

证券研究报告—深度报告
金融工程
量化投资
金融工程专题研究

2016年12月05日

专题报告
相关研究报告:

《国信技术面量化选股系列: 价格路径对收益的影响》——2016-09-09
 《金融工程专题研究: 基于上、下行 Beta 风险的选股策略研究》——2016-08-29
 《金融工程专题研究: 使用估值指标进行股债配置》——2016-08-29
 《机器学习专题研究: SVM 算法选股以及 Adaboost 增强》——2016-05-31
 《机器学习专题研究: Adaboost 算法下的多因子选股》——2016-05-31

联系人: 陈镜竹

电话: 0755-82130833-701336
 E-MAIL: chenjz@guosen.com.cn

证券分析师: 邹璐

电话: 0755-82130833-701418
 E-MAIL: zoulu@guosen.com.cn

证券投资咨询执业资格证书编码: S0980516080005

证券分析师: 黄志文

电话: 0755-82133928
 E-MAIL: huangzw@guosen.com.cn

证券投资咨询执业资格证书编码: S0980510120059

基于 K-Means 聚类的多因子特征检验

● K-Means 聚类

在欧氏距离的基础上, 聚类对于市值和波动率的分割很清晰, 结合簇数以及分割的大小关系, 我们选择将全 A 股分为 5 簇。虽然划分的依据是市值和波动率, 但是划分后的 5 个类别的盈利、增长、换手率等因子的数值特征也是不同的。波动率对于营收增长具有显著的相关性, 高波动率股票的整体营收增长率更高。而对于 ROE 因子, 则可以看到显著的市值特征, 大市值股票的整体 ROE 更高。而对于 EP 和换收益, 市值与波动率特征同时显著。

● 因子特征评价

以 IC 作为聚类检验的标准, 在整体全 A 股的 IC 不显著的因子中, 不同聚类簇的 IC 会体现出高于整体的显著水平。对于换手率、动量、波动率等价值因子, 大市值的因子显著性是低于小市值组合的, 其中以大市值、低波动的组合最为明显。而对于小市值组合中, 高波动率组合的因子显著性较好, 包括财务因子和价值因子都较为明显。对于财务因子而言, 显著性更多的体现在大市值的两个组合之中。

● 策略构建

构建因子打分选股策略, 在同样的选股数量 (均为总数的 30%) 下, 聚类之后的股票多空组合收益显著比全 A 组合稳定。月度的夏普率接近全 A 组合的 2 倍。

独立性声明:

作者保证报告所采用的数据均来自合规渠道, 分析逻辑基于本人的职业理解, 通过合理判断并得出结论, 力求客观、公正, 结论不受任何第三方的授意、影响, 特此声明。

内容目录

引言.....	5
K-Means 聚类.....	5
定义.....	5
过程.....	5
特征因子.....	7
均值方差统计.....	9
聚类因子特征评价.....	11
显著性特征.....	11
因子单向收益.....	15
聚类-因子选股策略.....	16
结论.....	19
国信证券投资评级.....	21
分析师承诺.....	21
风险提示.....	21
证券投资咨询业务的说明.....	21

图表目录

图 1: 聚类分析-随机生成中心	6
图 2: 聚类分析-生成初始簇	6
图 3: 聚类分析-变更中心	6
图 4: 聚类分析-最终分类	6
图 5: 聚类分析结果图示	6
图 6: 全 A 股聚类, cluster=3	7
图 7: 全 A 股聚类, cluster=4	7
图 8: 全 A 股聚类, cluster=5	7
图 9: 全 A 股聚类, cluster=6	7
图 10: 全 A 股聚类, 2011 年 1 月	8
图 11: 全 A 股聚类, 2012 年 1 月	8
图 12: 全 A 股聚类, 2013 年 1 月	8
图 13: 全 A 股聚类, 2014 年 1 月	8
图 14: 全 A 股聚类, 2015 年 1 月	9
图 15: 全 A 股聚类, 2016 年 1 月	9
图 16: 因子 EP	11
图 17: 1 月流动性因子	11
图 18: 因子 ROE	11
图 19: 营收增长率因子	11
图 20: 聚类划分图示	13
图 21: 1 号簇 IC 绝对值排序	14
图 22: 2 号簇 IC 绝对值排序	14
图 23: 3 号簇 IC 绝对值排序	14
图 24: 4 号簇 IC 绝对值排序	14
图 25: 5 号簇 IC 绝对值排序	14
图 26: 全 A 股 IC 绝对值排序	14
图 27: 因子 EP 单向多空收益净值	15
图 28: 因子 BP 单向多空收益净值	15
图 29: 一月换手率因子单向多空收益净值	15
图 30: 一月动量因子单向多空收益净值	15
图 31: ROE 因子单向多空收益净值	16
图 32: 营收增长率因子单向多空收益净值	16
图 33: 等权打分多空收益	17
图 34: 聚类组合收益对比	17
图 35: IC 加权打分多空收益	18
图 36: 聚类组合收益对比	18
图 37: 聚类组合收益对比	19

