

Sep Alpha

aglv

2025 年 3 月 16 日

1 问题

1.1 处理涨跌停股票因子值时

本质上还是为了让回测效果更加接近实盘效果：

方法 1：因子在第 i 日的值作为第 $i+1$ 日的调仓依据，若某股票在第 $i+1$ 日涨跌停，则将第 i 日因子变成 NaN，即不将其进入分组回测的股票池

缺点：用到了未来数据

方法 2：若第 i 天涨跌停，则将第 i 天的因子值变成 NaN

缺点：此时第 $i+1$ 天股票不一定涨跌停，可以进行交易，违背了只想剔除不能交易的情形的初衷

2 换手率分布均匀度

2.1

1. 获取分钟计算成交量 vol

2. $\text{turnvol} = \text{vol} / \text{vol.sum}()$ 针对每天计算换手率比，每天的 turnvol 总和为 1

3. 计算每天换手波动 $\text{turnstd-daily} = \text{std}(\text{turnvol})$

4. 均匀度： $\text{turnstd-daily} / \text{turnstd-daily.rolling}(20).\text{sum}()$

2.2

以 5 天时间期的相对交易 vol 来记作换手情况，改进因子：

$$\text{TurnMin} = \frac{\text{VolMin}}{\text{VolDay.rolling}(5).\text{mean}()}$$

2.3

不以标准差来判定换手均匀度，改用尾部分布：

1.

$$\text{TurnDay} = \text{Vol.groupby}('Date').\text{sum}() \quad \text{当日成交量和}$$



图 1: 去涨停股票, 除 20 日平均值

$$TurnMin = \frac{VolMin}{VolDay}$$

2.

$TurnCvar = TurnMin[TurnMin > TurnMin_{0.9}].mean()$ 分位数 0.9 以上的数据的均值

2.4

猜想换手率分布不均背后反应的是当天信息丰富度和交易情绪, 考虑结合当天收益率来考虑

$$\text{新因子} = TurnStd * abs(Ret)$$

Ret 因子为当天的收益率

2.5 将 1 分钟级别换手率数据变成 10 分钟级别换手率数据

适当增大窗口期, 提升计算稳定性

同时改变判定换手波动的指标, 为了体现时序特征, 将计算标准差改成计算一阶差分

$$NewFactor = Vol10Min.diff().abs().sum()$$

(但是效果并不好)

3 价量配合

3.1 基本形态

3.2 日内价量相关系数

认为日内若价格上涨时伴随着放量的股票总体上涨趋势大, 日内下跌时放量的股票具有下跌趋势

考虑因子:

$$PV_{corr} = corr(p_i, v_i)$$

价量关系	成因及影响
量增价涨	一般出现在上升通道，投资者情绪高涨，成交量持续放大，股价不断被推升
量增价平	成交量放大而股价不涨，有主力对倒、底部震荡吸筹等情况
量增价跌	一般出现在下降通道，主力出货，抛盘严重，杀伤力十足，短线风险极大
量缩价涨	既可能由于股价偏高，投资者跟进意愿不强，缺少市场人气，抑或者由于筹码集中，主力惜售，需留意后续变化
量缩价平	成交量缩小，而股价不跌，保持横盘。若股价处于上升通道，可能出现缩量横盘式洗盘，一旦突破有望大涨；若股价处于高位震荡，往往是主力出货形态，一旦跌破箱体即有暴跌可能
量缩价跌	一般呈现两种洗盘形态：一是上升通道中的短线洗盘形态，二是下降通道中的中线调整洗盘
量平价涨	横量上涨，市场人气较好，主力控盘能力强
量平价平	滞涨现象，多空不明，不宜操作，后续走势分情况：若股价已经充分下跌，重新从低位拉升，可能为上升途中休整；若股价经历大涨后从高位下跌，可能为下跌途中休整
量平价跌	一般出现在下降通道，股价加速赶底，成交量温和缩量，但已不足引起市场恐慌，没有更多筹码抛出

