

一. 因子处理方法

1. F-score打分法

将因子取值进行打分， >0 为1 <0 为0

每天计算一次，形成大因子。但也可以结合几天的表现统一算一下，这样可以减少调仓的频率

2. 标准化/归一化方法

z-score标准化、归一化、排序正态映射、softmax映射

3. 排序法

将因子排序后映射到-1到1上，也可以将上下的进行权重加和为一的操作。目前采取的是排序法。

4. 总结

因子处理可以剔除因子中多余的信息，在线性模型中更有助于简单的组合。虽然可能会丢失一部分信息，但是也在某种程度上缓解了过拟合。

二. 因子降维方法

1. 主成分分析法



知乎盐选 | 基于主成分分析的股票多因子量化投资策略研究

https://www.zhihu.com/market/paid_magazine/1329607856405159936/section/1329607952672837632?origin_label=search

将每个因子在时间尺度上做平均，得到每个因子每个股票上面的取值（当然也可以直接stack，但是可能计算量会比较大），然后进行主成分分析。

这样可以很好的进行降维，提取可能会有用的信息。如果一个主轴表现非常差，那就相当于从原始数据中剔除了这一部分表现非常差的数据。

问题在于无法进行风格调整与优化。因此可以分不同时间段进行这个操作，看主成分的变化规律，探究因子交互关系的变化。

2. 线性组合法

可以直接对因子进行等权重组合，这种方法比较简单，但是不能对风格差距很大的因子直接进行。

3. 逐步回归法

4. 与其他方法的结合

有的方法是利用IC筛选因子，一个个加入，直到加入的因子和已经有的因子相关性非常高。这相当于线性组合法（改良版）与逐步回归法的结合。

5. 总结

因子降维主要针对共线性非常强的因子。这些因子的存在可能会让模型不稳定，并且增加过拟合的可能。但是降维的过程中可能会丢失比较多的信息，因此设置过滤信息的标准是必要的，比如引入ICIR筛选，或者对主成分进行探究，剔除表现不良的主成分。

此外，降维本身也是对于因子库整体分析必要的手段，可以让我们更加了解因子的作用与分布。

三. 因子组合方法

0. 与因子降维的关系



多因子权重优化方法比较

在多因子量化投资体系中，具有稳定的预期收益，可解释的经济驱动理论，与其他因子的低…
<https://zhuanlan.zhihu.com/p/31753606>

很多因子表示了非常相似的信息，可以进行信息的提取。组合需要最大化每个因子的溢价，并尽量降低在某一个因子上面的暴露。降低维度的方法有等权重、PCA、逐步回归等等。

在降低维度之后，可以针对之后的因子进行进一步合成，如IC加权、IR加权、ICIR加权、最优化等等 $W^* = \sigma * \Sigma_{IC}^{-1} * IC$ 。

1. 风格因子暴露法：



多因子模型选股策略，是如何将多个因子联合在一起的？ - 知乎

引言之前两篇写了Barra的两个主要用途：组合业绩归因和风险预测，这篇写一写如何用Bar…
<https://www.zhihu.com/question/43439162/answer/2512744575>

作为逐渐转化为风险因子的Alpha因子，这些因子虽然时时刻刻有区分能力，但是方向可能会发生变化。如果方向长时间不变就可以看成Alpha因子。因此，可以（1）风险因子可以进行轮动，不断调整组合权重来进行优化；（2）加入更多Alpha因子提升解释能力

$$\begin{aligned} \max \quad & w' X_{obj_fct} \\ \text{s.t.} \quad & w' X_{oth_fct} = 0 \\ & w' \vec{1} = 1 \\ & w > 0 \end{aligned}$$

其中 X_{obj_fct} (n*1维)是所有股票在目标因子上的暴露，相当于让组合在目标因子上面暴露最大，在其他因子上面暴露最小。
结合barra对组合的风险预测，我们也可以把最小化组合风险的目的代入优化目标，具体形式如下：

$$\begin{aligned} \max \quad & w' X_{obj_fct} - w' V \quad \text{Q} \\ \text{s.t.} \quad & w' X_{oth_fct} = 0 \\ & w' \vec{1} = 1 \\ & w > 0 \end{aligned} \quad , \text{ 其中 } V(n*n \text{ 维}) \text{ 是 } n \text{ 只股票的方差-协方差矩阵}。$$