表 1: 聚类编号	9
表 2: 不同类别组合因子均值、方差	9
表 3: 因子 IC 均值	12
表 4: 聚类因子显著个数统计	16
表 5: 组合收益统计	18
表 6: IC 加权组合收益统计	19

引言

聚类算法是无监督学习中的代表算法，其核心思想是在特征向量构造的空间中，计算观测值之间的距离。距离较近的观测值认为属于同一个聚类（cluster），距离较远的观测值则认为属于不同的聚类。

与多因子结合起来，我们认为不同因子特征的股票组合，应该具有不同的收益特征。如果直接与收益关联，则类似于我们前期报告尝试过的 SVM 等监督学习算法，本篇报告希望转换一个新的角度，通过因子特征分析，不与收益直接关联，而是反过来检验聚类后的风险因子特征，利用聚类这种无监督学习的思想，观察不同聚类簇的风险因子特征是否具有差异。

K-Means 聚类

定义

K-Means 算法是最为经典的基于划分的聚类方法，其基本思想是：以空间中 k 个点为中心进行聚类，对最靠近他们的对象归类。通过迭代的方法，逐次更新各聚类中心的值，直至得到最好的聚类结果。

在模型训练中，K-平均聚类分析的迭代算法属于最大期望算法（Expectation Maximization, EM）中的一种。它以聚类内部距离的平方和最小为目标函数（objective function），通过迭代计算目标函数和聚类中心（centroid）来寻找目标函数的极小值。

给定需要聚类的簇的数量 k，则目标函数为：

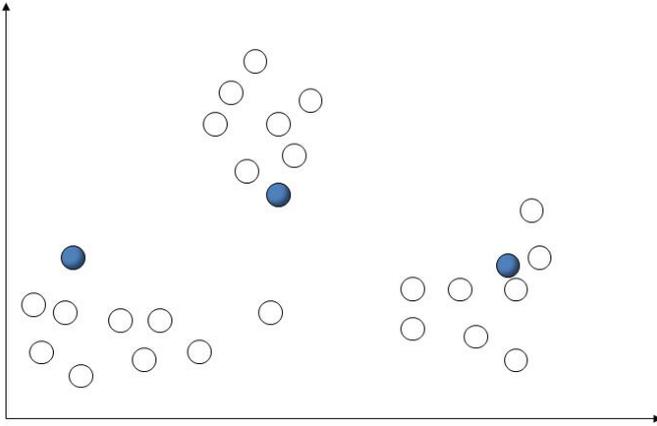
$$\operatorname{argmin} \sum_{i=1}^k \sum_{x \in S_i} \|x - u_i\|^2$$

其中 u_i 是簇 S_i 的均值（中心）。

过程

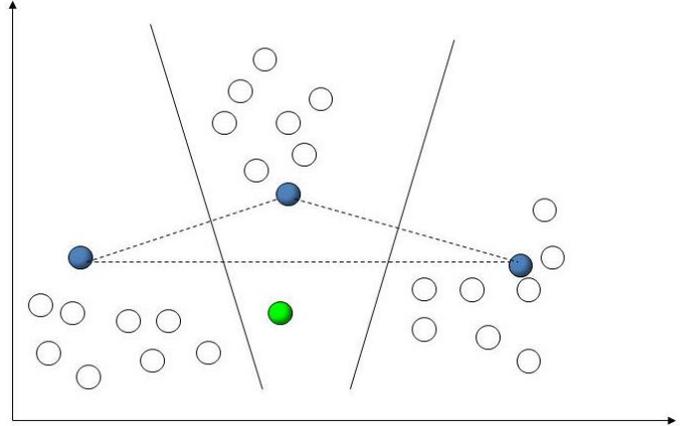
1. 在样本中随机生成 k 个点作为初始中心。
2. 计算所有样本点距离 k 个中心点的距离（一般采用欧氏距离），以距离最近对各个样本点做划分。
3. 根据划分出的簇，计算 k 个簇的中心，取代初始的 k 个中心。
4. 以新的中心重新划分簇，并通过反复迭代，找到最后稳定的中心点。
5. 以最终 k 个中心点，划分样本点，得到结果。