3.3 日内价量差分相关系数

效果下降：考虑到收益率序列呈现高度自相关性，在某段时间可能倾向于连续上升或者连续下降状态，而其一阶差分序列在零值附近摆动，波动性大，观察不到其单调性

3.4 分离日内和隔夜涨跌幅

通常上日内涨跌呈现反转效应，隔夜涨跌通常具有动量效应：

考虑到：决定隔夜涨跌的往往是机构，广大散户跟风现象严重，而日内交易散户成交量高

3.5 换手率与价的关系

- 1、伴随着换手率的增大，反转效应越强？选股能力越强
- 2、隔夜的意外信息，可能反映在‘昨日量’和‘今日价’的错配之中
- 3、将收益分成日内收益和隔夜收益，并用换手率分组检测，
- 4、价差与换手率相关性：日内价差 CO，隔夜价差 OYC，换手率 V，日内换手率 IV，昨日换手率 YC

3.6 换手率与价格结合因子

想捕捉量增价涨的行情

idea1: 直接计算相关系数

$$factor1 = corr(VolMin, RetMin) \quad corr \text{ 可以是 pearson 或者 spearman}$$

(尝试过后发现 pearson 比 spearman 效果好)

idea2: 认为直接计算分钟级别的相关系数结果不稳定, 将数据从分钟级聚合成十分钟级
(回测结果相对平稳度增强, 但是不多)

结果: 几乎无变化



(a) No-agg



(b) agg-10min

idea3: 改用日频数据计算

1、取日频数据的 rolling15 天的相关系数

2、取 close 和 vol.shift(i) 的 corr, $i \in \text{range}(5)$

结果: 日度调仓的回测曲线一般, 加大调仓周期后回测效果明显增强

idea4: 不计算总体数据的相关系数, 只聚焦于价涨时段的量的变化
回溯 15 日中取 ret 排名前 5 日的 vol 比例和:

$$\text{VolPercent} = \text{Vol} / \text{Vol.sum}()$$

`index = close.pctchange().sort()[-5:].index` (取这段 rolling 时间的涨幅的极大值)

`return → VolPercent[index].sum()` (取上涨行情中的成交量在这一整段 rolling 的时间内的比例)

结果: 计算速度有点慢, 考虑计算上规避排序

$$\text{id1} = \text{Ret.idxmax}(), \text{id2} = \text{Ret.drop(id1).idxmax}() \dots$$

or 转用分钟级数据计算, 聚合成十分钟数据, 去除开盘前十分钟数据, 找涨幅最大的十分钟内的成交量占比:

4 NewFactor-9.19

1、设置 Ret 的阈值来筛选样本

2、进行更精细的计算

3、不使用排序来对极端值进行计算

4. 由于不太好先验地确定 Ret 的阈值, 可以考虑使用一个激活函数来提高更大的值的权重值, e.g. Sigmoid 函数

4.1 考虑使用激活函数

$$Corr_{pearson}(X, Y) = \frac{(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{Std(X)Std(Y)}$$

记 \hat{X}, \hat{Y} 为 X, Y 做 z-score 后的随机变量, 即:

$$Corr_{pearson}(X, Y) = \hat{X}\hat{Y}$$

且 $\hat{X}, \hat{Y} \sim N(0, 1)$

具体计算中, 我们希望捕捉价和量的某种同步性, X, Y 实际上是 Price 和 Vol, 且希望给增强较快的区间给予更大的关注, 我们用 $\delta\hat{X}$ 代表 Ret

$$\Delta\hat{X} = \frac{X_1 - X_2}{X_1} = 1 - \frac{X_1}{X_2}$$

其中 X_1, X_2 是 \hat{X} 的两个 copy, 均服从标准正态分布

则 $\frac{X_1}{X_2}$ 服从柯西分布, (此要求 X, Y i.i.d, 这个假设不一定成立, 但是我们先暂且这样假设)

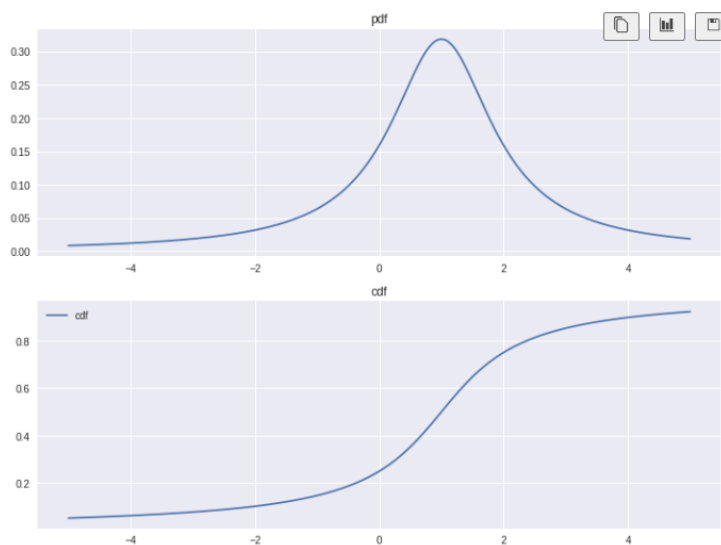
$$\frac{X_1}{X_2} \sim Cauchy(0, 1), \quad pdf : f(x) = \frac{1}{\pi(1+x^2)}$$