另外，如果是要做指数增强，在指数的成分股^Q中进行选股，那么优化的形式就会变成：

$$\begin{aligned} \max \quad & w' X_{obj_fct} - w' V \quad \text{Q} \\ \text{s.t.} \quad & w' X_{oth_fct} = w'_{index} X_{oth_fct} \quad , \text{ 其中 } w_{index} \text{ 是指数成分股 的权重矩阵。} \\ & w' \vec{1} = 1 \\ & w > 0 \end{aligned}$$

$$\begin{aligned} \min \quad & w' X_{resvol} \\ \text{s.t.} \quad & w' X_{oth_fct} = w_{zz_500} X_{oth_fct} \\ & w' \vec{1} = 1 \\ & w > 0.0005 \\ & w < 0.02 \end{aligned}$$

备注：由于做的是低波，所以组合在波动率因子上的暴露越小越好。

这种方法感觉比较适合指数增强或者控制一定风险的策略，可以用于最后对于风格因子中性化的调整。

行业中心化的意义？

2. 因子打分法

层次分析法、熵权法都属于自定义权重法。这种方法比较依赖于对因子的评价，只不过更加主观一些。

3. 因子特征权重法

主要考虑因子有效性、稳定性、相关性



【天风金工吴先兴团队·专题报告十九】半衰IC加权在多因子选股中的应用

【天风金工吴先兴团队·专题报告十九】半衰IC加权在多因子选股中的应用 2017-07-22 ...

https://www.sohu.com/a/159221447_619347

- (1) 等权重合成，不稳定，并且因子相关性较高
- (2) IC_IR加权合成：权重就是IR，考虑了稳定性
- (3) 最优ICIR加权合成：以组合IC最大化为目标，求出权重

权重向量W的解析解：协方差矩阵的逆与IC均值向量的乘积，

$$W = \Sigma^{-1}IC$$

权重向量的解析解

- (4) 半衰期IC加权IR加权

因子的预测能力是不断变化的，

假设序列长度为N，半衰期为H，半衰期权重wi：

$$\omega_i = \frac{2^{\frac{i-N-1}{H}}}{\sum_{t=1}^N 2^{-\frac{t}{H}}}$$

此处的wi是用于半衰期IC的权重。

之后将得到的半衰期IC用于因子加权，ICIR半衰期加权同理。

4. 模型法



量化多因子投资模型的三种常见建模方法

在我的上一篇文章中，我简单介绍了多因子投资模型需要回答的三个基本问题，可作为该篇...

<https://zhuanlan.zhihu.com/p/146219743>

(1) 线性模型

——FF的三因子模型：

通过时间序列的一元线性回归得到每只股票对每一个因子的敏感度：计算出SMB，每一只股票的月收益对SMB做回归，相当于一元线性回归，得到的就是因子敏感度。（时间区间比较长）结果是每一只股票在长期收益率有多少能被这个因子解释，是一个n*1的向量。

通过将同一时间点上的敏感度进行多元横截面线性回归，得到每个自变量系数：将每个月股票收益作为因变量，敏感度作为自变量，进行回归可以得到因子暴露beta。就是说在长期的敏感性基础上，目前这个因子暴露的收益有多少，是一个数值。相当于利用长期敏感性求每个时刻的因子收益（暴露）。

但是第二步的横截面回归其实还有一种理解，就是敏感度就是因子值，因此直接进行横截面的第

第二步，利用收益对因子值进行回归，就可以求出因子暴露，也就是beta，然后用它去进行预测。这个时候主要的目的在于预测。

(2) 改良的线性模型

——线性回归法存在一些问题（多重共线性），因此我们可以利用岭回归、Lasso、WLS等等。

——如果想要针对未来多天的收益，同时降低多重共线性的影响，可以考虑基于典型相关分析的PLS（主成分对主成分回归）

知

【量化策略】如何用PLS方法合成因子，实现一个年化41%的策略？

因本文涉及公式较多，建议大家前往优矿量化社区，获取完整代码 导读A. 研究目的： 本文…
<https://zhuanlan.zhihu.com/p/38368825>

