跳跃因子系列研报复现-下

一、概述

在上篇中，文章参考现有的文献和研报资料，汇总了股票价格跳跃的主要研究架构，以及直接使用跳跃波动和跳跃检验统计量相关的因子定义，并对所构建的单因子效果进行了简单的周度测试。

本文将沿用上篇介绍的跳跃因子检验框架，进一步介绍跳跃识别与跳跃收益结合、使用跳跃信息对振幅进行改进，从而构建的选股因子，并使用LGBM模型对两篇文章涉及的跳跃因子，进行简单合成，提高因子整体表现。

二、跳跃因子构建

1.累计跳跃收益类因子

构建累计跳跃收益类因子，该类因子将存在显著（正向/负向）跳跃交易日的当日收益累加，表征信息冲击在过去一段时间中对股价造成的累计变化效果。因子定义中涉及的跳跃检验示性变量定义参考上篇，合计构建12个因子。

|  |  |  |
| --- | --- | --- |
| 因子名 | 因子描述 | 因子定义 |
| JR | 观察期内累积跳跃收益 |  |
| JR+ | 观察期内累积正向跳跃收益 |  |
| JR- | 观察期内累积负向跳跃收益 |  |
| JAR | 观察期内累积绝对跳跃收益 |  |
| JAR+ | 观察期内累积绝对正向跳跃收益 |  |
| JAR- | 观察期内累积绝对负向跳跃收益 |  |

2..日内跳跃收益

日内跳跃收益相关因子的构建，涉及日内跳跃的识别，结合上篇中提及的跳跃检验统计量𝑇𝐽𝑂,𝑡，识别日内跳跃的步骤如下：

（1）给定单只股票的日内5min频简单收益率{𝑅1,𝑅2, . . . ,𝑅𝑁}和对数收益率 {𝑟1, 𝑟2, . . . , 𝑟𝑁}，计算跳跃检验统计量𝑇𝐽𝑂。若不拒绝原假设：则当日不存在价格跳跃，则当日跳跃收益为0， 并结束检验；若拒绝原假设，则将结果记录为𝑇𝐽𝑂\_0；

（2）将初始简单收益序列和对数收益序列中每个值依次替换为相应序列的中位数， 并基于替换后的序列计算跳跃检验统计量，则可得𝑇𝐽𝑂\_i (𝑖 = 1, . . . , 𝑁)；

（3）计算统计量差值| 𝑇𝐽𝑂\_0| − |𝑇𝐽𝑂\_i|(𝑖 = 1. . . 𝑇)，比较替换某一时段收益率后的差异。根据似然比检验的想法，差值越大则对拒绝原假设的贡献越大。因此，将差值中最大值对应的时段识别为发生跳跃；

（4）将已识别价格跳跃时段的简单和对数收益率替换为相应序列的中位数，并重复步骤（1）-（4），直至𝑇𝐽𝑂\_0不拒绝原假设；

计算过程中，加入隔夜收益，并视其为一个单独的时间间隔，因此，每个交易日有49个时间间隔（N=49）。跳跃检验的置信度取5%。

根据识别出的日内跳跃时刻信息，结合隔夜收益和日内5分钟频收益率，并取过去20个交易日均值，构建合计13个日内跳跃收益因子。

|  |  |
| --- | --- |
| 因子名 | 因子描述 |
| jump | 过去20个交易日的跳跃收益之和 |
| no\_jump | 过去20个交易日的非跳跃收益之和 |
| pos\_jump | 过去20个交易日的正跳跃收益之和 |
| neg\_jump | 过去20个交易日的负跳跃收益之和 |
| co\_jump | 过去20个交易日的隔夜跳跃收益之和 |
| oc\_jump | 过去20个交易日的日内跳跃收益之和 |
| co\_pos\_jump | 过去20个交易日的隔夜正跳跃收益之和 |
| oc\_pos\_jump | 过去20个交易日的日内正跳跃收益之和 |
| co\_neg\_jump | 过去20个交易日的隔夜负跳跃收益之和 |
| oc\_neg\_jump | 过去20个交易日的日内负跳跃收益之和 |
| oc\_pos\_jump\_std | 过去20个交易日的日内正跳跃收益的标准差 |
| oc\_pos\_jump\_avg | 过去20个交易日的日内正跳跃收益的指数加权平均收益 |
| oc\_pos\_jumpt\_avg | 过去20个交易日每日日内正向跳跃收益乘上当日的换手率，再取指数加权平均收益 |