我们希望在 Ret 上选取一个恰当的激活函数, 从而将此函数作用在 Ret 上的值作为求相关系数时 \hat{X}, \hat{Y} 做内积时的权重, 因此我们需要考察 Ret 的分布进而选取一个好的激活函数

依照上面的分析, 随机变量 $\delta\hat{X}$:

$$pdf : f(x) = \frac{1}{\pi(1+(1-x)^2)}$$

$$cdf : F(x) = \frac{1}{2} - \frac{1}{\pi} \arctan(1-x)$$

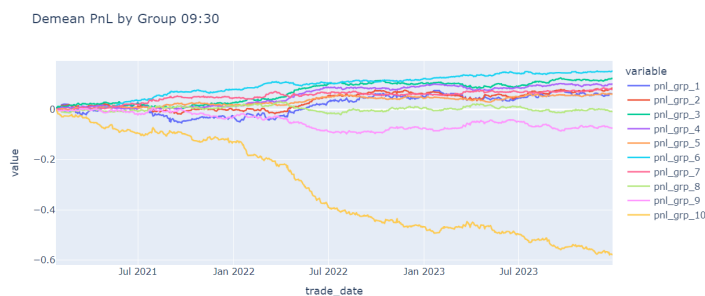


想法 1: $w = \text{sigmoid}(\delta X)$, $Corr = w\hat{X}\hat{Y}$

想法 2: 只取上涨行情 $w = \text{ReLU}(\delta X - 1)$

想法 3: 对下跌行情施加惩罚 $\text{sigmoid}(\delta X - 1)$

尝试 2: 结果一般般, 空头因子组表现明显, 计算速度有点慢, 考虑想法 4:



尝试 2

想法 4:

$$factor = corr(\sum_{i=1}^{win_num} sigmoid(ret), vol)$$

结果: 因子结果与尝试 2 结果趋同, 空头组表现强烈

想法 5: 在想法 4 的基础上, ret 分别带入隔夜 ret 和日内 ret

发现使用隔夜收益效果更好

想法 6: 考虑更换其他激活函数

激活函数:

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

$$relu(x) = \max(0, x)$$

$$relu_2(x) = 1 \text{ if } x > 0 \text{ or } 0$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$LeakyRelu(x) = \max(x, \alpha x) \quad \alpha \text{ 通常是个很小的正数}$$

发现对 vol 序列做类 z-score 后回测结果有小幅度提升

问题: 为什么多进程池的计算速度反而慢

4.2 考虑使用其他种类内积

e.g. p-内积:

$$\langle x, y \rangle = \left(\sum_{i=1}^n (x_i y_i)^p \right)^{\frac{1}{p}}$$

结果: 没什么用

考虑用向量夹角刻画相关性:

$$factor = \cos \langle ret, vol \rangle = \frac{\langle ret, vol \rangle}{|ret| * |vol|}$$

结果: 效果跟 corr 基本一样

5 分离隔夜和日内收益因子

$$NightRet_i = \frac{open_i}{close_{i-1}}$$

$$IndayRet_i = \frac{close_i}{open_i}$$

发现 a 股市场隔夜收益具有动量效应，日内收益具有反转效应 考虑将每日收益拆分为隔夜收益



(c) 隔夜



(d) 日内

与日内收益，加强动量类或反转类因子效果

特别地，通过收益率取对数的形式将两部分收益率以加法形式分开

$$\ln(Ret) = \ln\left(\frac{close_i}{close_{i-1}}\right) = \ln(NightRet) + \ln(IndayRet)$$

E.g.: 以 CVaR 因子为例对比原始收益率和隔夜收益率作为样本的因子效果:



(e) 隔夜收益率



(f) 原始收益率

使用隔夜收益率替代原始收益率计算动量因子 CVaR，效果有明显提升