

Aglv-Alpha

aglv

2025 年 3 月 16 日

1 数据处理

尝试在数据段提升效果:

1.1 价量数据日内分时段

将日内加量数据分成不同时段:9:30-10:00、10:00-收盘等等....

尝试将日内特殊时段的收益组合提升信息效果

ideal1: 组合上午下午收益,

$$ret = ret1 - ret2, ret = ret1 * ret2, ret = ret1 / ret2.....$$

1.2 数据降维

对数据降维, 考虑常用的机器学习的降维方法?pca、kmeans

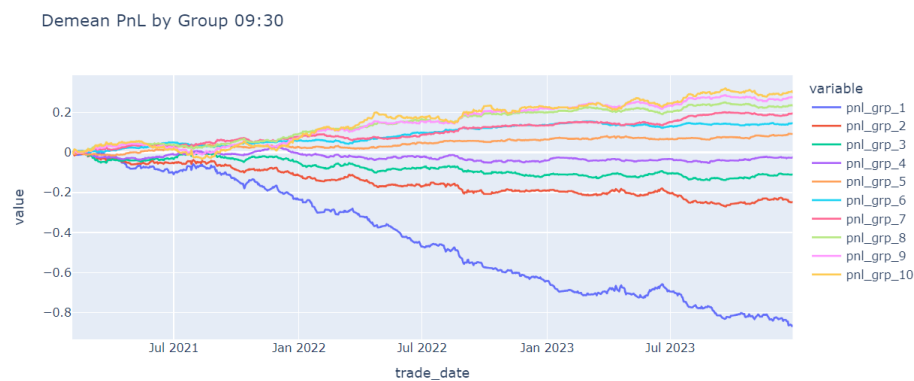
1.3 选股池过滤

以某个指标或者其他风格因子来实现对某只股票的因子

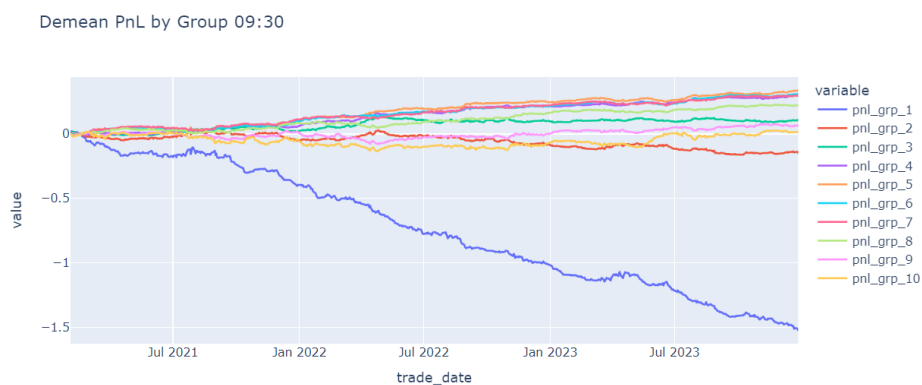
2 数据处理尝试

2.1 尝试使用风格因子

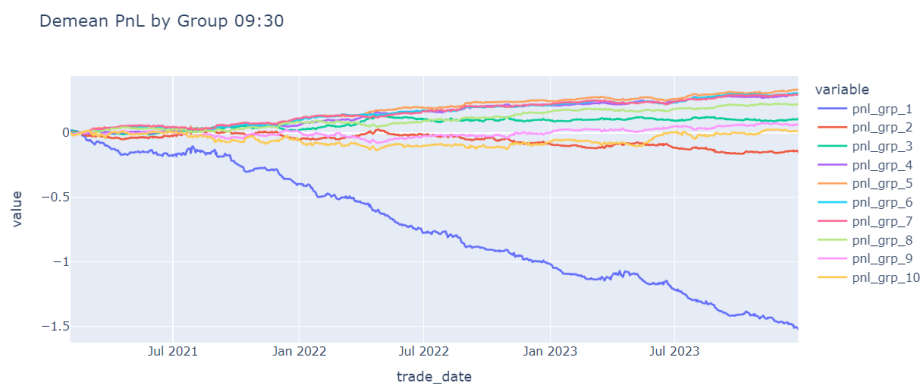
使用 ret20 风格因子对隔夜收益率下的 CVaR 因子进行增强: 对于动量类因子 cvar: 没有进行风格以因子加权:



