2020年6月7日星期日

**FF1993分组研究两特征与return关系A股复刻：**

1. 考察时间区间：2006年7月2020年6月（2006年6月30日首次构建分组）
2. Rebalance date：每年6月30日（保证届时去年年报已出）
3. Size特征构建逻辑：以2019年6月30日rebalance group为例， 获取2019年6月30日各股总市值market\_cap（其他选择：流通市值circulating\_market\_cap）
4. BM（VALUE）特征构建逻辑：上例前提下，获取2018财年的年报BOOK EQUITY与2028年12月31日收盘总市值相除
5. 重构分组时点的股票筛选逻辑：18年12月31日停牌、待退市、ST股票筛除，无18年12月31日市值数据的股票筛除，18年财报发布日晚于19年6月30日的股票筛除，年报无book equity数据的股票删除，19年6月30日停牌、待退市、ST股票删除、无19年6月30日市值数据股票筛除，以上条件交集股票形成有效股票列表，获取该列表股票的19年6月30日至20年6月30日月度收盘价计算月收益，对于期间有停牌一整月的股票筛除
6. 组合收益率平均方法：市值平均法
7. 分组采用的breakpoint对象：以构建组合日所有有效股票特征的整体（非全市场股票整体也非某交易所股票整体（用NYSE单交易所股票整体考虑size影响【FF1993pg12】））

**过程细节优化：**

1. 2019年6月30日重构组合，保持一年时间至次年2020年6月30日，期间收益归属2019年自然年口径，有半年交错。 考虑一年rebalance两次，把错口移正。即2018年12月31日拿2018财年三季度报的数据构建组合，2019年6月30日2018年财年年报数据构建组合。计算本自然年两次rebalance组合的各半年return，与本年口径一致，不发生偏移。

**缺陷：**对于资产负债表数据，三季度报可以拿到最新的（如book market），但是涉及利润表相关特征只有三季度水平，无效！！！，该优化方案对FF VALUE CHAR有效

1. Breakpoint分组法改为小市值为主的交易所（理由见图标观察2.2）

**结果及图表：**

1. 平均市值（单位：十亿）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| mkt\_cap(billion) | | | | | |
| Size/BM | HIGH | 2 | 3 | 4 | LOW |
| BIG | 86.32 | 48.33 | 52.97 | 47.58 | 36.65 |
| 2 | 8.36 | 8.32 | 8.22 | 8.33 | 8.44 |
| 3 | 4.91 | 4.90 | 4.89 | 4.89 | 4.93 |
| 4 | 3.28 | 3.28 | 3.29 | 3.26 | 3.28 |
| SMALL | 2.15 | 2.10 | 2.07 | 2.09 | 2.03 |

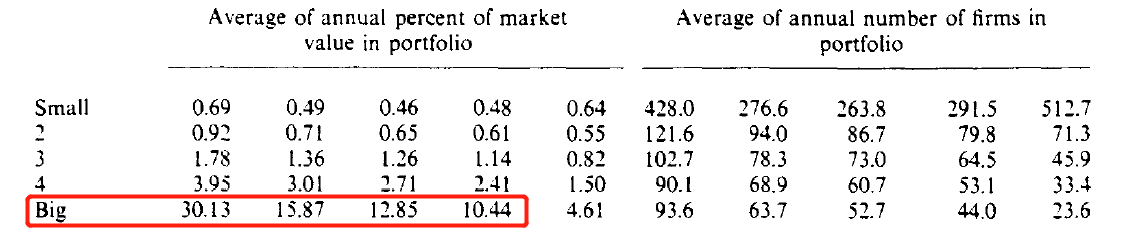
观察：

1. 纵看，最大市值组和2、3、4、5号市值组指数级差距，2、3、4、5市值组差距不大
2. 横向看，2、3、4、5市值大组内，BM小组的市值差距不大；但1号市值大组内，BM小组的市值差距较大，即低市净率大市值组的市值非常大，考虑银行股的原因
3. 平均市值占比、组合内平均股票数量

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| mkt\_cap\_distribution | | | | | | stock\_num | | | | | |
| Size/BM | HIGH | 2 | 3 | 4 | LOW | Size/BM | HIGH | 2 | 3 | 4 | LOW |
| BIG | 30.14% | 10.37% | 10.53% | 12.01% | 13.22% | BIG | 104 | 66 | 58 | 62 | 95 |
| 2 | 2.08% | 2.00% | 2.11% | 2.01% | 2.46% | 2 | 76 | 74 | 77 | 71 | 85 |
| 3 | 1.27% | 1.37% | 1.26% | 1.25% | 1.11% | 3 | 77 | 84 | 80 | 76 | 67 |
| 4 | 0.83% | 0.92% | 0.87% | 0.90% | 0.66% | 4 | 74 | 83 | 80 | 83 | 64 |
| SMALL | 0.39% | 0.55% | 0.62% | 0.59% | 0.49% | SMALL | 52 | 78 | 89 | 91 | 74 |

