

回测框架说明

🐱 数据读取说明

✅ mat文件可以用以下代码读取

```
import scipy.io as scio #加载用到的库

data=list(scio.loadmat('xxx.mat').values())[3] #xxx.mat为你的mat文件名
#由于通过 scio.loadmat函数读入的变量，其类型是一个字典
#该字典有4个key-value对，其中第4个value为需要的数据
```

🐱 数据预处理

✅ 剔除上市未满 60 个交易日的次新股、ST 股、非正常交易股票，这三个信息分别在“AllStock_DailyListedDate.mat”（上市交易日数）、“AllStock_DailyST.mat”（是否 ST 股）、“AllStock_DailyStatus”（交易状态）excel 中：“上市交易日数”中，每个格子都是截至昨日收盘的上市交易日数；“是否 ST 股”中 1 代表当日是 ST 股；“交易状态”中“1”表示当日正常交易，“0”是非正常交易。

⚠️ 如何将日度状态月度化？

- 1 交易状态：如果过去一个月内，正常交易的天数比非正常交易的天数多，则认为该月交易状态正常；否则就认为交易状态不正常，予以剔除。
- 2 st 股：如果过去一个月内，是 st 股的天数比不是 st 的天数多，则认为该月是 st 股，予以剔除；否则就认为不是 st 股。
- 3 上市交易天数：如果该月最后一个交易日时，上市天数不足 60 天，就认为该月上市天数不足 60 天，否则就认为上市天数达到了 60 天。

⚠️ 市值中性化处理

👉 为什么要做市值中性化？

📌 因为有些因子值的大小，会受到股票市值大小的影响，若要剔除这一部分的影响，就需要做市值中性化。

👉 如何对因子做市值中性化处理？

- 1 计算出月度的因子值。
- 2 计算股票每月月末的流通市值。计算方式为：每月最后一个交易日的流通股本*每月最后一个交易日的收盘价。
- 3 在每个月月末，各自做一次带有截距项的横截面回归，取残差，残差即为中性化后的因子值。因变量 y 为因子值，自变量 x 为该月月末流通市值的对数。

🐱 因子回测

- 1 每月月底，回看过去 K 个交易日（K 一般取 20），计算因子值，以反转因子 Ret20 为例，即为每只股票过去 20 个交易日的累计收益率（t 日收盘价 / t-20 日收盘价-1，复权价）；

2 按照因子值对所有样本从小到大排序，等分为 5 组（或 10 组），分组 1 因子值最小，分组 5（或分组 10）因子值最大；

3 下个月月初，5 个组（或 10 个组）都以月初开盘价，等权（此处指，买入每只股票的钱都相等，例如第一组的股票有 300 只，每只股票都买入 10000 元，这样计算本月收益时，该月该组的收益即为这 300 只股票的收益率均值，此处收益率指 $(\text{当月收盘价} - \text{当月开盘价}) / \text{当月开盘价}$ ）买入组内相应股票，持有至月底，以月末收盘价平仓（都用复权价）；

4 每个月都重复上述操作；

⚠ 特别注意：每月月底平仓时，不卖出月末收盘跌停股；每月月初建仓，不买入月初开盘涨停股。

👉 涨停和跌停怎么判断呢？

📌 当每月第一个交易日的开盘价，高于上月最后一个交易日的收盘价的 1.098 倍时，我们认为该股票发生了月初开盘涨停。（此处忽略近年来创业板 20% 和北交所 30% 的涨跌停限制，一律视为 10%，下同）

📌 当每月最后一个交易日的收盘价，相对于该月倒数第二个交易日的收盘价，下跌了 9.8% 时，我们认为该股票发生了月末收盘跌停。

👉 发生月末跌停或月初涨停时，具体要怎么操作呢？

📌 月末跌停处理，具体来说：例如 2020 年 5 月时，600519.SH 这只股票在第 1 组，恰巧 2020 年 5 月的最后一个交易日收盘时跌停了，那么 600519.SH 这只股票将不能卖出去，因此在 2020 年 6 月时，第一组里将依然有 600519.SH 这只股票；而根据 2020 年 5 月的因子值，计算得出 2020 年 6 月时，600519.SH 这只股票理应在第 4 组，那么在 2020 年 6 月时，第 4 组也将买入 600519.SH 这只股票。总结来说，2020 年 6 月时，600519.SH 这只股票将同时位于第 1 组和第 4 组中。

📌 月初涨停处理，具体来说：例如 2020 年 10 月收盘时，根据因子值的大小，我们判断出 600519.SH 这只股票在 2020 年 11 月时，理应位于第 3 组，但是在 2020 年 11 月第一个交易日开盘时，600519.SH 这只股票涨停了，那么我们 2020 年 11 月时，将不再买入 600519.SH 这只股票。总结来说，2020 年 11 月时，600519.SH 这只股票，被从我们的股票中剔除了。