使用 ret20 风格因子加权后:

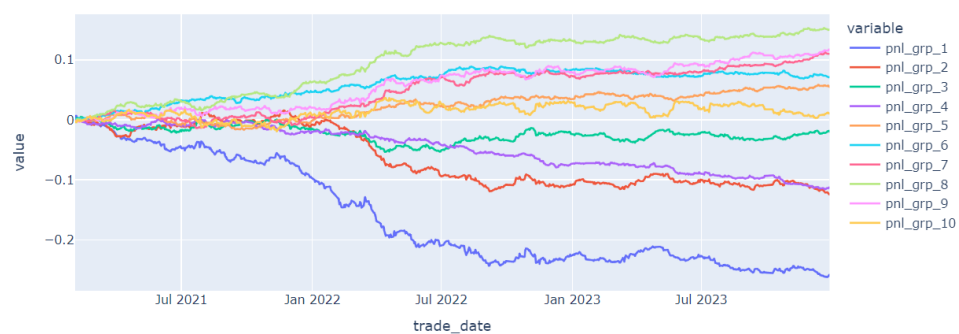


使用 vol20 风格因子加权后:

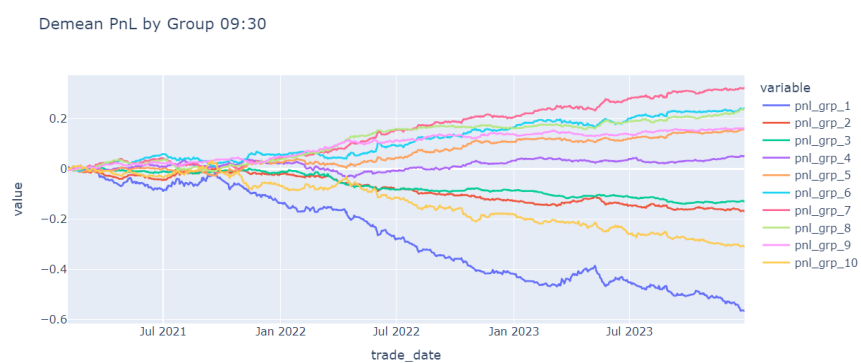


尝试用 ret20 因子对 TurnStd 因子进行加权, 效果几乎么有提升

对于尾部分布型因子 Asymp:



使用 Vol20 因子加权后回测效果变稳定:



组别分隔明显但是顺序不清晰?

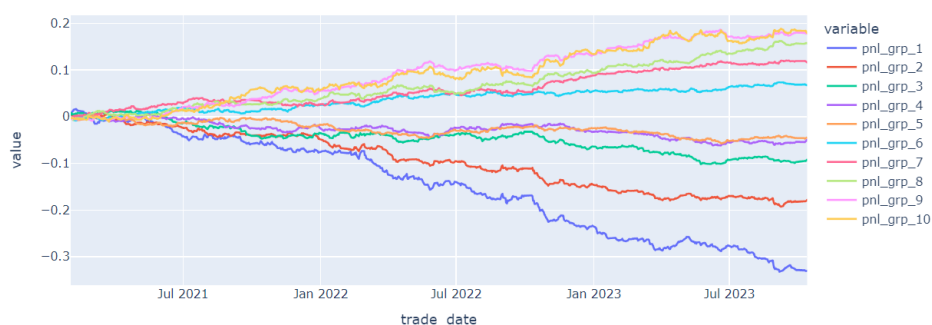
3 早盘午盘分歧因子

配合换手率数据描述单只股票的分歧程度

$MornRet30$ = 上午开盘 30min 的收益率

$AfnoRet30$ = 下午收盘前 3min 的收益率

$factor = MornRet30.rolling(20).corr(AfnoRet30)$



结果: 还行

早上开盘跳跃收益和午间开盘跳跃收益:

$factor = corr(\text{隔夜跳跃收益}, \text{午间跳跃收益})$

结果: 基本没用

4 早盘预期因子

idea: 单只股票头一日的涨跌情况 (or 隔夜收益 or 尾盘收益) 可能代表了对第二日的预期, 从开坡价可以反应实际情况对预期的回应程度

结果: 试过几个, 都不太行

5 基于成交单的因子

5.1 大卖单大卖单以及其差值因子

5.2 成交集中度因子

对数数据调整?

