使用python（spyder）, 所有t值使用Newey-West调整

一．数据处理：缩尾处理：左右1%。将低于样本1%的用1%的数值赋值，大于99%的用99%赋值（市场收益率和无风险收益率不做此处理）；剔除ST、\*ST股票。

此步骤单独在一个.py文件中，生成数据处理结束后的表格。后续步骤使用处理后的数据。

1. 因子。Rm-Rf 、SMB（市值)、HML（账面市值比）、RMW(净资产收益率）、CMA（总资产增长率）

此步骤单独在一个.py文件中

1. 单变量排序描述性统计

在每个时间t将股票按 市值、账面市值比、净资产收益率、总资产增长率 分别从小到大分为10组。计算每个组合的平均年化波动率、PB、ROA（组合中的股票等权重），之后在所有t将年化波动率、PB、ROA取平均。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | small | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | big |
| MV/BM/ROE/INV |  |  |  |  |  |  |  |  |  |  |
| 年化波动率 |  |  |  |  |  |  |  |  |  |  |
| PB |  |  |  |  |  |  |  |  |  |  |
| ROA |  |  |  |  |  |  |  |  |  |  |

计算每个组合的月平均收益率

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MV/BP/ROE/INV | small | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | big |
| 收益率 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|
|

1. 独立双重排序：

将市值和BM、ROE、INV分别独立双重排序，构造三个5\*5个组合，计算每个组合的月均收益率和t值（等权重）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | BM/ROE/INV | | | | | |
|  |  | small | 2 | 3 | 4 | big |
| MV | samll |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| big |  |  |  |  |  |

三．模型

2\*3 独立双重排序

市值按50%分两组 其他指标按30%、70%分三组 计算每组收益率（等权重加权）

|  |  |  |  |
| --- | --- | --- | --- |
|  | BM | | |
| 市值 | S/L | S/M | S/H |
| B/L | B/M | B/H |

账面市值比高减低

HML=(S/H+B/H)-(S/L+B/L)

|  |  |  |  |
| --- | --- | --- | --- |
|  | ROE | | |
| 市值 | S/W | S/N | S/R |
| B/W | B/N | B/R |

ROE高减低

RMW=(S/R+B/R)-(S/W+B/W),

|  |  |  |  |
| --- | --- | --- | --- |
|  | INV | | |
| 市值 | S/C | S/N | S/A |
| B/C | B/N | B/A |

总资产增长率低减高

CMA=(S/C+B/C)- (S/A+B/A),

共得到18个组合

SMB=()

其中=

三个模型分别放在三个.py文件

1. 模型1

-=+(-)+++

其中是通过市值和BM、ROE、INV分别5\*5的独立双重排序法构建的75个投资组合的收益率（等权重加权）。是t期市场收益率，为 t 期无风险收益率，SMB、HML、RMW 以及CMA是独立双重排序构建的多空对冲组合的收益率，代表了因子的收益率。

使用FM两步回归。对每个资产组合i，首先以资产组合的收益率作为被解释变量，以排序法构建的对冲组合收益率SMBt、HMLt、RMWt 、CMAt作为解释变量，通过时序回归得到因子暴露、、、、。之后以第一步估计的因子暴露作为解释变量，在每个时间点t对进行截面回归，得到t个因子溢价λ和截距项的估计值，将t个估计值取均值，可得到因子溢价和定价误差。

1. 模型二：

=+++++

首先将每期t截面上的各类指标标准化，使均值为0，方差为1，其中、、、是使用标准化后的公司指标，作为解释变量。为三个2\*3 独立双重排序形成的18个投资组合的收益率。回归可得到、因子收益率、、、。

考虑市值和BM、ROE、INV分别5\*5的独立双重排序法构建的75个投资组合，在t期对于某资产组合，可通过其四个公司特征（等权重加权）和、、、以及计算出该资产t期的预期收益率：

++++，该预期收益率与真实收益率之差就是该资产在 t 时刻的。用此法可以求出每个资产每期的，将它们在时序上平均就得到该资产的平均。

1. 模型3

-=+ (-)+++++

模型三使用了模型二的因子溢价、、、，并加入市场因子(-)，通过时序回归得到因子暴露为、、、、，为定价误差。

四．异象检验

1个月异常换手率异象：过去20日日均换手率和过去240日日均换手率之比

市现率

前6个月回报动量

前11个月回报动量

每期t将以上指标单变量排序 分为10组 最大组收益率减最小组收益率作为

模型一：

-=+(-)+++

将Rmt-Rft、以及独立双重排序得到的SMBt、HMLt、RMWt 以及CMAt作为解释变量，-作为被解释变量，做时序回归得到常数项

模型二：

-=++++

使用、、、作为解释变量，-作为被解释变量，做时序回归得到常数项 为定价误差

模型三：

-=+ (-)+++++

Rmt-Rft、、、、作为解释变量，-作为被解释变量，做时序回归得到常数项