

指数增强策略介绍

2021.3

(仅供内部使用)

林翊





目录

- 部分因子介绍
- 数据处理
- 打分流程介绍
- 分层抽样流程介绍
- 回测详情与设定
- 区间收益特征

我们的投资观

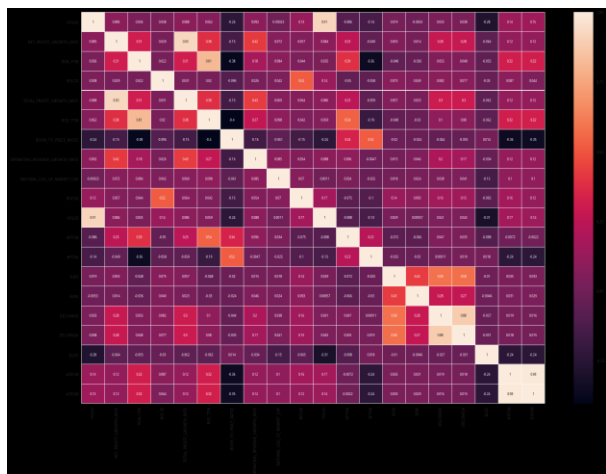
- 关于主动与被动投资：主动投资与被动投资对应，主观投资与量化投资对应。主动投资既可以采用主观方法也可以采用量化方法，而被动投资多采用量化方法。从发达市场的经验来看，采用主观方法的主动投资长期跑赢指数的概率不高，主要受人类片面的认知和情感偏差等因素拖累。因此经过多年发展，发达国家权益投资目前以指数投资和量化投资为主（大类配置不在谈论之列，如有兴趣请移步我们开发的全球大类配置策略）。
- 指数ETF与指数增强：指数ETF产品市场接受度高、流动性好、跟踪误差小，但收益相较于指数增强策略偏低。指数增强产品意在适当放宽跟踪误差的要求的同时来获取稳定的超额回报。
- 为什么国内现阶段可以“增强”：纵观全球ETF市场可以发现，越成熟的市场指数增强产品的alpha越低。因为成熟市场机构化程度高、投机相对较少，因此很多绩差股票的流动性较差，而充足流动性为量化策略有效执行提供了土壤。国内市场目前投机者提供的流动性充足，因此中短期内指数增强产品能够提供较好的alpha，但长期看这种超额会越来越薄。
- 指数增强产品用途：指数增强产品可作为很多策略的基石使用，比如指增+对冲（中性策略）；指增+新股申购；指增+信用债（不同的配置比例对应不同的风险收益特征）；指增+可转债等等。

模型使用的部分因子

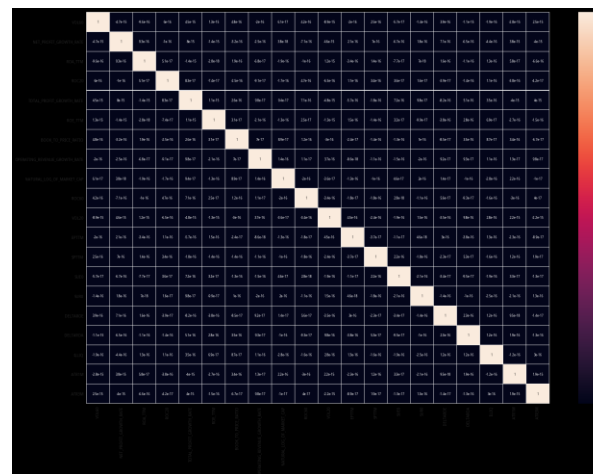
- 模型使用的因子主要有七大类二十余个：规模因子、估值因子、成长因子、盈利因子、流动性因子、波动率因子和技术类因子。
- 规模因子：对市值采用对数处理，得出的涨幅更客观；
- 估值因子：P/B、P/E等；
- 成长因子：净利润、收入的增长率；
- 盈利因子：ROE、ROA等；
- 波动率因子：60日波动率；特质波动率（剔除Fama-French三因子预测收益的部分）等；
- 技术因子：过去N日的涨幅，利用其反转效应。

基础数据处理

- 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，如果行业缺失则予以删除。
- 去极值：极值数据会影响拟合效果，对后续预测带来偏差。本策略采用MAD（Median Absolute Deviation 绝对中位数法）去极值，3倍sigma范围内数据保留。大于3倍sigma的数据，按照其顺序，均匀插值到 3-3.5 倍sigma范围内。
- 标准化：各因子量纲差别很大，为能够使因子之间能够比较，我们需要对每个因子进行标准化处理。本策略采用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。
- 市值和行业中性化：不同行业及市值对因子的影响较大，为了使因子在同一量纲上比较，需要对市值和行业进行中性化处理。处理后的因子值=原始因子值- β_{tai} *对数市值- β_{taj} *行业哑变量。之后再对处理后的因子值进行去极值和标准化处理。
- 因子多重共线性处理：多重共线性会导致参数估计方差较大，即估计不准确。为解决此问题需要对因子进行正交化处理，本策略采用“对称正交”处理，相较“施密特正交”对因子更加公平。