3.振幅改进因子-飞蛾扑火

传统振幅因子，衡量了一天之内股价变化最大的幅度。但振幅仅仅刻画了价格变动的幅度，却对变化的过程未加区分，这也导致了传统振幅因子的效果不甚理想。。使用股票的日内跳跃信息对传统的振幅因子进行改进，分别构建跳跃度因子和修正振幅因子，并进行等权合成，构建“飞蛾扑火”复合因子。

3.1.跳跃度因子：

（1）给定单只股票的日内min频简单收益率{𝑅1,𝑅2, . . . ,𝑅𝑁}和对数收益率 {𝑟1, 𝑟2, . . . , 𝑟𝑁}，计算每个时刻t的泰勒残项：

根据泰勒展开式，假设x对应股票的min对数收益率，则对应min简单收益率，则泰勒残项2等于2\*(min简单收益率-min对数收益率) – min对数收益率^2；

（2）对当日所有时刻的泰勒残项取时序均值，作为这一天个股股价跃程度的代理变量，记为“日跳跃度”因子，计算过去20个交易日的均值和标准差并等权求和，记为“月跳跃度”因子。

3.2.修正振幅因子：

基于前文对股价跳跃的论述，可以认为实际上吸引博彩偏好型投资者的，并不是“波动”而是“跳跃”，即波动大的个股，包含“跳跃”的可能性更大一些，因此表现出了一定的有效性。所以要增强改进包含波动信息的因子，一个行之有效的途径就是识别出波动中的跳跃。

因此，依据振幅能否吸引到博彩偏好型投资者（飞蛾），将振幅分为两类：“火把”型振幅和“太阳”型振幅。具体而言，如果 t 日股价发生了明显跳跃，则认为这一天可以吸引到博彩偏好型投资者，然而投资者追逐这种类型股票犹如飞蛾扑向“火把”，大概率将蒙受损失。反之，如果 t 日股价没有发生明显跳跃，则认为这一天不能吸引到博彩偏好型投资者，将其比作“太阳”型振幅，对于博彩偏好型投资者来说，虽然明亮，却不会去追逐它。

3.2.1修正振幅因子1：

（1）计算每天的“日跳跃度”因子及振幅。振幅为(t日最高价 - t日最低价)/(t-1日收盘价)；

（2）每天计算“日跳跃度”因子的截面均值，认为“日跳跃度”小于截面均值的股票，属于未发生跳跃或跳跃程度较小的股票，这一天的振幅为“太阳”型振幅，将其振幅值乘以-1；反之如果“日跳跃度”大于截面均值，则认为这种股票属于发生跳跃或跳跃程度较大的股票，这一天的振幅为“火把”型振幅，取值不变。记变化后的振幅为“翻转振幅1”因子，并取20日均值。

3.2.2修正振幅因子2：

（1）使用t-1日的最低价与 t 日的最高价，计算当日股票的“泰勒残项”；

（2）每天计算“泰勒残项”的截面均值，认为“泰勒残项”小于截面均值的股票，属于未发生跳跃或跳跃程度较小的股票，这一天的振幅为“太阳”型振幅，将其振幅值乘以-1；反之如果“泰勒残项”大于截面均值，则认为这种股票属于发生跳跃或跳跃程度较大的股票，这一天的振幅为“火把”型振幅，取值不变。记变化后的振幅为“翻转振幅2”， 并取20日均值。