观察：

1. 1号size大组占选择股票总市值的76%
2. 横看，各size大组的组合股票数量相同385只左右，不按FF NYSE breakpoint分组法已可以保证在1号大组市值占比与FF分法接近。FF NYSE 分组法能使1号size大组的skew变小，故可考虑改进breakpoint来源，考虑小市值市场作为breakpoint，此举会让小市值组合的分配的股票数量增加

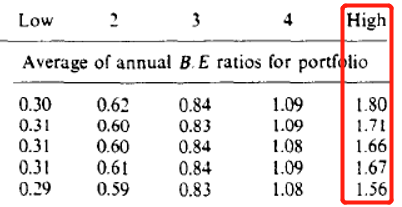
计算机生成了可选文字:
TheymentioninoneOftheirpapersthat"microcaps",thosestocksthatmakeup60％Ofthe
investableuniversebynumberonlymakeupsomethinglike3％Oftotalmarketvalue.The
variabilityincharacteristicsOfthesestockswouldmeantheydeterminemostOftheportfolio
breakpointswhenyouperformsortsTheuseOfNYSEbreakpointspreventsthis.
shareimprovethisanswerf01℃W'
addacomment
answeredFeb3，18att58
，；3980107
TheCRSPtapecoversNYSEstockssince1926TheAmexandNssdaqstockdatakicksinlater
（1960s）．Thereforeifyouaregoingtouseoneconsistentmethod忆breakstocksintomarketvalue
groupsitneedstobebasedonNYSEcriteria.Ifyoubaseiton削CRSPstockstherewillbea
breakorinconsistencyintheyearwhennon-NYSEstocksareaddedandthenumberCRSP
stocksincreasesgreatlyandmostOfthenewstockscominginarefairlysmallcap.()fyouplotthe
median()rotherquantile)削CRSPstocksMVthereisabigdropatthattime.Nothinghas
changedintheUSeconomy，itisjustthatalotOfsmallcapandmicrocapstocksarejoiningthe
CRSPdatabasethatwerepreviouslynotincluded).

1. 平均BM/市净率

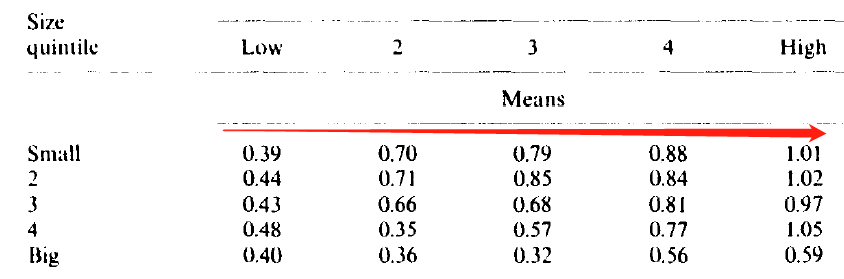
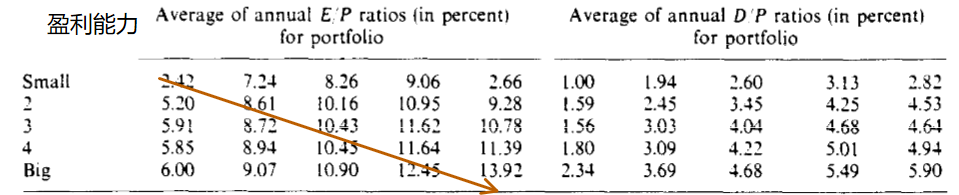
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BM | | | | | |
| Size/BM | HIGH | 2 | 3 | 4 | LOW |
| BIG | 0.7873 | 0.4793 | 0.3638 | 0.2721 | 0.1574 |
| 2 | 0.7521 | 0.4775 | 0.364 | 0.2726 | 0.1624 |
| 3 | 0.7232 | 0.4819 | 0.3634 | 0.2715 | 0.1651 |
| 4 | 0.7127 | 0.4799 | 0.3636 | 0.2731 | 0.164 |
| SMALL | 0.6868 | 0.4813 | 0.3626 | 0.2724 | 0.1614 |
| 市净率（1/BM） | | | | | |
| Size/市净率 | LOW | 2 | 3 | 4 | HIGH |
| BIG | 1.2701 | 2.0862 | 2.7486 | 3.6747 | 6.3545 |
| 2 | 1.3296 | 2.0943 | 2.7475 | 3.6684 | 6.1557 |
| 3 | 1.3827 | 2.0751 | 2.7517 | 3.6835 | 6.0561 |
| 4 | 1.4031 | 2.0837 | 2.7504 | 3.6616 | 6.0971 |
| SMALL | 1.4561 | 2.0776 | 2.7577 | 3.6713 | 6.197 |