🐱 绩效指标计算

1 得到 5 个分组（或 10 个分组）每个月的收益率；

2 每个月计算多空对冲收益：分组 1 的当月收益减去分组 5 的当月收益；

3 计算多空对冲净值：假设每个月的多空对冲收益率为 r_1, r_2, \dots, r_n （此处 n 为回测期的总月数），则每个月的累计净值为 $1 * (1 + r_1) * (1 + r_2) * \dots * (1 + r_n)$ ；

4 根据多空对冲净值序列，计算多空对冲年化收益，方法为：假设净值序列的最后一个数为 X ，回测时间共有 N 个月份，年化收益为 R ，则有 $(1 + R)^{(N/12)} = X$ ，反解得到 R ；注意 N 为有收益的月份，比如从 2019 年 1 月开始回测，1 月底回看过去计算因子值，2 月份才开始有收益，所以 N 的数量应该从 2 月份开始算；

5 多空对冲年化波动率 σ ，即为 $\text{std}(\text{每月多空对冲收益率序列}) * \sqrt{12}$ ；

6 多空对冲信息比率 $= R / \sigma$ ；

7 多空对冲月度胜率 $= \text{每月多空对冲收益率序列中大于 0 的数量} / \text{总交易月数 } N$ ；

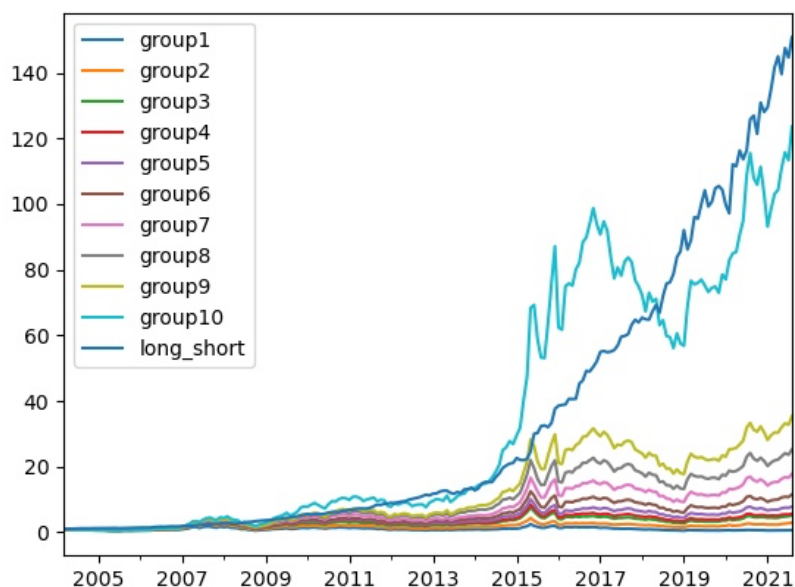
8 多空对冲净值的每月回撤 $= \text{当月净值} / \max(\text{起点至当月净值}) - 1$ ，最大回撤 $= -\min(\text{每月回撤})$ ；

9 IC、年化 ICIR 的计算：每月月底，利用过去一段时间的信息，计算得到所有股票的因子值，记为一列向量 $\{F\}$ ，下个月所有股票收益记为 $\{Ret\}$ ，则当月的信息系数 IC 等于向量 $\{F\}$ 与向量 $\{Ret\}$ 的相关系数；月度 IC 均值=所有月份相关系数的平均数，年化 ICIR=(所有月份相关系数的平均数/所有月份相关系数的标准差)* $\sqrt{12}$ ；

10 RankIC、年化 RankICIR 的计算：每月月底，利用过去一段时间的信息，计算得到所有股票的因子值，记为一列向量 $\{F\}$ ，下个月所有股票收益记为 $\{Ret\}$ ，则当月的 RankIC 等于向量 $\{F\}$ 与向量 $\{Ret\}$ 的秩相关系数；月度 RankIC 均值=所有月份秩相关系数的平均数，年化 RankICIR=(所有月份秩相关系数的平均数/所有月份秩相关系数的标准差)* $\sqrt{12}$ 。

🐱 绘图

✅ 将 5 个组（或 10 个组）以及多空对冲的净值走势，绘制为一张图，示例如下。



🐱 Barra 相关系数

✅ 分别取 10 个 barra 因子每月最后一个交易日的值，作为该 barra 因子在该月的因子值；每个月将 10 个 barra 因子分别与你的因子在横截面上求相关系数，得到每个月 10 个相关系数；在时间序列上分别对这 10 个因子求平均值，得到 10 个平均数，这 10 个平均数就是你的因子和 10 个 barra 因子的相关系数。通常来说，我们希望这些相关系数越低越好。

🐱 Barra 纯净化和行业中性化

✅ 用你的因子和 10 个 barra 因子以及 28 个申万一级行业哑变量做回归，取残差，即为纯净化之后的因子。其中你的因子为因变量 y ，10 个 barra 因子和 28 个申万一级行业哑变量为被解释变量 x 。用得到的新因子重新做一次回测。