和中证 1000 则是分十组计算超额。

后续章节若不做特殊说明则指标均按照此处定义。

3.1 风险因子表现

首先, 我们展示 ABCM 模型在两个数据集上所生成风险因子在全市场上样本外的绩效表现 (因子方向统一调成 RankIC>0 的方向), 以及数据集 1 对应风险因子组中 id 为 0 和 1 的这两个风险因子时序 RankIC 曲线和多空净值曲线 (多头空头分别为分十组的 top 组合 bottom 组) 分别如下所示:

图 4: 数据集 1 生成风险因子绩效表现 (20170101~20240930)

id	RankIC	Abs(RankIC)	ICIR	RankIC胜率	自相关系数
0	2.67%	13.73%	0.16	59.89%	95.38%
1	0.67%	9.88%	0.05	52.40%	88.77%
2	0.64%	11.82%	0.04	50.27%	95.41%
3	2.87%	8.73%	0.26	63.64%	89.05%
4	1.06%	10.33%	0.08	58.83%	72.93%
5	0.03%	8.44%	0	51.34%	93.53%
6	2.50%	9.57%	0.22	57.75%	90.60%
7	3.17%	17.92%	0.15	52.94%	97.20%
8	0.97%	8.07%	0.1	53.48%	85.08%
9	2.95%	12.93%	0.18	58.29%	79.47%
10	0.17%	8.93%	0.01	51.34%	85.66%
11	0.62%	7.96%	0.06	53.48%	72.52%

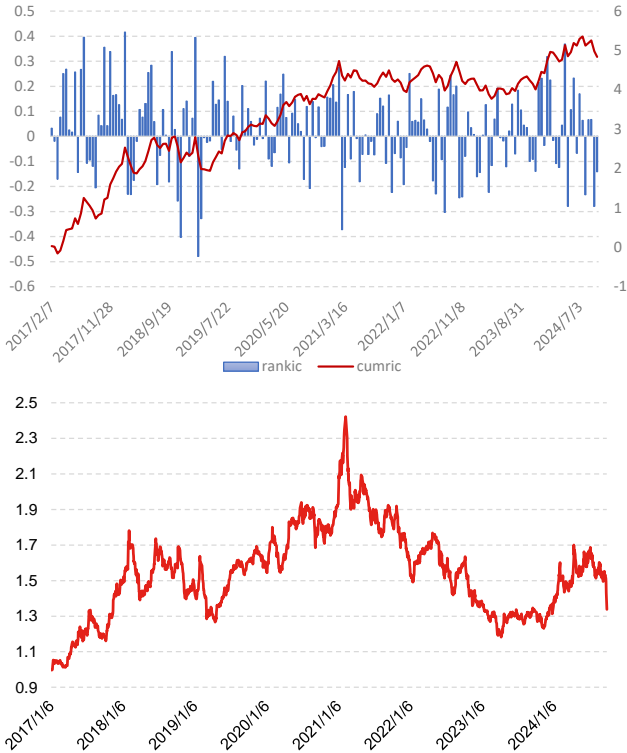
数据来源: wind、上交所、深交所、东方证券研究所

图 5: 数据集 2 生成风险因子绩效表现 (20170101~20240930)

id	RankIC	Abs(RankIC)	ICIR	RankIC胜率	自相关系数
0	1.53%	13.37%	0.09	50.80%	96.77%
1	0.31%	8.39%	0.03	49.73%	93.03%
2	2.54%	9.47%	0.23	60.43%	94.27%
3	2.70%	15.11%	0.15	55.61%	97.14%
4	0.60%	9.39%	0.05	49.73%	95.15%
5	2.99%	10.12%	0.23	56.15%	92.64%
6	1.85%	8.94%	0.23	56.15%	91.88%
7	2.74%	8.10%	0.29	61.00%	93.50%

