

2024 《量化投资》，《金融数据分析》作业 2

请用 Python 完成以下题目。作业请用.ipynb 编写并转换成 HTML 文档格式提交至我的邮箱：zhaobo@nankai.edu.cn

命名格式：姓名-2-quant

截止日期：2024 年 4 月 22 日

1. 考察中证 1000 成分股¹在 2017-01 至 2024-02 月频率的 BM, 市值 (mktcap)。数据可从优矿平台下载。无风险收益率可从锐思数据库下载，或者从优矿平台调取。无风险收益可以用 Shibor 或者其他你认为合理的数据（例如用 1 个月中债国债收益率曲线）
 - 先按 mktcap 分成三组。分组时间点：t 年 7 月，用 t 年 6 月的 mktcap 分组并保持分组不变至 t+1 年 6 月。同时计算每组的简单平均收益率和按市值加权的收益率（每只股票分配的权重是其市值占分组内总市值的比例）。对于中证 1000 成分股，mktcap 分组的投资组合是否有显著不一样的收益率？
 - 做 BM 和 mktcap 的条件分组 (dependant sort)。也即，先用 BM 分组之后，再在每个 BM 分组内部做 mktcap 的分组。分组时间点可以自行决定。BM 分成三组，mktcap 分成两组。同时计算简单平均和按市值加权的收益率，这 6 组是否显著不一样？
 - 做 Fama-MacBeth 回归，结论是否和排序法有显著差异？
2. 不同行业可能有不同的分化走势。应当如何构建行业因子？查阅资料（请列出主要参考文献）、自选数据并做出分析。
3. 从优矿中选择你感兴趣的几个因子，做分组、FM 回归分析。

¹成份股构成可能会随时间发生变化。可以考虑此复杂性并相应做出股票池调整，也可以不考虑。