**暂以华泰五因子模型为例**

输入变量：

1. 股票池定义：
2. 个股实际收益；
3. 市场因子指标；
4. 规模因子指标；
5. 估值（BP）因子指标；
6. 盈利因子指标；
7. 投资因子指标；

具体步骤：

1. **因子构造**

多因子模型的因子收益率，是沿着某一个或几个因子维度，将股票分层。这里我们考虑在每个月末，按照因子大小将所有股票分为2-3层，计算下个月每层内部股票的平均收益率，然后对不同层股票的平均收益率作差，以用来衡量该种因子的收益效应。以五因子模型为例，除市场因子以外， 将：

* + 1. 规模因子按中位数分为两层，分别记为大（B）和小（S）；
    2. BP因子按30%和70%位数分为三层，分别记为高（H）中（N）低（L）；
    3. 盈利因子按30%和70%位数分为三层，分别记为强（R）中（N）弱（W）；
    4. 投资因子按30%和70%位数分为三层，分别记为保守（C）中（N）强（A）；

因此得到的组合例如大市值股票和高BP因子的组合记为BH，以此类推。我们给出一下四个因子收益率的具体定义和计算公式：

|  |  |  |
| --- | --- | --- |
| 因子收益率定义 | 分组方式 | 计算公式 |
| SMB: 小市值组股票减大市值股票平均月收益率 | 规模与BP两维度分组 |  |
| 规模与盈利两维度分组 |  |
| 规模与投资两维度分组 |  |
| 以上三种的均值 |  |
| HML: 高BP减低BP | 规模与BP两维度分组 |  |
| RMW: 强盈利减弱盈利 | 规模与盈利两维度分组 |  |
| CMA: 保守投资减强投资 | 规模与投资两维度分组 |  |

1. **模型回归分析**

采用Fama和French（2015）提出的股票分类方法，沿着某两个因子维度将股票分成的25宫格的资产组合（每月调仓）。这里我们采用规模-估值（BP）、规模-盈利、规模-投资这三个组合维度将所有股票分成3组25宫格的资产组合，将每一组25宫格的资产组合分别进行独立的回归分析。

同样以五因子模型为例，由于市场因子必须包含在每个模型中，我们只需要把市场因子与剩下4个因子中任选3个一起组成4个不同的四因子模型，最后和五因子模型一起比较以测试五因子模型的有效性。

一般情况下假设有个因子的模型给出如下：

我们用GRS-Test来检验N个资产回归组合的Alpha是否联合为0. 具体构造如下：

* + 1. N个资产组合的收益率时间序列回归方程，即上述模型中i从1变化到N. 在五因子模型的例子中
    2. 模型的截距向量为：

收益率残差时间序列为：

残差矩阵为：

* + 1. 协方差矩阵： 是一个 矩阵

因子均值向量：，其中.

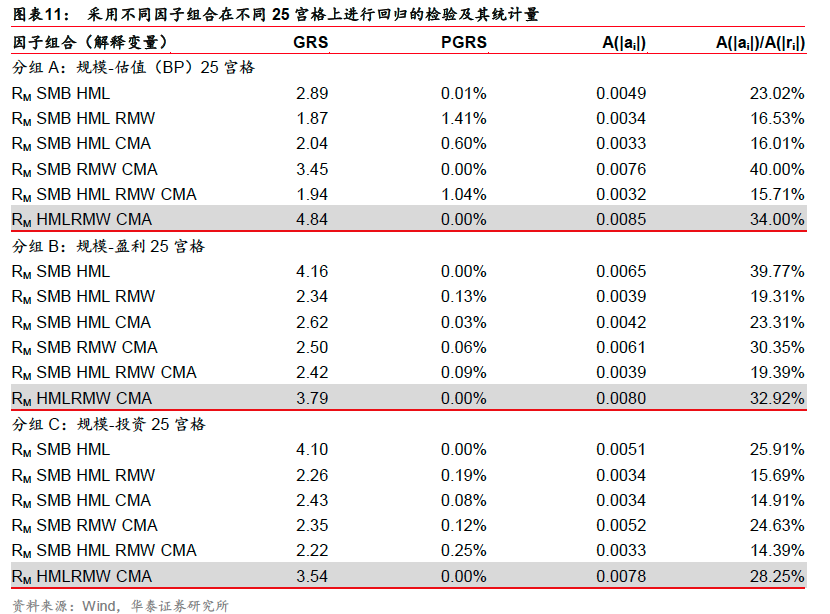
因子矩阵：, 均值：,

因子的无偏估计协方差矩阵：

* + 1. 最后

在模型足够好的理想情况下，Alpha理论值为0，也就是说GRS检验量越大，拒绝Alpha联合为0所犯错的概率越小。

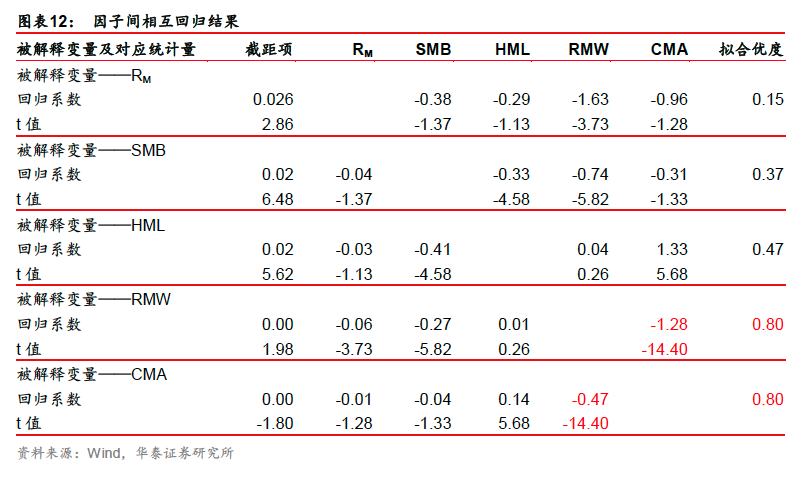
如果我们对25个组合分别做回归，同时给出拒绝25个Alpha联合为0的原假设对应的犯错概率（p值），记作PGRS可以得到类似下表



这里函数代表对某变量在25个投资组合中取均值，其中为因子模型的截距，中为CAPM模型的截距。

1. **因子多重共线性检验预处理**

将每一个因子对剩下的所有因子分别做回归得到回归系数，t检验量和拟合度（）得到类似下表



从回归系数和t值分析是否存在共线性情况，例如上表的例子中有如下分析和处理：

观察上表发现，CMA与RMW有很强的共线性，我们将投资因子中与其他因子共线性的部分去除，令

这里

针对修正后的因子我们得到了新的五因子模型：

1. **目标**

首先通过第二步模型回归分析得到解释力度最好的多因子模型，再通过第三步的共线性分析更新模型中与其他因子有较强相关性的因子，从而使得该因子对模型的解释度更直观。