将2个修正后的振幅因子等权合成，获得“修正振幅”因子。

3.3飞蛾扑火因子

将“月跳跃度”因子和“修正振幅”因子等权合成，得到“飞蛾扑火”因子（moth\_to\_fire）。

三、单因子测试结果汇总

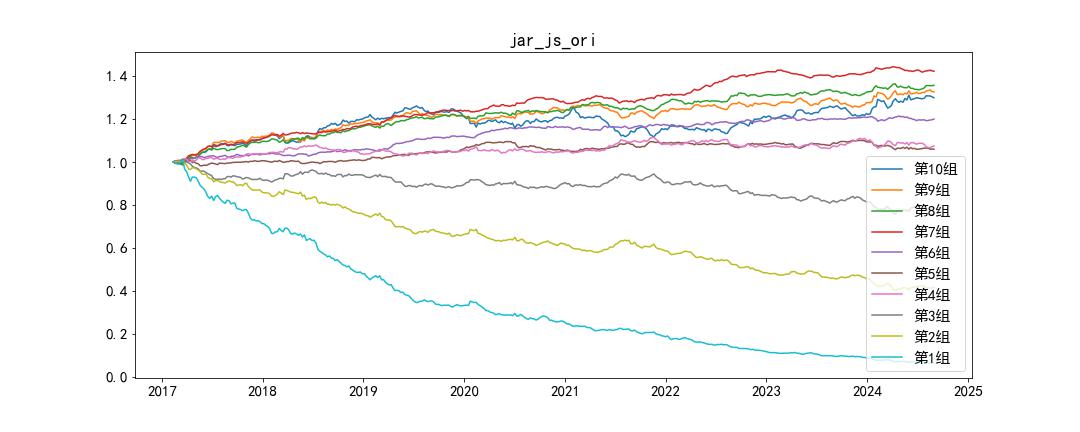
本文计算因子的日频数据，并进行周频回测，股票池为全部A股，回测起止日期为2017.2-2024.8。

受限于篇幅，本文仅列出前文涉及因子的全样本区间回测结果，各类别因子仅选取表现较好的因子进行展示，仅供参考。因子构建完整代码见知识星球，欲获取全部因子的回测统计图表，请联系本文作者。

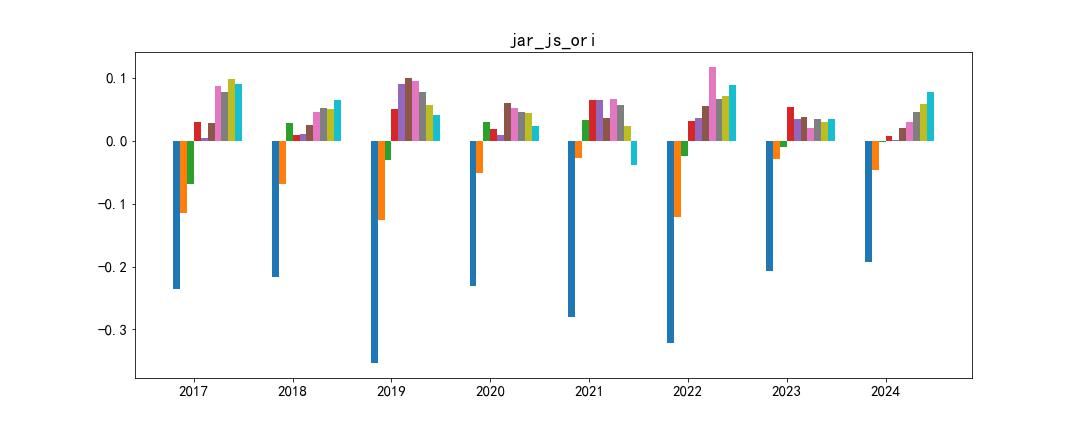
1. 累计跳跃收益类因子

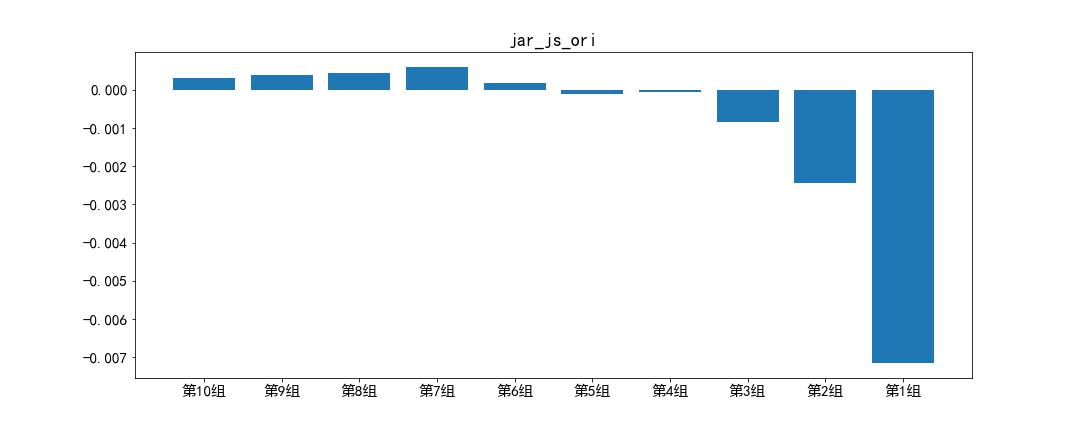
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 因子名称 | t值绝对值平均值 | IC平均值 | IRIC | IC>0概率 |
| jr\_bns | 3.384 | -0.033 | -0.558 | 27.72% |
| jr\_pos\_bns | 4.384 | -0.040 | -0.557 | 29.02% |
| jr\_neg\_bns | 3.214 | 0.006 | 0.086 | 55.70% |
| jr\_js | 6.042 | -0.066 | -0.617 | 26.17% |
| jr\_pos\_js | 6.362 | -0.071 | -0.623 | 26.17% |
| jr\_neg\_js | 3.152 | 0.024 | 0.454 | 71.50% |
| jar\_bns | 4.682 | -0.039 | -0.485 | 31.09% |
| jar\_pos\_bns | 4.384 | -0.040 | -0.557 | 29.02% |
| jar\_neg\_bns | 3.207 | -0.006 | -0.086 | 44.30% |
| jar\_js | 6.414 | -0.073 | -0.626 | 26.94% |
| jar\_pos\_js | 6.362 | -0.071 | -0.623 | 26.17% |
| jar\_neg\_js | 3.125 | -0.024 | -0.454 | 28.50% |