正交化前因子
相关性



正交化后因子
相关性

股票得分及因子权重（参数估计）流程

一、计算因子的IC：IC可简单理解为 $t-1$ 期的因子值与 t 期收益率的相关系数。但考虑到不同股票的涨跌幅有其自身特点（小盘股弹性大、大盘股弹性小），因此本策略选择秩相关，即 $t-1$ 期的因子值排名与 t 期收益率排名的相关系数。

二、计算因子IR：计算时点前推12期计算IR（过去12期平均IC/IC的标淮差）

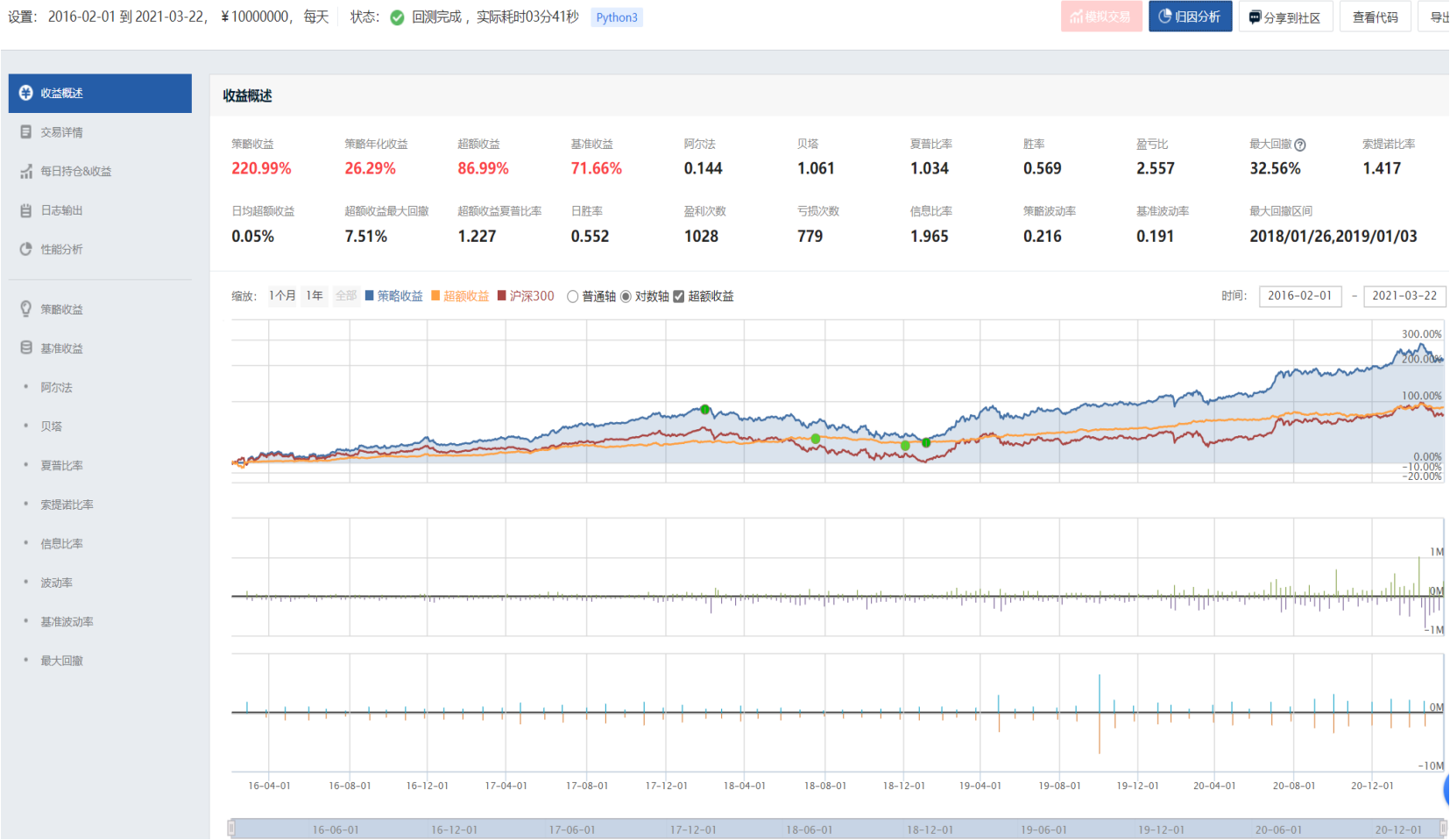
三、计算因子权重：每个因子IR/本期IR之和

四、计算每只股票的得分： $\text{sum}(\text{每只股票因子值} * \text{因子权重})$ ，备用。

分层抽样

