**简报**

先贴个**任务**在这里



拆分任务：

1. 通过tushare取沪深300成分股最近一个月行情数据

2. 通过eps和市值排序

3. 分组收益率回测

**1. 获取行情数据：**

**注册tushare，不花钱买积分的话只有120积分，因此只能获得日线行情，无法获得市值，eps，hs300表等数据**

* **从中证指数网站上下载excel格式的hs300成分股列表并另存为csv文件**
* **从tushare上获取这300支股票的过去一个月内20个交易日的日线数据，包括**
* **去baostock网站上获取这300支股票最近一期财报上的eps（2022年三季度财报）和市值（最后一天的）**
* **发现baostock网站上缺乏‘sz.002555’的数据，手动翻阅财报补上**

**2. 通过eps和市值排序**

**双排序，应该意思是说分层抽样。有两个因素因此有两个不同的顺序。以下只以一种情况为例分析。**

* **先按照eps切成五份，每份有60个**
* **再按照市值切五份，每份12个**

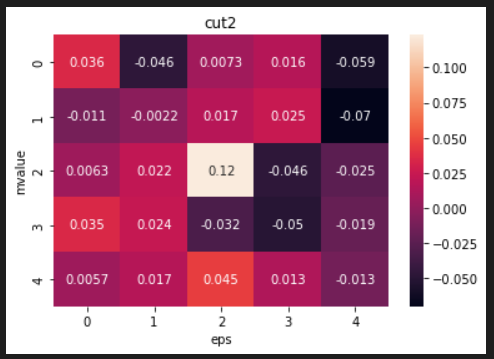
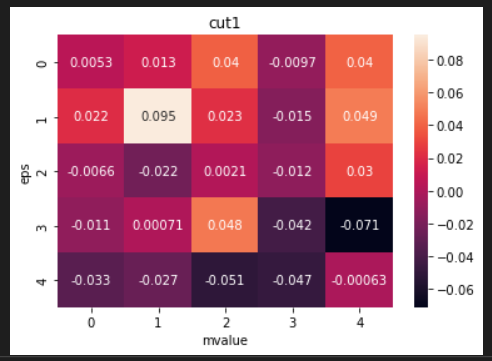
**3. 分组收益率回测**

**应该是说，要我把这25个组合作为投资组合计算一个月内的收益情况。同样以一种情况为例**

* **一共有5\*5个投资组合。计算它们的每日平均收益率作为组合的每日收益率**
* **计算累计收益率（1+dr）cumprod()-1**
* **最后一个交易日（'20230324'）就是一个月的累积收益率了**

**4. 分析数据**

**排序是从小往大的，切片也是**



**左图是先切eps的情况。结果一个月内的收益率和eps是负相关的？貌似不和常理。收益和市值相关性不大。**

**右图是先切市值的情况。市值和收益率倒是呈一定的正相关性。和eps相关性不大。**

**思考：**

**可能因为eps和市值本身存在一定的关联性。因此这个分层抽样不是很合理？**