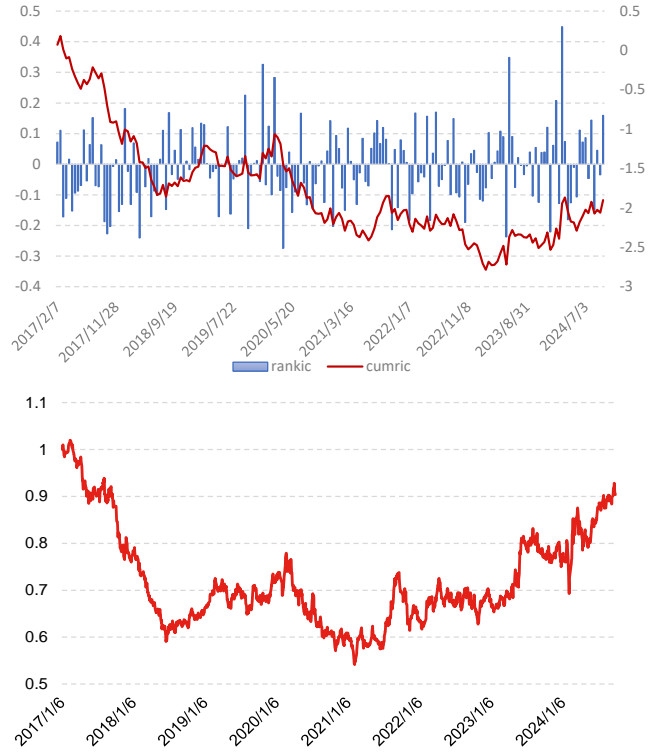
数据来源: wind、上交所、深交所、东方证券研究所

图 6: 因子 id0 时序 RankIC 曲线和多空净值曲线
(20170101~20240930)



数据来源: wind、上交所、深交所、东方证券研究所

图 7: 因子 id1 时序 RankIC 曲线和多空净值曲线
(20170101~20240930)



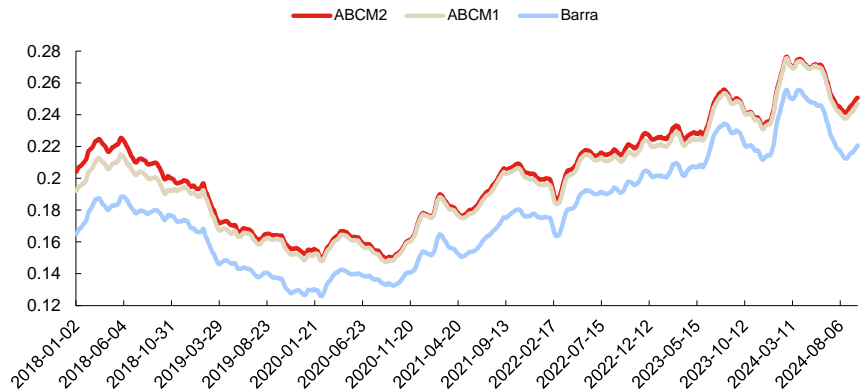
数据来源: wind、上交所、深交所、东方证券研究所

通过上述结果我们可以看出:

1. 两个数据集生成风险因子的 RankIC 基本上都在 3% 以内、ICIR 基本上都在 0.2 以内而 RankIC 胜率基本上都在 60% 以内, 意味着各个风险因子均不具有显著的选股能力, 且对未来收益率预测的方向波动十分剧烈;
2. 两个数据集所生成的风险因子 Abs(RankIC) 均较高, 取值基本上都在 8% 以上, 这说明各个因子对未来收益率的解释能力较强;
3. 数据集 1 所生成风险因子的自相关系数均在 70% 以上, 而数据集 2 生成的风险因子自相关系数均在 90% 以上, 说明短期来看两个数据集产生的各个风险因子衰减速度均较弱, 并且频率越低的数据集产生的风险因子自相关性越高;

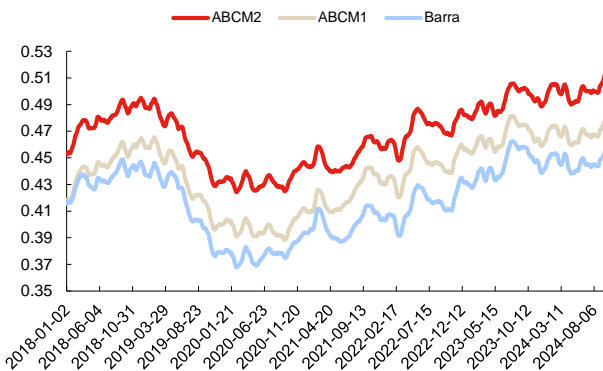
综上所述, 两个数据集所生成风险因子均能较好的满足选股能力弱、自相关性高、对未来收益率解释能力强等预期目标。接着我们将分别对比 ABCM2、ABCM1 和 Barra 模型在各个股票池上的 Rsquare 的表现。