（1）jar\_js因子分10组回测结果



（2）jar\_js因子10分组分年表现



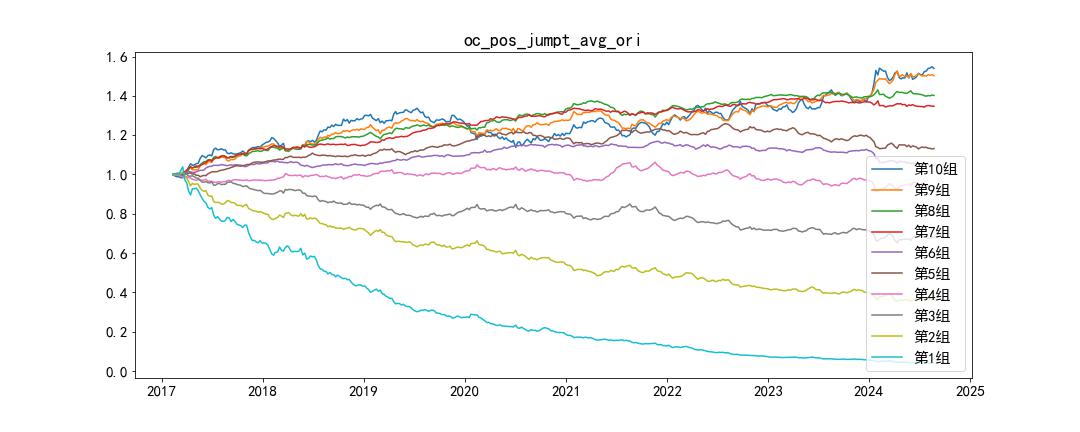


以回测区间内效果较好的因子过去20天累积绝对跳跃收益jar\_js为例，周度换仓的RankIC达-0.073，年化RankICIR为-4.51，IC胜率（<0的比例）达72%，样本内分层单调性尚可，且体现出显著的空头负向超额获取能力。