(3) 机器学习模型

机器学习的主要优势在于信息的提取与降维

（机器学习梳理的文献）



链接

<https://dachxiu.chicagobooth.edu/download/ML.pdf>

《量化投资策略：多因子到人工智能》

非常全的一本书，包括目前主流的各种预测模型以及训练方法，附源码。

(4) 排序学习模型



链接

<https://www.doc88.com/p-51061700525415.html>

原创

排序学习（Learning to rank）综述_anshuai_aw1的博客-CSDN博客_排序学

Learning to rank(简写 LTR、L2R) 也叫排序学习，指的是机器学习中任何用于排序的技术…
https://blog.csdn.net/anshuai_aw1/article/details/86018105

过去是推荐系统和信息检索系统的组成部分，可以进行因子的去相关（独特的特征筛选机制），并且最大程度减弱过拟合问题（排序比较模糊，可以有效减弱传统模型对收益率完美拟合的追求）。因此可以尝试利用这种机器学习方法。（排序学习只是一种思想，在Xgb, Lgbm, 深度模型中都可以存在）

第一个链接是一篇硕士论文。这篇论文中采取滚动60天预测后5天复利的方法（P32）。在每个时刻利用IR进行因子筛选，具体的流程就是根据IR从大到小逐渐加入因子，直到新加入的因子相关性与存在的超过0.5。

5. 整体票池筛选法

知

混合”还是“整合”——多因子组合的构建

前言与前几篇类似地，笔者翻译了一篇来自海外的网络研讨会议纪要，较为简洁地为各位读…

<https://zhuanlan.zhihu.com/p/66652071>

整合代表传统的多因子策略，不好解释不太明朗，但是对于证券集中度较低，可能需要降低维度的时候，整合可能更加方便一些。

混合代表排序筛选，比较清晰透明，时候证券集中度较高，因子数量较少的时候采用，适合进行风格调整，smart beta策略。这里比较常用的有：（1）每个因子构件组合之后取并集；（2）按照从属关系进行（比如在某一个部分用一部分因子，或者长期用一个因子，短期调仓用另一个）；（3）因子打分，如果因子之间存在concensus就多买一些，否则少买一些。

6. 总结

因子组合的核心在于预测，因此可以引入主观的知识，或者利用模型为我们进行学习。但是，不同的方法对应不同的策略特征。可能需要针对策略的需要选择更加合适的模型或者优化目标。

四. 样本训练方法

1. 训练与测试样本的切割

如果需要训练模型，则会涉及到参数的不断变化。一般不会依赖一套参数使用很长的时间。因此，在每一个时间窗口，都要进行因子筛选、模型训练（交叉训练）等，更新组合权重。因此，一般采取滑动窗口进行训练。训练窗口的长度，调整模型的频率需要进行确定。这样，其实得到的是一个单因子策略。上文提到的不同模型会对应不同的投资风格，其实可以看作这个学习出的大因子的回测特征。

如果不用训练模型，直接生成权重（如特征权重、打分等），则可以生成一个大因子，也得到了单因子策略。

2. 对于训练模型的思考

（1）多数文献都没有提到划分不同样本进行训练和学习。有一些因子在某些股票上明显有更好的表现。在这个时候，一方面可以把他们挑出来，对这一部分样本建立模型。也可以将因子进行调整，让其只在我们需要的部分体现分类特征。

（2）对于过拟合问题，在模型中可能需要剔除一些过拟合现象严重的因子。这个时候可以采取遗传算法等复杂度稍低的优化算法探索最佳的因子组合。但是这样非常费时费力。因此，从因子本身表现评价方面入手可能会更好一些。