图 1: 聚类分析-随机生成中心



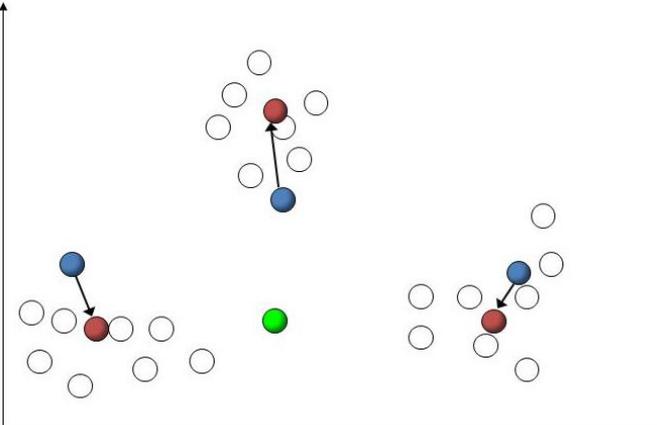
资料来源: wikidot

图 2: 聚类分析-生成初始簇



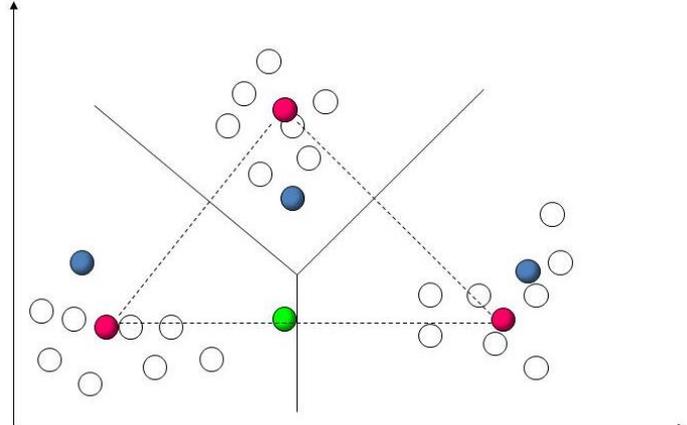
资料来源: wikidot

图 3: 聚类分析-变更中心



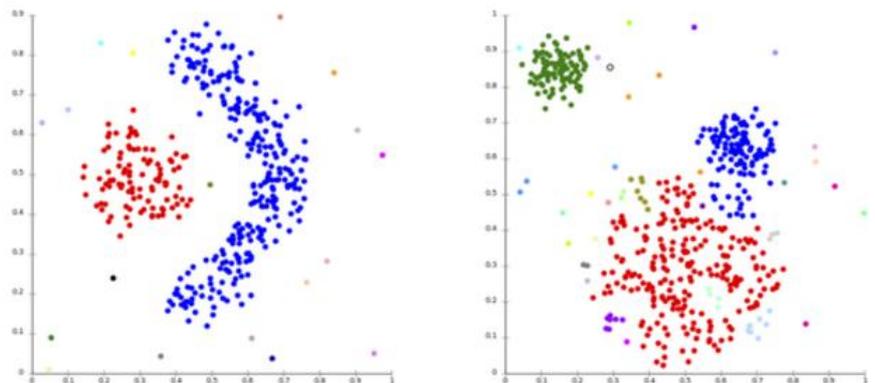
资料来源: wikidot

图 4: 聚类分析-最终分类



资料来源: wikidot

图 5: 聚类分析结果图示



资料来源: Wki 百科

特征因子

特征因子的选择实际上是聚类（乃至机器学习领域）的核心工作，特征工程是一个长期的过程，本篇报告将略过这一过程，利用我们之前的研究结果，直接给定所需的特征因子。

结合报告《国信技术面量化选股系列：价格路径对收益的影响》，我们实际上在波动率因子上观察到了与收益较为显著的关系。虽然在开篇已经说明，本次报告选择特征因子的思路并不与收益直接关联，但我们仍然希望通过与收益具有稳定相关性的特征分类，进而获得不同的风险因子特征。