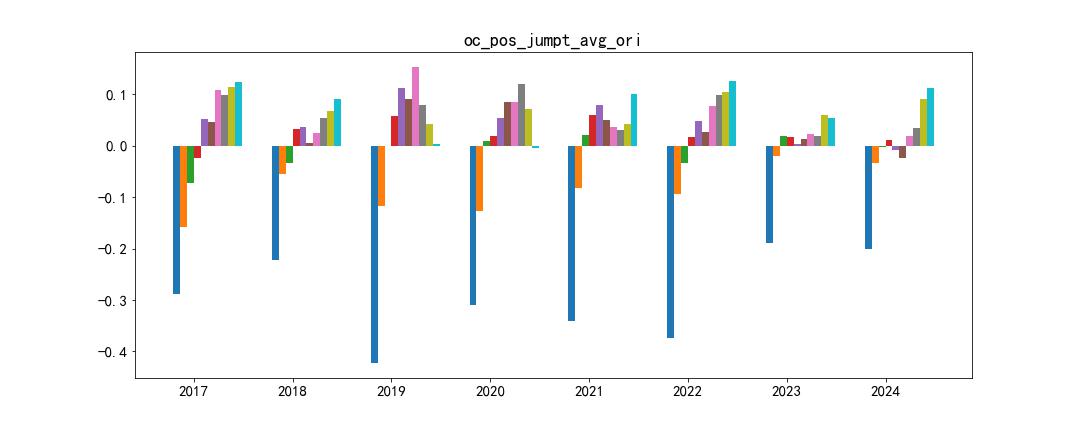
2. 日内跳跃收益类因子

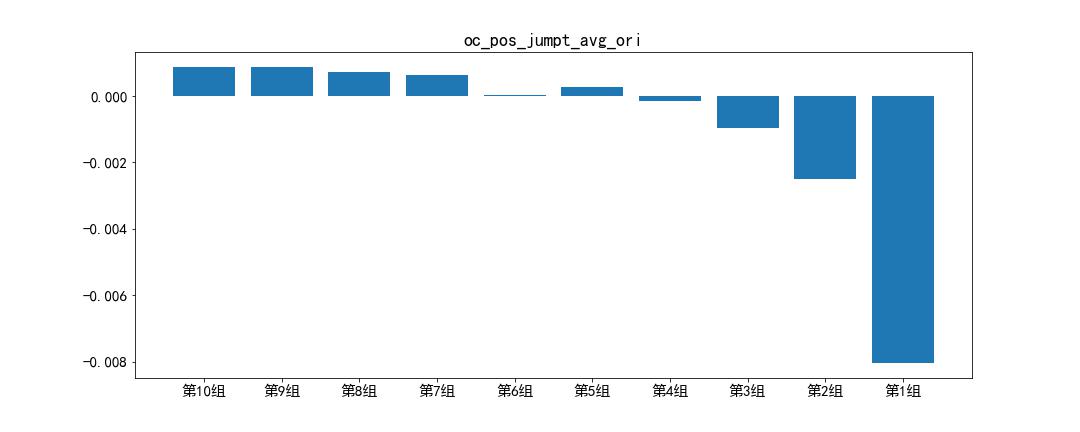
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 因子名称 | t值绝对值平均值 | IC平均值 | IRIC | IC>0概率 |
| jump | 4.773 | -0.054 | -0.642 | 25.65% |
| no\_jump | 6.270 | -0.003 | -0.026 | 49.22% |
| pos\_jump | 6.668 | -0.076 | -0.632 | 26.17% |
| neg\_jump | 5.771 | 0.047 | 0.394 | 67.88% |
| co\_jump | 2.824 | 0.008 | 0.130 | 56.48% |
| oc\_jump | 5.065 | -0.063 | -0.722 | 22.02% |
| co\_pos\_jump | 4.079 | -0.022 | -0.321 | 38.08% |
| oc\_pos\_jump | 6.428 | -0.074 | -0.622 | 26.17% |
| co\_neg\_jump | 4.482 | 0.037 | 0.438 | 69.17% |
| oc\_neg\_jump | 5.215 | 0.031 | 0.276 | 63.21% |
| oc\_pos\_jump\_std | 6.530 | -0.077 | -0.654 | 25.65% |
| oc\_pos\_jump\_avg | 6.365 | -0.077 | -0.654 | 26.23% |
| oc\_pos\_jumpt\_avg | 7.696 | -0.081 | -0.528 | 29.61% |

（1）oc\_pos\_jumpt\_avg因子分10组回测结果



（2）oc\_pos\_jumpt\_avg因子10分组分年表现



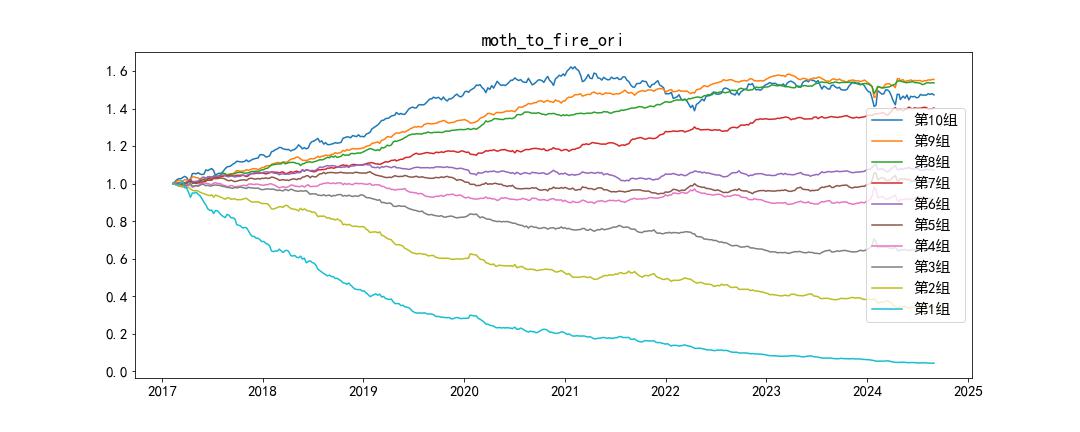


日内跳跃收益类因子中，oc\_pos\_jumpt\_avg表现较好，该因子周度换仓的RankIC达-0.081，年化RankICIR为-3.81，IC胜率（<0的比例）约70%，样本内分层单调性整体较好，且同样体现出显著的空头负向超额获取能力。