factor1: 除以 rolling 一段一段的标准差

factor2: rolling 一段时间的最大值除以 rolling 一段时间的标准差

factor3: rolling 一段时间的最小值除以 roling 一段时间的标准差

$$\text{大单买入强度}_{i,t} = \frac{\text{mean}(\text{大买单成交额}_{i,t-j})}{\text{std}(\text{大买单成交额}_{i,t-j})}$$

$$\text{大单净买入强度}_{i,t} = \frac{\text{mean}(\text{大买单成交额}_{i,t-j} - \text{大卖单成交额}_{i,t-j})}{\text{std}(\text{大买单成交额}_{i,t-j} - \text{大卖单成交额}_{i,t-j})}$$

5.3 基于小买单数据

小单流入额对数正态分布？

5.4 超常资金流向

测试发现买入卖出的方向性并不是很重要，于是将他试做一种成交量的数据，尝试去掉总体成交量变化的印象来获得超常资金金额

$$ExceBuy = Buy - Buy.shift(1) * (vol/vol.shift(1))$$

结果：非常奇怪的是，使用超常资金本身比拿他做标准化后的数据效果还要好

5.5 与价量相结合

1、猜想收益率数据可能与 ret 有直接的正相关性，考虑做正交化后的残差作为大单的 alpha 用来计算各个以你这：

结果：效果一般

$$factor = B_i/B_{i-1} - Vol_i/Vol_{i-1}$$

2、考虑计算 $\text{corr}(\text{BuyOrder}, \text{Ret})$

结果：效果不太行

6 主力流入启动因子