另一维的特征因子我们选择市值因子，同样是基于市值因子与收益率的显著关系。

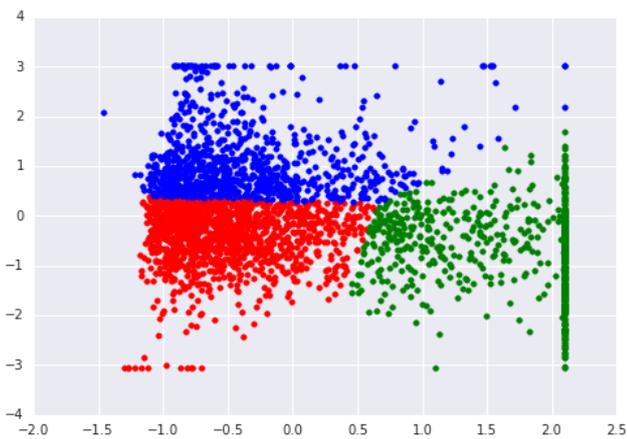
以 2016 年 9 月的全 A 股数据作为样本：

X 轴：总市值，

Y 轴：股票过去一年的波动率，日频，

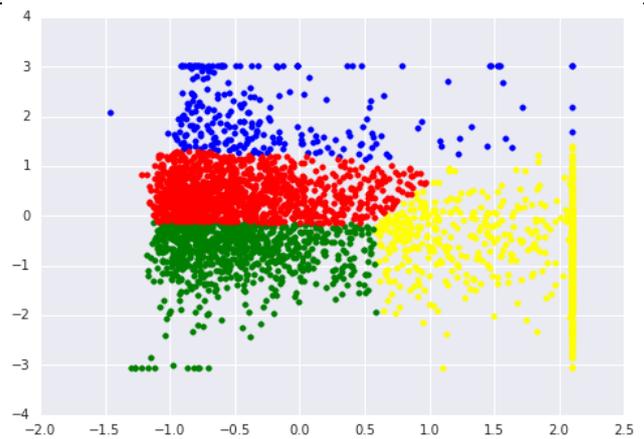
经过去极值、标准化之后的聚类结果如下：

图 6: 全 A 股聚类, duster=3



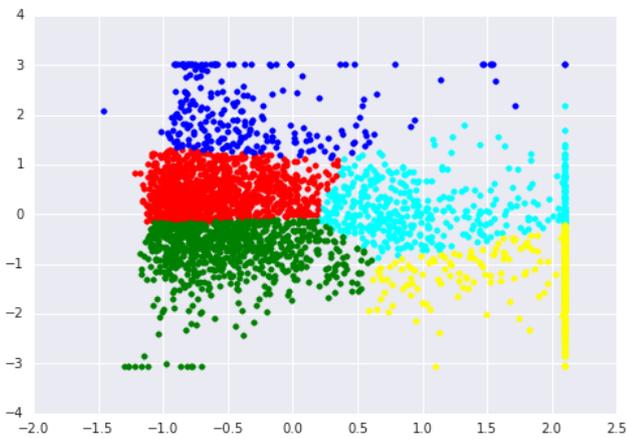
资料来源: wind, 国信证券经济研究所整理

图 7: 全 A 股聚类, duster=4



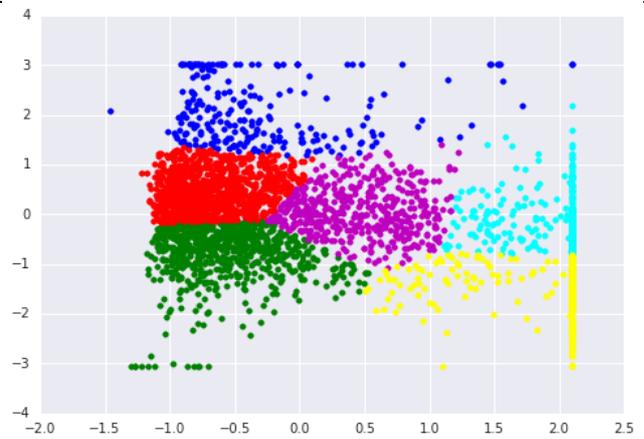
资料来源: wind, 国信证券经济研究所整理

图 8: 全 A 股聚类, duster=5



资料来源: wind, 国信证券经济研究所整理

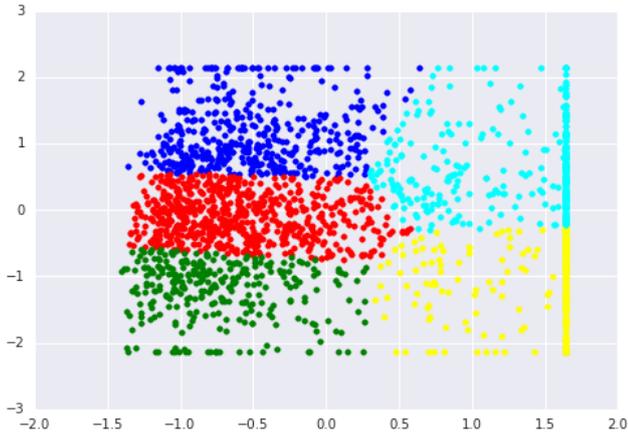
图 9: 全 A 股聚类, duster=6



资料来源: wind, 国信证券经济研究所整理

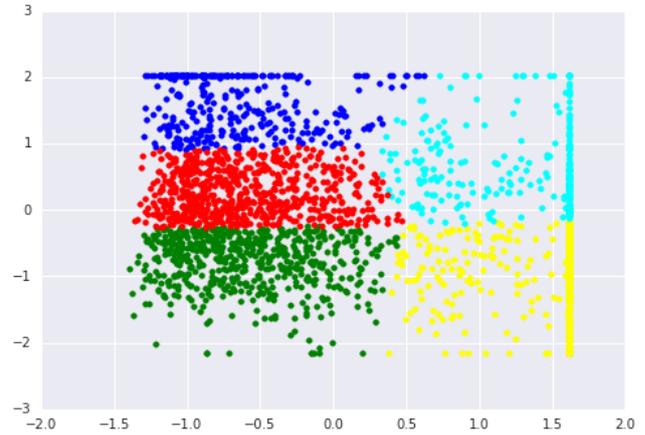
可以看到，在欧氏距离的基础上，聚类对于市值和波动率的分割很清晰，结合簇数以及分割的大小关系，我们最终选择将全 A 股分为 5 簇，也就是取 $k=5$ 。接着查看历年的 $k=5$ 时的聚类结果：