观察：

1. 无平均市净率破净
2. 纵看，市净率最低的大组，市净率随市值减小递增趋势较其他大组明显，其他市净率大组无明显单调趋势。该特征同美股（FF1993 TABLE1）



1. 文章用E/P D/P指标验证BM、size与盈利能力关系，推论：经济萧条，大公司相对小公司更能存活赚钱，且在经济繁荣时期小公司也不那么赚钱（基于未知原因），故size小🡪赚钱能力差🡪更高的风险补偿🡪return高，；BM大🡪赚钱能力低🡪更高风险补偿🡪return 高。 但文章美股结果显示BM高 赚钱能力高，为什么return会高？文章逻辑链存疑（FF1993 TABLE1）。思考：return不一定只和赚钱能力有关，并非所有因子与return关系的逻辑链必须通过赚钱能力一环转接到return，待解决逻辑链：市净率高🡪return高

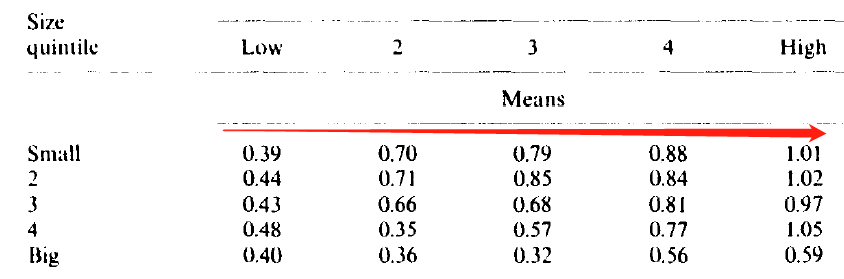


1. 平均收益率

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| return | | | | | |
| Size/BM | HIGH | 2 | 3 | 4 | LOW |
| BIG | 0.0105 | 0.0097 | 0.0092 | 0.0108 | 0.0110 |
| 2 | 0.0139 | 0.0138 | 0.0117 | 0.0111 | 0.0115 |
| 3 | 0.0156 | 0.0170 | 0.0159 | 0.0138 | 0.0131 |
| 4 | 0.0197 | 0.0196 | 0.0189 | 0.0184 | 0.0159 |
| SMALL | 0.0247 | 0.0241 | 0.0256 | 0.0248 | 0.0212 |

观察：

1. 该时段内存在FF的size结论，即size越小的组合return越高。且size对return的影响更‘显著’（从严格单调性的角度），即FF的size结果非严格单调，而此处各BM大组内（纵看）均严格单调



1. BM在FF表中并非严格单调，但不看中间看两端呈现明显递增且diff较大，反观A股同样不严格单调，在横条SIZE1大组内，BM与return无肉眼弱线性关系；size2、3、4、5大组有肉眼极弱线性关系即BM越大，return越高（注：观察表3各组的平均BM发现，各组BM差距非常明显，但差距显著的各分组的return并无较大差别）。欲得结论：A股BM不能作为一个因子（待不交叉分组再做一次观察结果）

**思考：**

1. 通过分组观察return单调性，可以论证特征对return是否有影响，但是该方法和直接回归观察显著性的关联是什么？可以说单调必然相同期间的回归就是显著的吗？可以把return range照搬到回归的loading range大小吗？（当前理解和猜测：分组论证特征有效，回归应显著，分组论证无效回归显著时无意义，只要分组无效就不看回归结果）
2. 考虑不交叉分组，直接考虑单特征的分组能否得到特征对return的单调影响？未来对某特征的发掘直觉上应该做单特征分组？（后文更新已论证可行）
3. 全时段有效能代表某时段有效？全时段无效能代表某时段必然无效？考虑分段分组验证

