

Flab QR 笔试题

笔试题：基于量价与基本面字段构建交易策略

题目背景

data.zip提供了一批从 2015-01-01 至 2025-07-30 的日频特征数据，涵盖了价格、成交量、波动率等量化指标，以及部分基本面的feature。

数据以 `.parquet` 文件形式提供，每个文件对应一个特征（feature），文件名即特征名。

目标变量为 `ret_21D`，表示未来 21 个交易日的收益率。

任务目标

在不引入任何未来信息（未来数据或泄漏未来走势）的前提下，利用给定的特征数据设计并实现一个策略，使得策略在样本区间内获得**可观的 PnL 曲线**。

你可以自由选择方法，不局限于机器学习，可以结合统计、信号处理、因子合成等方法。

数据说明

- 时间区间：2015-01-01 至 2025-07-30
- 数据频率：日频
- 特征类别：
 - 量价类**（成交量、成交额、波动率、价量关系等）
 - 基本面类**（财务指标、估值等）
- 目标变量：
 - `ret_21D`：未来 21 个交易日收益率（**注意**：不能直接使用目标期内的信息来构建特征来防止引入未来信息）

要求与约束

1. 无未来信息

特征构造、信号计算必须严格遵循时间顺序，防止数据泄漏。

2. 策略形式

- 可以是打分排序、离散信号组合、基本面择股等多种选股方式
- 允许使用多因子组合与权重优化

- 可自由选择调仓周期（每日、每周等）

3. 评价指标

- 最终策略的 PnL 曲线走势
- 年化收益率、夏普比率（Sharpe Ratio）
- 最大回撤（Max Drawdown）

4. 可使用的工具

Python（pandas、numpy、scikit-learn、statsmodels、lightgbm、xgboost、ta-lib 等） / Chatgpt等大语言模型

5. 其他数据

- a. 可使用例如yahoo finance等免费数据网站下载任何其他数据用于构建策略模型和回测
- b. 需在report中说明数据金融学意义，以及为什么需要这些数据

提交内容

1. 策略代码（Jupyter Notebook / python文件）

- 数据读取、特征处理、信号构建、回测流程

2. 结果分析报告（Jupyter Notebook导出的pdf）

- 方法说明（特征选择及其特征工程、信号生成逻辑）
- 回测结果（PnL 曲线、主要绩效指标表）

加分项

- 稳健性检验（根据信号类别考虑不同时间区间、不同股票池的表现）