图 10: 全 A 股聚类, 2011 年 1 月



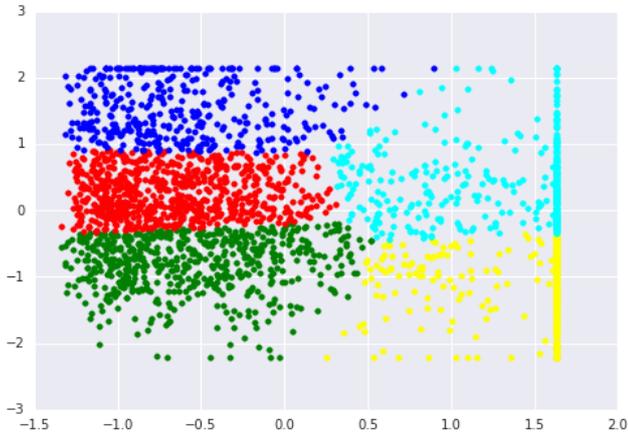
资料来源: wind, 国信证券经济研究所整理

图 11: 全 A 股聚类, 2012 年 1 月



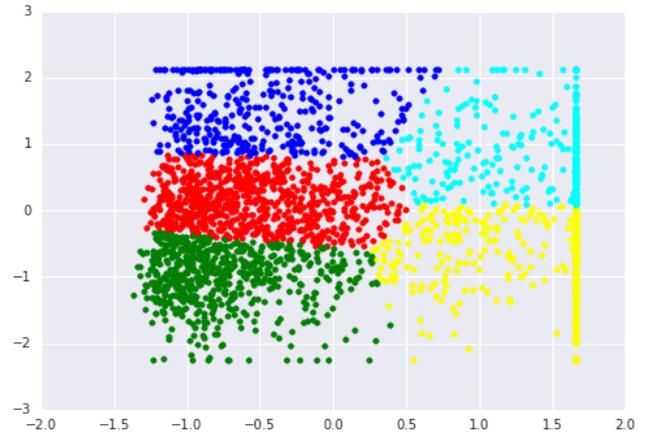
资料来源: wind, 国信证券经济研究所整理

图 12: 全 A 股聚类, 2013 年 1 月



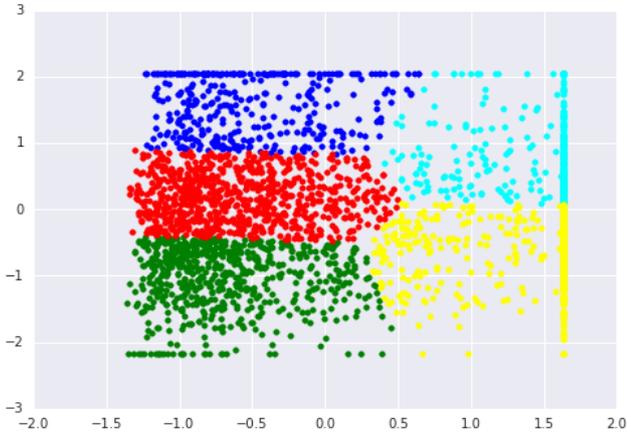
资料来源: wind, 国信证券经济研究所整理

图 13: 全 A 股聚类, 2014 年 1 月



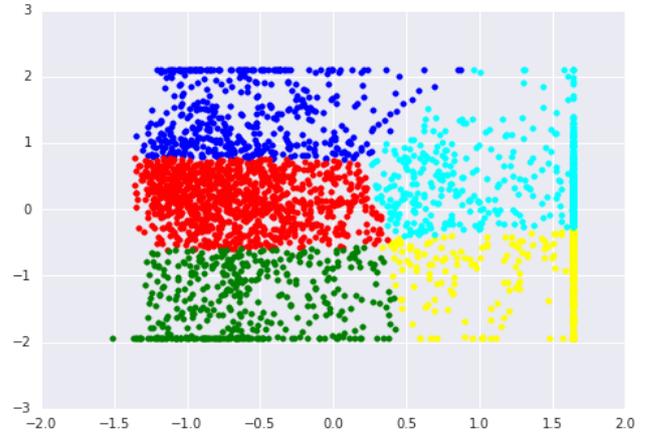
资料来源: wind, 国信证券经济研究所整理

图 14: 全 A 股聚类, 2015 年 1 月



资料来源: wind, 国信证券经济研究所整理

图 15: 全 A 股聚类, 2016 年 1 月



资料来源: wind, 国信证券经济研究所整理

从各年的聚类结果来看, 划分 5 类的效果也是稳定的, 按照从左至右, 从上之至下的顺序编号:

表 1: 聚类编号

聚类	市值特征	波动率特征
1	低	高
2	低	中
3	低	低
4	高	高
5	高	低

资料来源: 国信证券经济研究所整理

均值方差统计

将股票组合按照最新一期的市值、波动率划分之后, 不同类别间的因子统计特征呈现出差异化, 以国信金工多因子 (详细因子定义请参考国信金工多因子系列报告) 的 69 个因子数据来看:

表 2: 不同类别组合因子均值、方差

	均值 (0)					方差 (1)				
	1	2	3	4	5	1	2	3	4	5
ETOP	-0.36	-0.31	0.02	0.09	0.87	0.51	0.82	1.06	0.84	1.20
ETP5	-0.46	-0.31	0.05	-0.04	0.82	0.62	0.75	1.08	0.81	1.16
EXTE	-0.12	-0.01	0.04	0.09	-0.04	1.00	1.11	1.05	0.92	0.74
VFLO	0.02	-0.02	-0.11	0.23	-0.09	1.10	1.05	1.03	0.98	0.73
VERN	-0.02	-0.07	-0.03	0.20	-0.03	1.05	1.07	1.04	0.95	0.73
PAYO	-0.09	0.01	-0.02	-0.08	0.17	0.99	1.11	0.98	0.91	0.86
VCAP	0.16	-0.03	-0.16	0.32	-0.16	1.10	0.96	0.89	1.12	0.91
AGRO	0.11	-0.20	-0.30	0.50	0.19	1.05	0.87	0.84	1.16	0.96
EGRO	-0.16	-0.15	-0.03	0.13	0.34	0.95	1.02	1.01	1.02	0.84
DELE	-0.01	-0.05	-0.10	0.11	0.09	0.90	1.05	1.03	1.08	0.83
S_SalseG	0.20	-0.06	-0.26	0.32	-0.03	1.08	0.94	0.88	1.16	0.90
C_SalseG	0.16	-0.07	-0.28	0.35	-0.02	1.04	0.94	0.89	1.15	0.90
T_SalseG	0.19	-0.05	-0.25	0.28	-0.03	1.05	0.93	0.90	1.17	0.93
S_ProfitG	0.01	0.00	-0.15	0.23	-0.06	1.08	1.01	1.02	1.02	0.79
C_ProfitG	0.06	-0.02	-0.19	0.26	-0.07	1.09	1.02	0.97	1.03	0.78