**阶段结论：**

1. 市值因子在全时段内（06年-20年）对return有影响，进一步考察单特征不交叉分组是否仍有线性规律，并考察子时段情况（两个角度：有效性（线性强度），影响力（最大最低分组return差值，考虑相对比值diff(最大最小组return)/diff(最大最小组平均市值)））
2. FF价值因子（BM/市净率）对return影响不大，进一步考察单特征不交叉分组及分时段检验，同1思路。另：价值因子的underlying考虑替换,即用其他特征替代BM作为价值因子，A股的‘价值股’概念阶段热门，如前几年的白马股周期，个人认为存在价值因子，但是这个因子不是BM，再找找那时候白马股周期的内在逻辑特征
3. 如果从分组检验return线性的角度判断BM因子，将BM排除，但如果回归时发现该因子显著怎么解释？以哪个判断为准？分组法还是回归法？个人倾向前者，但不排除这样的矛盾发生。1.线性回归的假设2.回归的魔性不可控3.回归的前提是能解释，分组法不通过就没解释力，强行回归显著也没意义

更新：

**用非交叉，单特征分组的结果：**

1. BM

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| BM | HIGH | 2 | 3 | 4 | LOW |
| mkt\_cap\_dist | 0.3471 | 0.1521 | 0.1539 | 0.1676 | 0.1793 |
| mkt\_cap | 28.0094 | 12.0307 | 12.1941 | 12.1669 | 13.2633 |
| bm | 0.7419 | 0.4802 | 0.3635 | 0.2722 | 0.1615 |
| stock\_num | 384 | 384 | 384 | 384 | 384 |
| return | 0.0121 | 0.0112 | 0.0102 | 0.0113 | 0.0109 |

观察：

1. 根据BM对股票分组，HIGH BM组的市值突出高（考虑银行股拉高）
2. 头尾return有较大差别，但中间过程线性程度较弱，abs[diff(return)/diff(bm)]，考察单位特征差带来的return差，0.002053087，而交叉分组的五个结果为

|  |
| --- |
| 0.00086 |
| 0.004077 |
| 0.004524 |
| 0.007035 |
| 0.006655 |

故通过单特征分组检验，BM特征有效性未变强（BM与return的线性程度），BM的影响力变弱

结论：单特征分组法与交叉分组法对BM有相同结论，即BM特征对return影响不大

1. Size

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Size | BIG | 2 | 3 | 4 | SMALL |
| mkt\_cap\_dist | 0.7626 | 0.1067 | 0.0626 | 0.0418 | 0.0264 |
| mkt\_cap | 59.05 | 8.34 | 4.90 | 3.28 | 2.08 |
| bm | 0.4199 | 0.3991 | 0.4121 | 0.4091 | 0.3790 |
| stock\_num | 384 | 384 | 384 | 384 | 385 |
| return | 0.0105 | 0.0122 | 0.0152 | 0.0188 | 0.0242 |

观察：

1. 市值二级分化较明显，组1平均市值指数级增长，且占据市场75%以上市值
2. return依然强单调，单位特征差带来的return差0.000240478；交叉分组的五个结果为

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0.00017 | 0.000313 | 0.000321 | 0.000307 | 0.000296 |

单特征分组检验得到的SIZE影响力与交叉分组情况差不多。

结论：单特征分组法与交叉分组法对SIZE的结论相同，SIZE在全时段有效且两方法影响力水平检验结果接近

**阶段结论：**

1. 考虑后续在做子区间的两特征对return的影响时以单特征分组法为主（省时方便），用交叉分组法偶尔抽检观察是否能得出相同结论。
2. 假设目前因子库里确定的有size因子，现检验一个新的因子，先做一次单特征分组法检验，若通过，再与因子库内的其他所有因子分别做交叉分组法检验，目的：记录两两因子间的相关性，在后续构建因子组合时，要刻意中和相关性。并思考两因子为何相关

2020年6月8日星期一

Size因子有效性研究，如何给出有效信号。即利用历史数据得到因子有效信号，并假设因子未来一定时间内有效

考虑方法：

对rebalance时点前window时间窗口内的分组计算return比较。

比较结果可以是严格单调递增/递减，头尾增减，即可能大市值更高收益，有时候小市值更高收益统计四个情况

同重构时点，不同window下的结果是否不同

出现严格单调，则在下一期做相关操作。逻辑链：严格单调则必然有size效应，至于越大越好还是越低越好看单调方向

优化考虑：又一年确定一次分组到每个月确定一次分组，甚至分组更新频率缩小到每天（目前不可信，代码不支持）