图 8: 全市场风险因子滚动 243 日 Rsquare (20180101~20240930)



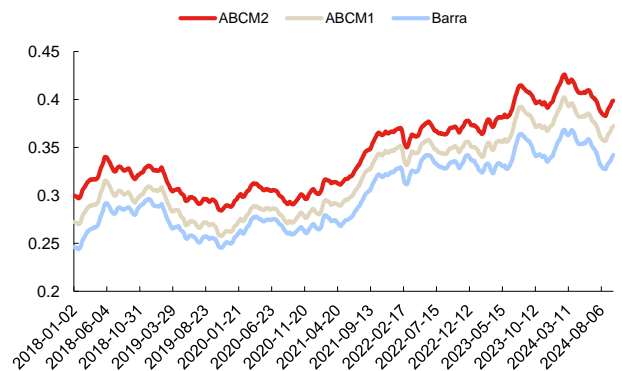
数据来源: wind、上交所、深交所、东方证券研究所

图 9: 沪深 300 风险因子滚动 243 日 Rsquare (20180101~20240930)



数据来源: wind、上交所、深交所、东方证券研究所

图 10: 中证 500 风险因子滚动 243 日 Rsquare (20180101~20240930)



数据来源: wind、上交所、深交所、东方证券研究所

在各个股票池上, ABCM 模型生成风险因子和 Barra 风险因子的 Rsquare 分别为:

ABCM2: 沪深 300 47.02%、中证 500 34.28%、全市场 20.53%;

ABCM1: 沪深 300 43.90%、中证 500 31.92%、全市场 20.08%;

Barra: 沪深 300 41.97%、中证 500 29.86%、全市场 17.82%;

上述图像和结果可以看出:

1. 在各个股票池上 ABCM1 模型生成的风险因子滚动 Rsquare 曲线均稳定位于 Barra 风险因子上方, 且相对于 Barra 风险因子, 回测区间内 Rsquare 指标 ABCM1 模型生成的风险因子显著提升 2%-3%。因此 ABCM1 模型生成的风险因子对未来收益率的解释能力整体强于 Barra 风险因子且该模型风险因子在小盘股里面增量更加明显。
2. 在各个股票池上 ABCM2 模型生成的风险因子滚动 Rsquare 曲线均稳定位于 ABCM1 风险因子上方, 且回测区间内, 相较于 ABCM1 模型, ABCM2 模型的 Rsquare 指标在沪深 300 上的提升幅度大于中证 500 大于中证 1000。因此我们认为低频数据集 2 生成的风险因子可能更加贴合大市值股票池上的规律。

3.2 生成的 alpha 因子表现

接着，我们展示 ABCM 模型在数据集 1 和 2 上生成 alpha 因子以及合成因子在中证全指上的绩效表现以及这些因子今年以来多头超额净值走势：

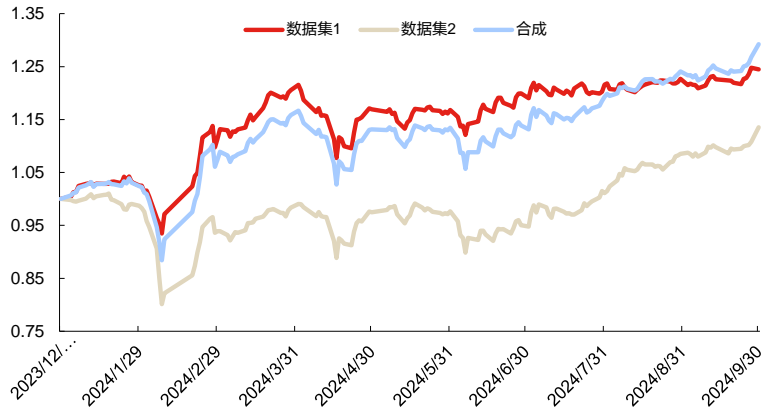
图 11: ABCM 模型生成 alpha 因子的绩效表现 (20170101~20240930)