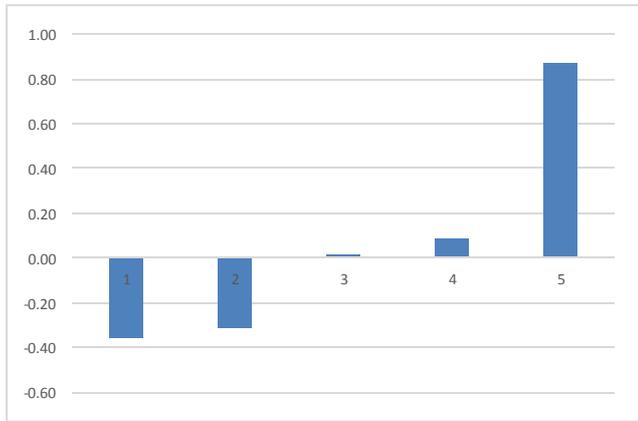
T_ProfitG	0.03	0.01	-0.10	0.16	-0.09	1.00	1.03	1.07	0.97	0.82
S_CFOG	-0.02	-0.06	0.15	-0.07	0.01	1.11	1.02	0.94	1.08	0.81
C_CFOG	-0.07	-0.01	0.09	-0.03	0.00	1.06	1.04	0.94	1.04	0.86
T_CFOG	-0.02	0.00	0.12	-0.07	-0.06	1.04	1.02	0.95	1.08	0.86
S_ROEG	-0.05	0.05	-0.08	0.10	-0.08	1.06	1.06	1.05	0.96	0.75
C_ROEG	0.00	0.04	-0.11	0.14	-0.09	1.08	1.06	1.01	1.01	0.71
T_ROEG	0.00	0.03	-0.05	0.05	-0.04	1.01	1.04	1.05	1.00	0.80
S_ROAG	-0.05	0.05	-0.08	0.10	-0.07	1.03	1.05	1.06	0.98	0.76
C_ROAG	0.00	0.03	-0.10	0.13	-0.08	1.05	1.05	1.02	1.01	0.74
T_ROAG	0.00	0.04	-0.06	0.06	-0.08	1.01	1.04	1.07	0.98	0.81
MLEV	-0.53	-0.13	0.22	0.07	0.48	0.66	0.91	1.05	1.02	1.08
BLEV	-0.37	-0.13	0.11	0.15	0.37	0.79	0.92	1.04	1.04	1.08
DTOA	-0.31	-0.05	0.12	0.12	0.15	0.92	0.96	0.98	1.02	1.09
STO_1M	0.89	0.15	-0.26	-0.16	-0.77	1.14	0.94	0.79	0.83	0.44
STO_3M	1.13	0.13	-0.36	-0.11	-0.83	1.11	0.86	0.70	0.82	0.42
STO_6M	1.31	0.16	-0.44	-0.07	-0.91	1.03	0.82	0.58	0.83	0.35
STO_12M	1.33	0.29	-0.49	-0.07	-1.02	1.00	0.80	0.49	0.83	0.35
STO_60M	0.94	0.41	-0.18	-0.16	-0.87	1.01	0.96	0.77	0.83	0.62
HALPHA	0.81	0.01	-0.52	0.45	-0.24	0.82	0.80	0.81	1.14	1.00
RSTR_1M	-0.34	0.14	0.24	-0.17	-0.06	1.24	0.96	0.90	1.03	0.67
RSTR_3M	-0.54	0.06	0.20	-0.03	0.19	1.22	0.95	0.83	1.10	0.69
RSTR_6M	0.02	0.08	0.00	-0.03	-0.15	1.29	0.95	0.84	1.09	0.85
RSTR_12M	0.74	0.07	-0.21	-0.04	-0.46	1.13	0.92	0.81	1.07	0.80
LNCAP	-0.47	-0.63	-0.50	0.88	1.55	0.46	0.46	0.49	0.57	0.84
LNCAPCB	-0.47	-0.63	-0.50	0.88	1.55	0.46	0.46	0.49	0.57	0.84
BTOP	-0.71	-0.15	0.43	-0.22	0.79	0.43	0.73	1.06	0.87	1.23
STOP	-0.59	-0.13	0.26	-0.12	0.58	0.55	0.90	1.03	0.96	1.11
EBITDAvsEV	-0.59	-0.27	0.27	-0.03	0.76	0.41	0.74	1.04	0.87	1.32
HILO	0.56	0.04	-0.23	0.09	-0.51	1.12	0.93	0.98	0.95	0.71
BTSG	0.91	0.22	-0.49	0.40	-0.66	0.96	0.77	0.74	1.02	0.94
DASTD	0.64	0.14	-0.31	0.07	-0.63	1.01	0.95	0.97	0.87	0.75
LPRI	0.81	-0.07	-0.45	0.13	-0.31	0.98	0.80	0.88	0.87	1.14
CMRA	1.06	0.28	-0.55	0.17	-0.96	0.95	0.75	0.68	0.89	0.72
VOLBT	-0.38	-0.11	0.09	0.01	0.44	0.91	0.96	1.02	0.97	0.99
SERDP	0.15	-0.13	-0.05	0.21	0.01	0.93	0.96	1.00	1.03	1.03
BETA	0.35	0.12	-0.25	0.18	-0.29	1.03	0.86	0.89	1.11	1.09
SIGMA	0.97	0.16	-0.48	0.43	-0.65	0.84	0.80	0.78	1.05	0.88
YLD	-0.37	-0.22	0.06	-0.04	0.87	0.68	0.79	1.05	0.93	1.19
YLD3	-0.50	-0.19	0.13	-0.10	0.93	0.57	0.83	1.06	0.84	1.17
S_GPM	0.10	-0.11	-0.11	0.14	0.15	0.97	0.91	1.03	1.09	1.01
C_GPM	0.13	-0.12	-0.12	0.15	0.14	0.95	0.91	1.01	1.10	1.03
T_GPM	0.10	-0.12	-0.09	0.14	0.14	0.95	0.91	1.01	1.11	1.03
S_NPM	-0.07	-0.19	-0.20	0.27	0.41	0.97	0.92	0.99	1.01	0.99
C_NPM	-0.05	-0.20	-0.19	0.27	0.44	0.96	0.91	0.97	1.02	1.02
T_NPM	-0.07	-0.22	-0.17	0.31	0.41	0.97	0.94	0.91	1.02	1.04
S_CTP	-0.19	-0.03	0.20	-0.16	0.13	0.98	1.05	1.01	0.96	0.89
C_CTP	-0.23	0.00	0.17	-0.14	0.15	0.96	1.06	0.95	0.99	0.92
T_CTP	-0.04	0.04	0.19	-0.22	-0.01	0.92	1.06	1.00	0.93	0.97
S_ROE	-0.14	-0.17	-0.20	0.23	0.49	0.93	0.96	0.99	0.99	0.95