- 指数增强策略的基础是指数，跟踪误差不能过大，所以需要尽可能的复制指数最重要的因子。为避免备选股票的行业过度集中，我们在每个行业里分成两组，每组中取股票得分最高的标的进行配置。每只股票权重为该分组中所有股票的市值权重。行业内的分组标准为随机选股并按市值等分。

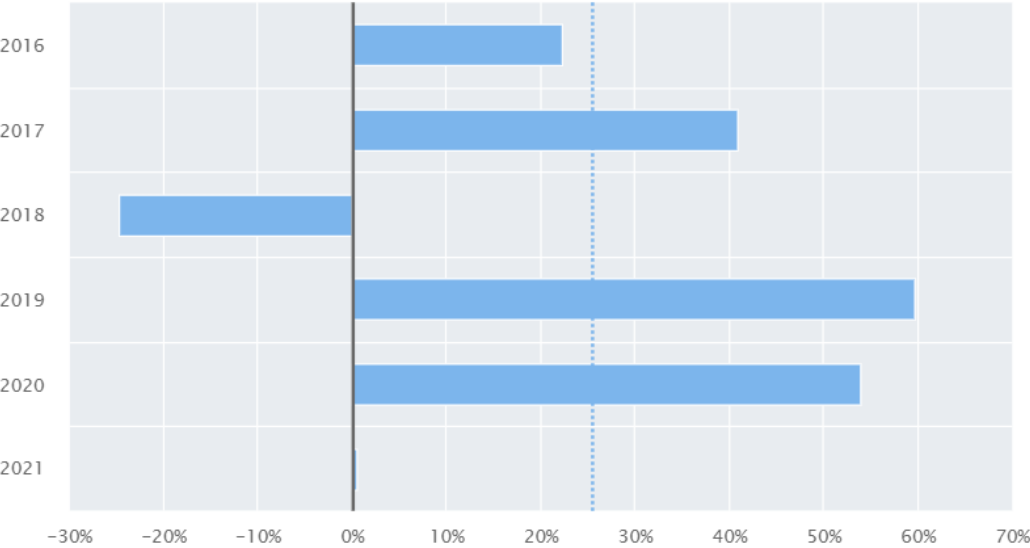
回测详情与设定



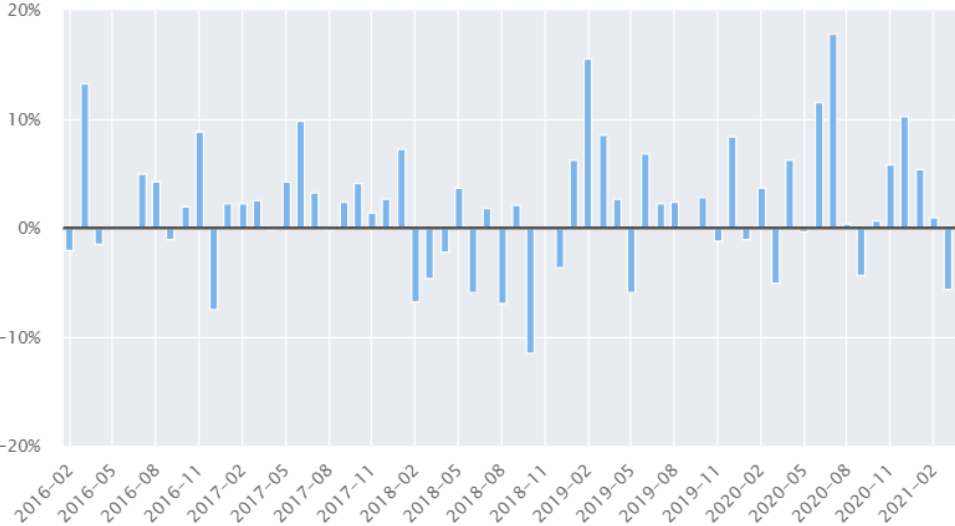
- 回测设定: 过滤st; 过滤上市时间不够三个月的股票; 过滤流动性较差即月度交易天数不足的股票; 考虑佣金, 未考虑冲击成本
- 超额收益比较稳健: 超额收益五年年化14.4%, 超额收益最大回撤7.51%

区间收益特征

年度收益



月度收益的时间序列



月度胜率72.1%