模型	RankIC	ICIR	RankIC胜率	Top组超额	Top组换手率	maxdd
数据集1	12.69%	0.96	86.63%	34.51%	50.91%	-11.36%
数据集2	11.09%	0.90	83.42%	26.57%	41.45%	-20.68%
合成	13.54%	1.03	87.70%	36.55%	45.16%	-14.90%

	年化收益	回撤区间	年化波动率	最大回撤
2017	26.66%	2017-05-18 - 2017-06-01	7.24%	-4.23%
2018	77.09%	2018-01-18 - 2018-02-06	9.02%	-3.96%
2019	26.40%	2019-05-28 - 2019-06-10	6.19%	-2.83%
2020	32.61%	2020-04-09 - 2020-05-18	8.24%	-5.11%
2021	39.55%	2021-10-14 - 2021-11-29	10.32%	-4.84%
2022	37.54%	2022-03-23 - 2022-04-26	10.95%	-5.86%
2023	19.07%	2023-05-09 - 2023-06-09	7.09%	-2.92%
2024	29.25%	2024-01-25 - 2024-02-07	20.98%	-14.90%
年化	36.53%	2024-01-25 - 2024-02-07	10.45%	-14.90%

数据来源: wind、上交所、深交所、东方证券研究所

图 12: ABCM 模型生成 alpha 因子今年多头超额净值走势 (20231229~20240930)



数据来源: wind、上交所、深交所、东方证券研究所

通过上述图表结果，我们可以看出：

- 即使在模型层面对 alpha 因子的风险信息进行剥离，无论数据集 1 还是数据集 2 生成的 alpha 因子以及合成因子仍然能有着较强的获取超额收益的能力。三个因子虽然都是预测未来十天收益率，但换手率相对常规机器学习因子显著更低，另外合成因子往年多头超额收益回撤仅有 5.86%，且今年该因子最大回撤相对常规机器学习因子不大，这说明 ABCM 模型生成因子的稳定性整体较强。
- 合成因子的 RankIC、ICIR、RankIC 胜率和多头年化超额显著强于数据集 1 和 2 各自生成的 alpha 因子，并且今年三季度，数据集 1 产生因子的多头超额曲线相对较平而数据集 2 产生的

alpha 因子斜率更陡, 这说明两个数据集产生的 alpha 因子各自信息差异较大, 放在一起具有强互补作用, 因而合成因子具有更好的表现。

- 与常规的机器学习因子不同, 在今年三季度各种机器学习因子多头均出现了较大回撤, 但该 alpha 因子的多头在今年三季度表现较为稳定甚至超额相对平时更高, 未出现大幅回撤, 这说明通过 ABCM 模型生成 alpha 因子获取 alpha 信息的来源与常规机器学习生成的 alpha 因子差异性较大, 与常规机器学习因子有着较强互补性。

3.3 生成的因子相关性分析

本节我们将对两个数据集各自生成的风险因子之间以及和 Barra 风险因子之间的相关性进行分析, 以下分别展示了生成因子之间相关系数矩阵以及生成因子与 10 个基于 Barra 风险因子之间相关系数矩阵 (这里 id 为 f0 到 f8 为数据集 2 生成的风险因子, alpha2 表示数据集 2 生成的 alpha 因子, id 为 0 到 13 为数据集 1 生成的风险因子, alpha1 表示数据集 1 生成的 alpha 因子) :