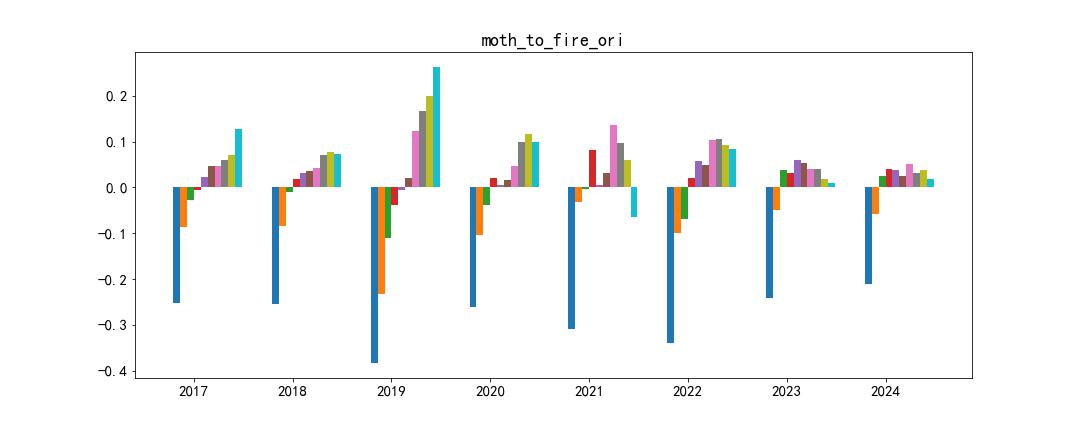
3.振幅改进因子-飞蛾扑火

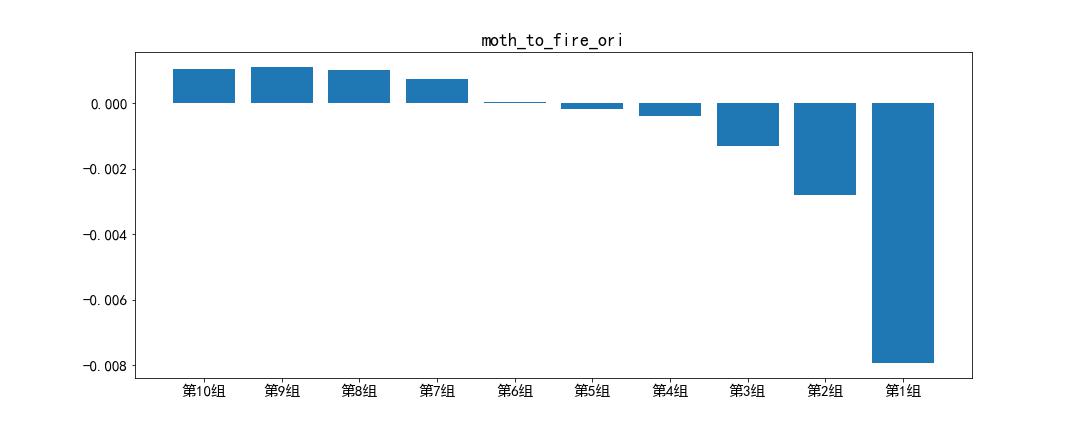
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 因子名称 | 统计区间 | t值绝对值平均值 | IC平均值 | IRIC | IC>0概率 |
| moth\_to\_fire | 2017 | 4.487 | -0.076 | -0.819 | 23.91% |
| 2018 | 5.024 | -0.080 | -0.985 | 15.69% |
| 2019 | 5.311 | -0.101 | -1.388 | 3.92% |
| 2020 | 5.820 | -0.073 | -0.839 | 21.15% |
| 2021 | 7.312 | -0.058 | -0.594 | 32.69% |
| 2022 | 6.397 | -0.081 | -0.994 | 20.00% |
| 2023 | 6.103 | -0.054 | -0.631 | 32.00% |
| 2024 | 6.241 | -0.051 | -0.449 | 32.35% |
| 总计 | 5.838 | -0.072 | -0.807 | 22.28% |

