

MultiFactor Barra Proposal

以华泰金工多因子研报为基础，对多因子模型进行复现，了解多因子模型在业界是如何使用的。传统多因子模型是多因子分析框架的基础，在本次内培结束后，我们希望通过此次的模型复现，达成可以快速阅读多因子相关研报，迅速构建其他衍生多因子模型的能力，从而获得深入分析各个因子对股票市场影响的手段。

整个多因子模型分为数据收集、单因子识别、因子收益模型、风险模型、二次规划确定权重、回测多因子绩效分析等部分，下面分别进行叙述。

步骤

一、数据处理

1. 基础数据采集（已经有数据）
2. 数据预处理（增删等处理，ST 股）
3. 数据标准化（确定需要计算的因子，十二大类里面每个挑选 2 种；使用原始数据计算选取的因子载荷；使用因子载荷原始值或者排序值进行标准化去量纲等的处理）

二、单因子识别

1. 单因子回归得到每个因子每期的因子收益（在一个 dict 里面，df 行为时间，列为股票）

$$\tilde{r}_j^t = \sum_{s=1}^S X_{js}^t * \tilde{f}_s^t + X_{jk}^t * \tilde{f}_k^t + \tilde{u}_j^t$$

2. 因子收益率序列 t 检验

$$t = \frac{\bar{x} - u}{\sigma / \sqrt{n-1}}$$

3. 计算因子 IC 值

个股第 T 期在因子 K 上的暴露度与 T+1 期的收益率的相关系数

4. 因子打分法回测

按照因子值对股票进行打分，构建投资组合回测，在某个截面期上，根据一个或几个因子值对个股进行打分，将所有个股按照分数进行排序，然后分为 N 个投资组合，进行回测

三、因子收益

1. 大类因子分析

同类型因子的相关性检验；因子取舍或者因子合成

2. 因子共线性分析

上一步的补充，针对经济含义不同的因子，若存在明显相关性，需进行取舍

3. 残差异方差分析

Breusch-Pagan test 和 White test

4. 多元线性回归

$$\tilde{r}_j^t = \sum_{s=1}^S X_{js}^t * \tilde{f}_s^t + \sum_{k=1}^K X_{jk}^t * \tilde{f}_k^t + \tilde{u}_j^t$$

- 1) 通过多元回归得到因子每期的收益序列

- 2) 对 T+1 期因子估计预期收益率：历史均值、指数加权移动平均、时间序列预测、滤波法

提取趋势项

3) 根据 T+1 期因子收益率和因子载荷, 计算 T+1 期每只股票的预期收益率向量

四、因子协方差

根据多元线性回归, 得到所有因子每期因子收益的历史序列值, 可以计算因子收益率之间的协方差矩阵, 此外, 还需要估计残差风险

五、二次规划, 确定权重

使用数据: T+1 期股票预期收益、因子收益协方差矩阵、预期残差风险, 结合投资组合的风险-收益目标, 以及约束条件, 进行股票选择和权重分配

$$\min_H \frac{1}{2} * H^T * Q * H + H^T * c$$
$$s.t. \quad A^T * H \leq b$$

目标: 收益目标和风险目标

约束条件: 个股上下限约束、行业权重约束、因子暴露约束

六、回测, 多因子绩效分析

业绩分析指标: 年化收益率、最大回撤等

收益率回归、业绩归因

时间安排: 6 周

第一周: 数据处理+单因子识别

第二、三周: 因子收益+因子协方差

第四、五周: 二次规划+回测

第六周: 梳理+展示

分工

每周晚上有固定的时间一起写代码和讨论, 根据步骤内容和顺序分工写代码