图 13: ABCM 模型生成因子相关系数矩阵

	f0	f1	f2	f3	f4	f5	f6	f7	alpha2	0	1	2	3	4	5	6	7	8	9	10	11	alpha1
f0	100.00%	8.41%	-5.67%	2.28%	-2.98%	3.75%	-11.36%	-26.89%	7.95%	39.43%	-21.30%	42.61%	6.61%	-16.83%	-13.83%	27.68%	4.97%	8.44%	-2.79%	-1.92%	9.91%	-0.21%
f1	8.41%	100.00%	21.01%	-1.73%	-2.20%	-10.42%	-9.34%	-20.31%	-3.13%	-2.54%	-12.72%	-1.78%	10.78%	7.57%	3.09%	-13.44%	-22.36%	-15.39%	5.40%	20.52%	12.56%	2.85%
f2	-5.67%	21.01%	100.00%	-6.44%	14.46%	2.49%	-10.95%	-15.90%	-0.39%	-25.04%	8.61%	-16.59%	20.99%	10.74%	-23.47%	3.30%	-7.47%	2.02%	1.49%	-11.60%	-1.28%	-1.36%
f3	2.28%	-1.73%	-6.44%	100.00%	9.42%	-3.55%	-3.29%	1.83%	6.55%	18.46%	17.70%	9.48%	4.46%	-6.49%	-0.24%	11.23%	-53.12%	-1.73%	30.74%	-5.76%	-8.01%	10.30%
f4	-2.98%	-2.20%	14.46%	9.42%	100.00%	11.66%	2.92%	-0.77%	-8.62%	4.78%	-8.29%	-20.66%	11.38%	23.03%	-1.38%	0.21%	7.36%	-4.88%	-2.10%	-3.07%	-0.12%	-4.66%
f5	3.75%	-10.42%	2.49%	-3.55%	11.66%	100.00%	0.20%	-8.33%	-5.22%	3.18%	23.47%	9.81%	8.14%	16.09%	-22.35%	-8.67%	5.68%	28.46%	-16.65%	-4.22%	-5.49%	6.41%
f6	-11.36%	-9.34%	-10.95%	-3.29%	2.92%	0.20%	100.00%	15.53%	12.24%	-2.84%	-14.07%	-5.42%	-33.95%	7.25%	5.24%	12.84%	-15.30%	0.79%	-28.85%	18.58%	-11.05%	11.80%
f7	-26.89%	-20.31%	-15.90%	1.83%	-0.77%	-8.33%	15.53%	100.00%	7.41%	5.37%	2.03%	-31.54%	-25.05%	-9.24%	-18.50%	-10.27%	-7.66%	-5.68%	15.72%	16.44%	-5.10%	0.46%
alpha2	7.95%	-3.13%	-0.39%	6.55%	-8.62%	-5.22%	12.24%	7.41%	100.00%	4.54%	-7.06%	-8.08%	-16.46%	-0.27%	-2.14%	22.02%	-13.64%	1.05%	-18.06%	1.78%	-2.20%	50.85%
0	39.43%	-2.54%	-25.04%	18.46%	4.78%	3.18%	-2.84%	5.37%	4.54%	100.00%	8.30%	-5.12%	-2.02%	-5.65%	1.31%	10.11%	2.58%	4.92%	10.84%	5.30%	-4.81%	-1.20%
1	-21.30%	-12.72%	8.61%	17.70%	-8.29%	23.47%	-14.07%	2.03%	-7.06%	8.30%	100.00%	-3.96%	-1.11%	3.29%	-9.33%	-13.68%	0.25%	-5.17%	9.53%	-18.81%	-8.55%	2.51%
2	42.61%	-1.78%	-16.59%	9.48%	-20.66%	9.81%	-5.42%	-31.54%	-8.08%	-5.12%	-3.96%	100.00%	-3.68%	-2.51%	-3.93%	4.04%	3.17%	2.57%	-1.89%	6.05%	-2.56%	5.62%
3	6.61%	10.78%	20.99%	4.46%	11.39%	8.14%	-33.95%	-25.05%	-16.46%	-2.02%	-1.11%	-3.68%	100.00%	6.97%	-2.79%	-6.32%	-2.76%	-3.73%	5.48%	-12.97%	3.49%	-11.47%
4	-16.83%	7.57%	10.74%	-6.49%	23.03%	16.09%	7.25%	-9.24%	-0.27%	-5.65%	3.29%	-2.51%	6.97%	100.00%	5.46%	4.81%	2.37%	2.38%	-8.14%	-9.70%	-3.10%	3.98%
5	-13.83%	3.09%	-23.47%	-0.24%	-1.38%	-22.35%	5.24%	-18.50%	-2.14%	1.31%	-9.33%	-3.93%	-2.79%	5.46%	100.00%	-0.98%	13.39%	-10.94%	-2.62%	-6.44%	-3.11%	-4.74%
6	27.68%	-13.44%	3.30%	11.23%	0.21%	-8.67%	12.84%	-10.27%	22.02%	10.11%	-13.68%	4.04%	-6.32%	4.81%	-0.98%	100.00%	17.27%	-0.05%	-4.93%	2.13%	-8.01%	5.61%
7	4.97%	-22.36%	-7.47%	-53.12%	7.36%	5.68%	-15.30%	-7.66%	-13.64%	2.58%	3.17%	-2.76%	2.37%	13.39%	-17.27%	100.00%	-3.44%	-22.61%	-4.89%	-0.50%	-21.07%	
8	8.44%	-15.39%	2.02%	-1.73%	-4.88%	28.46%	0.79%	-5.68%	1.05%	4.92%	-5.17%	2.57%	-3.73%	2.38%	-10.94%	-0.05%	-3.44%	100.00%	1.86%	6.46%	3.53%	5.71%
9	-2.79%	5.40%	1.49%	30.74%	-2.10%	-16.65%	-28.85%	15.72%	-18.06%	10.84%	9.53%	-1.89%	5.48%	-8.14%	-2.62%	-4.93%	-22.61%	1.86%	100.00%	2.93%	-3.38%	-8.40%
10	-1.92%	20.52%	-11.60%	-5.76%	-3.07%	-4.22%	18.58%	18.44%	1.78%	5.30%	-18.81%	6.05%	-12.97%	-9.70%	-6.44%	2.13%	-4.89%	6.46%	2.93%	100.00%	2.89%	-3.15%
11	9.91%	12.56%	-1.28%	-8.01%	-0.12%	-5.49%	-11.05%	-5.10%	-2.20%	-4.81%	-8.55%	-5.10%	3.49%	-3.10%	-3.11%	-0.81%	-0.50%	3.53%	-3.38%	2.89%	100.00%	-2.40%
alpha1	-0.21%	2.85%	-1.36%	10.30%	-4.66%	6.41%	11.80%	0.46%	50.85%	-1.20%	2.51%	5.62%	-11.47%	3.98%	-4.74%	5.61%	-21.07%	5.71%	-8.40%	-3.15%	-2.40%	100.00%