（1）moth\_to\_fire因子分10组回测结果



（2）moth\_to\_fire因子10分组分年表现

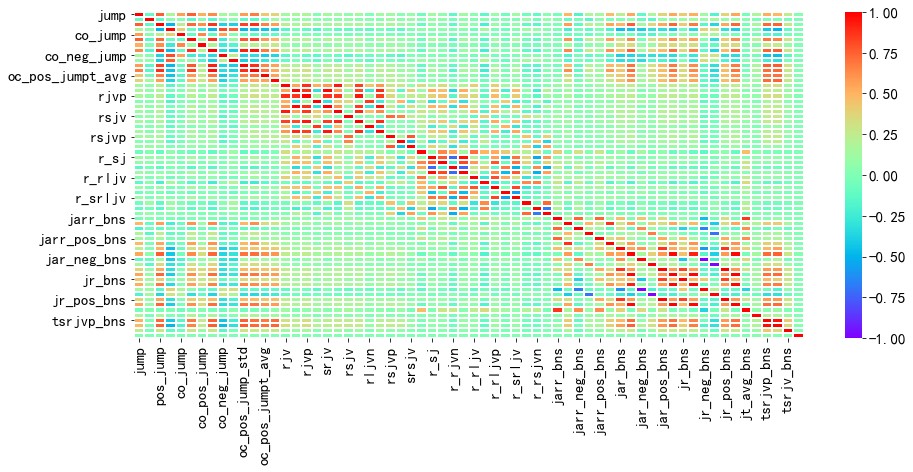




飞蛾扑火因子（moth\_to\_fire）全区间内的表现来看，周度RankIC达-0.072，年化RankICIR为-5.82，总体分层单调性良好，近年表现有所衰减。

四、跳跃系列因子合成

1.相关性分析



根据全部64个跳跃因子的相关系数分布情况，可以看出：

（1）同一类别的系列因子之间，因定义较为接近，存在较高的相关关联；

（2）日内跳跃收益因子与日频累积跳跃收益因子之间，同样存在一定的正相关关联；

（3）跳跃收益类因子与跳跃波动类因子之间的相关性相对较低。

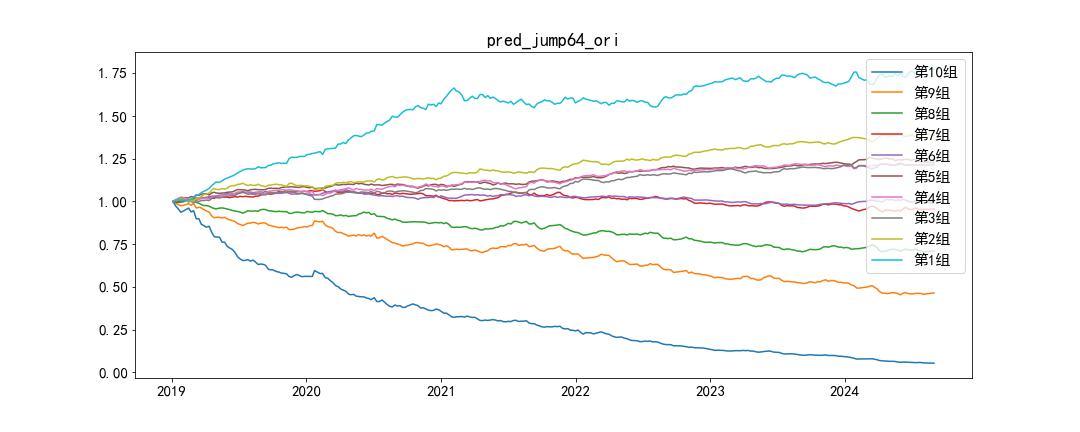
因而在后续的合成步骤中，考虑采用对因子间相关性要求更低的非线性合成方式，对全部跳跃因子进行复合。

