

简报

先贴个任务在这里

通过 tushare 取沪深 300 成份股最近一个月的行情数据 根据 eps 和市值排序做分组收益率回测

拆分任务：

1. 通过 tushare 取沪深 300 成份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 成份股列表并另存为 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

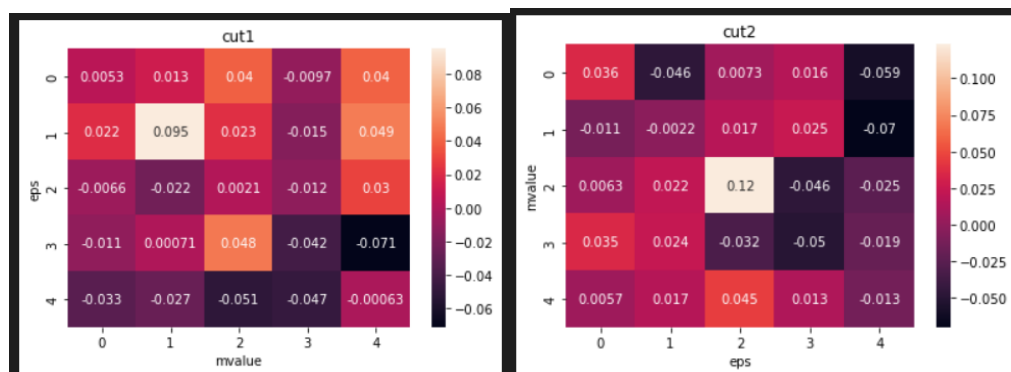
3. 分组收益率回测

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 $(1+dr) \text{ cumprod()}-1$
- 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？