观察到其实有些股票出现，有主力资金流入拉升开始上涨，再主力紫金撤回后小单合力还能继续保持上涨的态势，尝试刻画这一类行为

7 对买卖单数据标准化的问题

尝试平衡不同标的容量的问题：

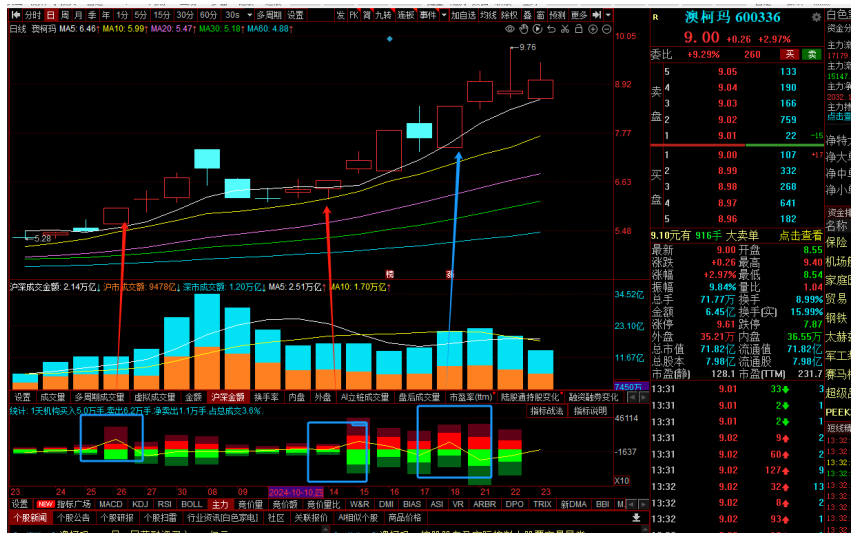
try1: 尝试除以前一段时间的均值

try2: 尝试除以成交量，但是这可能跟股票的性质有关？

try3: 对成交量做线性回归，只考虑增量

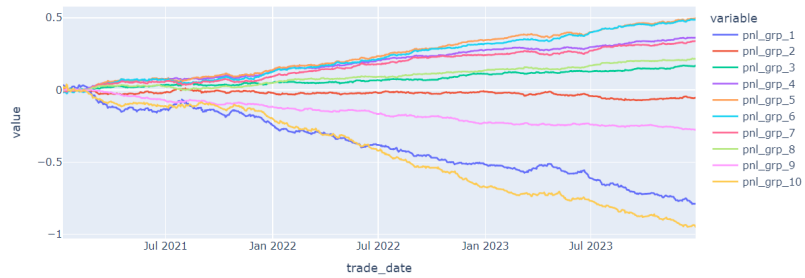
try4: 尝试用均线等指标刻画

小单买入负向预测性明显，尝试大单减去买入单后再除以成交量进行标准化



8 扭转因子

Demean PnL by Group 09:30

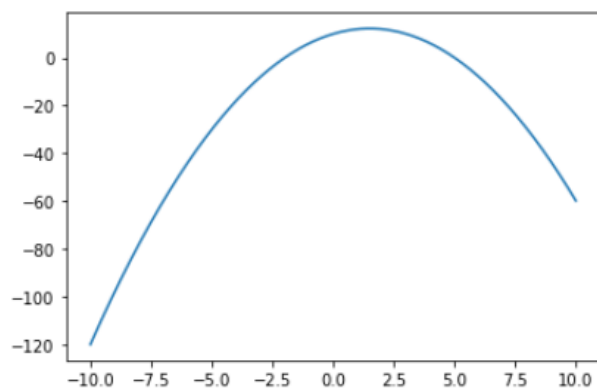


类似于像最大最小组收益均比较低, 收益与因子的关系是类似于开头朝下的抛物线形式, 而非线性的关系

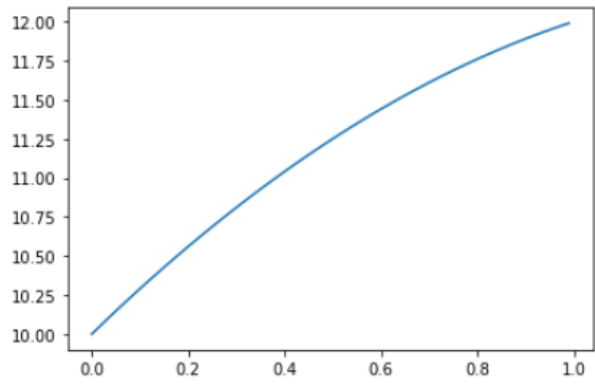
1、尝试减去均值后平方:

效果很差

原始函数, 记作 $f(x)$:



尝试寻找一个函数 g , 复合上原来的函数个 $g(f(x))$ 后图像类似如下:



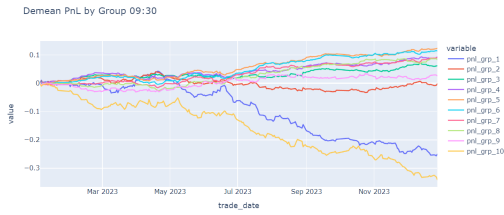
即总结为如下问题: 假设光滑函数 f 满足 $f''(x) \leq 0$, 找到一个次数比较低的函数 g , 使得:

$$\frac{d^2}{dx^2}g(f(x)) \rightarrow 0$$

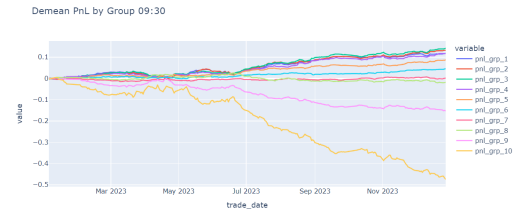
$$\frac{d}{dx}g(f(x)) \text{ 尽量大}$$

尝试 1:

对抛物形态分布的因子, 使用减去中位数除以标准化进行标准化之后, 再取绝对值
对 BuyActRet 因子进行尝试

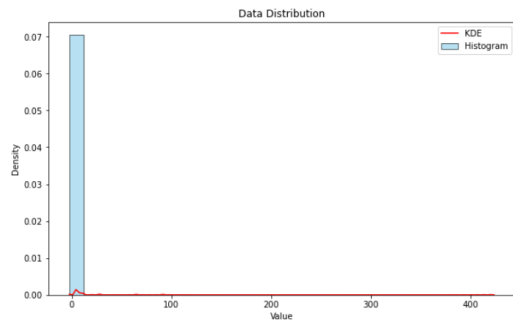


(a) origin

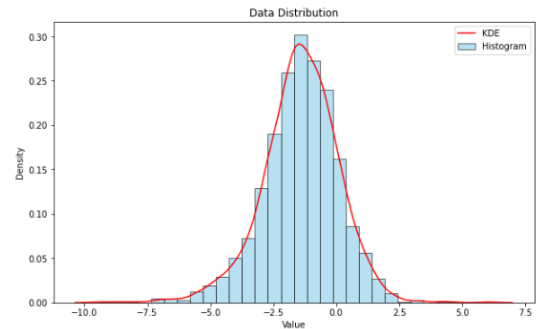


(b) transformed

尝试用数据分隔开训练集和测试集和 ic 来训练出复合函数 g , 基于 ExceBuyLOrder 因子:
对因子取对数均衡极限值



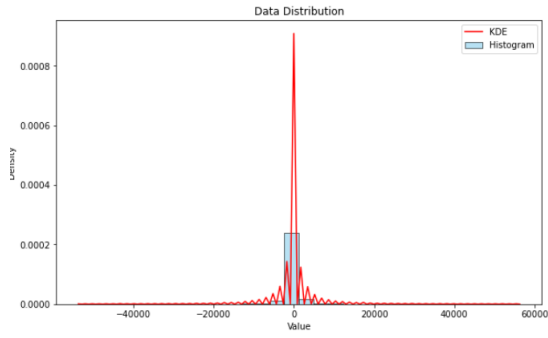
(c) origin



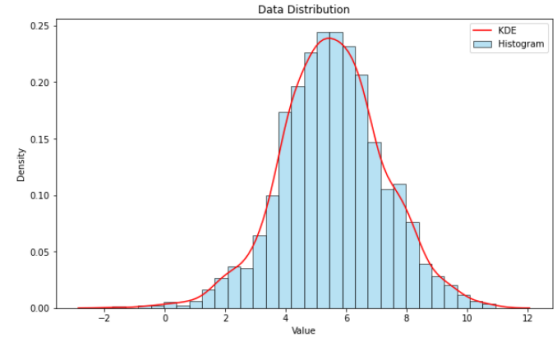
(d) log

直接对大买单因子做处理

$$PureBuy_i = B_i - S_i$$



(e) origin



(f) log

$$PureBuyPer_i = \ln(PureBuy_i / PureBuy_{i-1}) \text{ if } > 0 \text{ else NaN}$$

$$PureBuyPerZs = PureBuyPer \text{ 按照 axis=1 做 Z-score}$$

$$factor = -\cos(PureBuyPerZs)$$

9 大小单因子

B_i, S_i 记为第 i 天的大买/卖单

9.1 超额买

$$ExceBuy_i = B_i - B_{i-1} * \frac{Vol_i}{Vol_{i-1}}$$

可取一个对数

9.2 超额买比率

$$ExceBuyPer = \frac{B_i}{B_{i-1}} - \frac{Vol_i}{Vol_{i-1}}$$

9.3 净买单

$$NetBuy_i = B_i - S_i$$

可取一个对数

9.4 净买入收益率

$$NetBuyRet = (B_i - S_i) * Ret_i$$

9.5 BuyRet

某种程度上想描述主力的获利情况

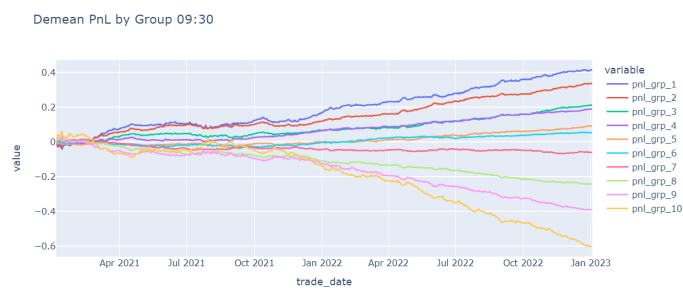
$$BuyRet_i = B_i * Ret_i$$

$$BuyPerRet = \frac{B_i}{B_{i-1}} * Ret_I$$

9.6 BuyLSRatioRet

主力相对散户获利情况

$$BuyLSRatioRet_I = \frac{BuyL_i}{BuyS_i} * Ret_I$$



9.7 对上面因子取对数，改变分布