数据来源: wind、上交所、深交所、东方证券研究所

图 14: ABCM 模型数据集 1 上生成因子与 Barra 风险因子相关系数

	Size	Beta	Volatility	Liquidity	Value	Certainty	Cubic size	Trend	Growth	SOE
0	30.82%	-30.85%	-3.10%	-19.26%	19.14%	11.28%	6.09%	69.58%	13.13%	3.93%
1	3.00%	29.71%	7.27%	19.17%	12.71%	-16.27%	6.76%	14.55%	-13.78%	9.78%
2	73.55%	-4.84%	-3.71%	-14.60%	11.31%	32.99%	56.95%	3.27%	15.17%	7.30%
3	11.36%	11.09%	24.93%	-9.52%	13.74%	27.06%	14.13%	19.42%	9.65%	0.20%
4	-15.61%	-4.31%	28.07%	16.99%	-9.07%	-16.18%	-4.15%	-4.33%	1.75%	-2.35%
5	6.14%	-13.19%	-2.11%	1.32%	7.49%	-30.25%	-34.68%	-15.14%	-2.85%	9.41%
6	10.88%	-8.52%	-27.49%	-16.14%	26.49%	23.24%	-7.51%	-4.68%	28.28%	0.18%
7	-2.76%	42.43%	44.40%	34.35%	-43.88%	1.07%	-4.32%	28.35%	26.90%	-8.25%
8	13.06%	17.14%	18.42%	12.79%	1.95%	17.90%	6.56%	8.65%	16.90%	-6.69%
9	9.24%	-13.55%	-1.50%	-9.96%	9.06%	-16.36%	6.65%	20.92%	-15.88%	5.90%
10	-3.76%	-21.65%	-7.79%	-16.61%	-21.84%	-9.99%	-17.51%	-6.07%	-24.03%	-1.76%
11	5.97%	-0.92%	12.96%	-8.44%	-12.05%	10.29%	-17.30%	5.02%	3.06%	-1.34%
alpha	-8.76%	-2.57%	-25.10%	-11.33%	18.10%	7.58%	-15.57%	15.60%	-3.31%	-2.16%

数据来源: wind、上交所、深交所、东方证券研究所

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/508104036133007006>