2.因子复合

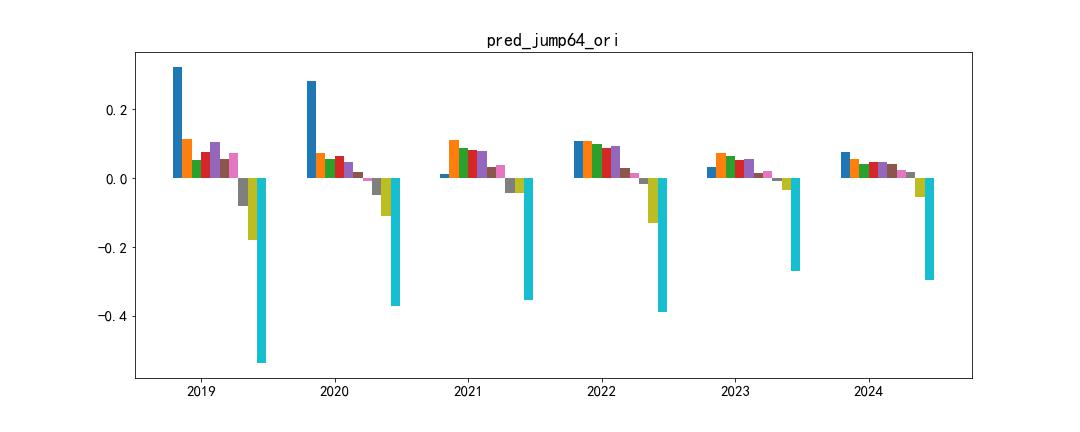
使用LGBM模型，以pearsonr的相反数作为训练目标，进行逐年滚动合成，复合因子的周度测试结果见下。

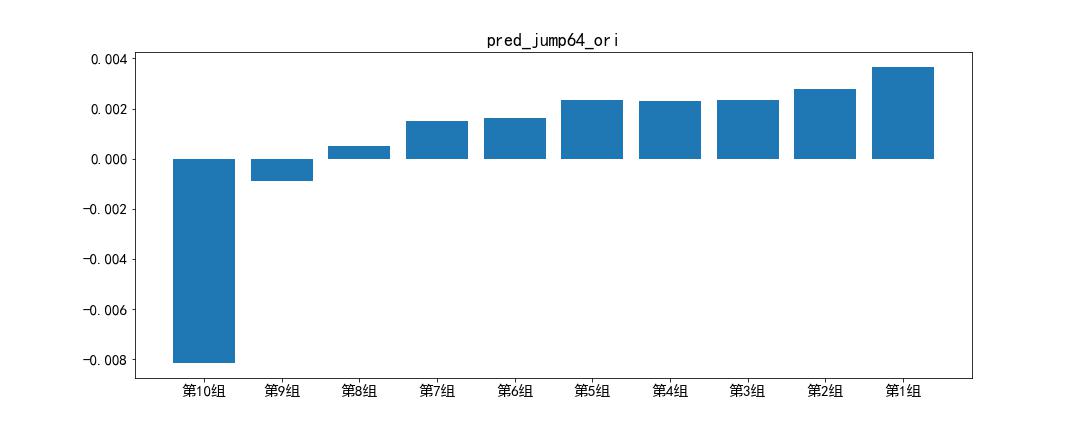
|  |  |
| --- | --- |
| 因子名称 | Pred\_jump64 |
| t值绝对值平均值 | 7.308 |
| IC平均值 | 0.092 |
| 年化IRIC | 5.976 |
| IC胜率 | 78% |

（1）复合因子分10组回测结果



（2）复合因子10分组分年表现





可以看出，经过合成后的复合因子，单因子的IC、稳定性及分组单调性都有了一定的提升。

五、总结

因篇幅较长，跳跃因子复现的全部内容分为上下两篇。上篇参考现有的文献和研报资料，汇总了股票价格跳跃的主要研究架构，以及直接使用跳跃波动和跳跃检验统计量相关的因子定义，并对所构建的单因子效果进行了简单的周度测试。

下篇进一步介绍了跳跃识别与跳跃收益结合、使用跳跃信息对振幅进行改进，从而构建的因子，并采用LGBM模型对两篇文章中涉及的跳跃因子，进行简单合成，从而提高因子整体表现。

股价跳跃模型已有不少相关的学术文献及研究报告可供借鉴，基于其逻辑构建的选股因子也体现出一定的筛选能力，通过对基于不同的切入点构建的跳跃系列因子进行合成，最终的因子表现相比单因子有一定的提升。

参考资料：

1. 《20220831-广发证券-金融工程专题-基于股价跳跃模型的因子研究-高频数据因子研究系列九》

2.《20221209-广发证券-金融工程专题-再谈股价跳跃因子研究-多因子Alpha系列报告之(四十六)》

3.《20220922-方正证券-多因子选股系列研究之六：个股股价跳跃及其对振幅因子的改进》

4.《20240725-招商证券-金融工程研究——如何识别股价的跳跃？》