资料来源：天软，wind，国信证券经济研究所整理

由于各因子值都经过了标准化处理，因此因子整体的均值为 0，方差为 1，但在表格中可以看到，划分 5 类的股票组合的因子均值和方差都存在明显区别于整体的差异。

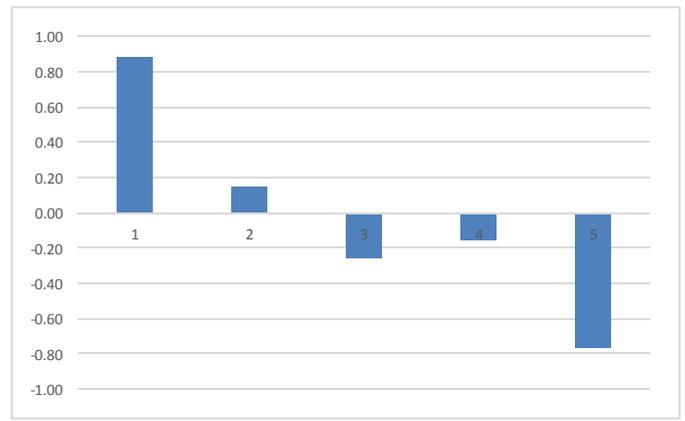
分别取 ROE、EP、1 月换手率、营收增长率 4 个不同类型的因子，观察不同组别的因子特征：

图 16: 因子 EP



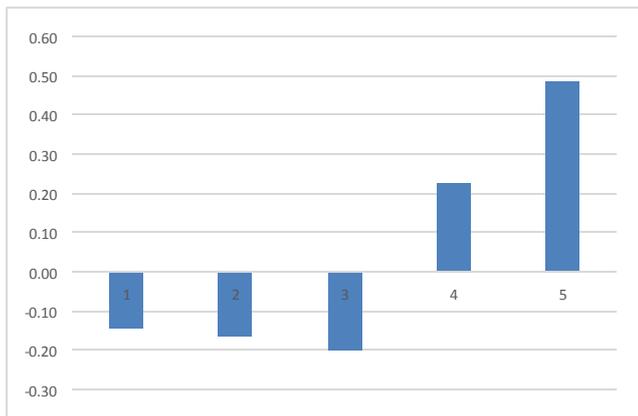
资料来源: wind, 国信证券经济研究所整理

图 17: 1月流动性因子



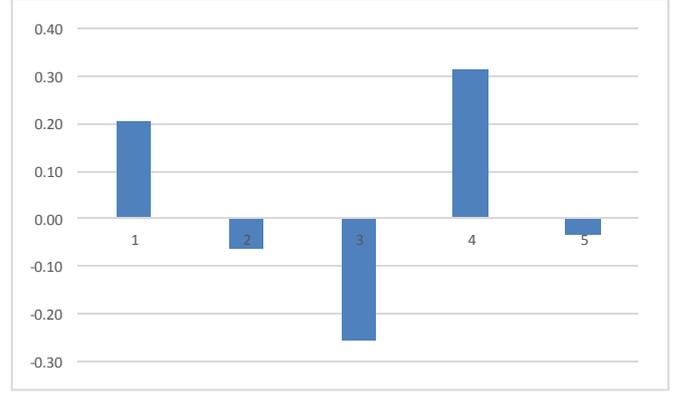
资料来源: wind, 国信证券经济研究所整理

图 18: 因子 ROE



资料来源: wind, 国信证券经济研究所整理

图 19: 营收增长率因子



资料来源: wind, 国信证券经济研究所整理

可以看到, 虽然划分的依据是市值和波动率, 但是划分后的不同类别的盈利、增长、换手率等因子的数值特征也是不同的。以营收增长因子为例, 波动率对于营收增长具有显著的相关性, 高波动率股票的整体营收增长率更高。而对于 ROE 因子, 则可以看到显著的市值特征, 大市值股票的整体 ROE 更高。而对于 EP 和换收益, 市值与波动率特征同时显著。

聚类因子特征评价

从前文可以看到, 通过聚类划分不同的簇之后, 不同簇之间的因子统计特征呈现差异化。但实际上, 部分统计特征的不同从因子的逻辑本身是有迹可循的。例如大市值的公司 ROE 可能要高于小市值公司, 高波动的股票换手率更高。通过聚类分析, 更重要的目的是希望得到更加深入的因子特征, 而不是因子间的相关性。因此, 我们进一步希望观察不同簇之间因子的收益特征是否也存在差异性。

显著性特征

以因子的 IC 均值作为观察显著性的指标, 2009 年-2016 年的月 IC 均值统计如

下:

表 3: 因子 IC 均值

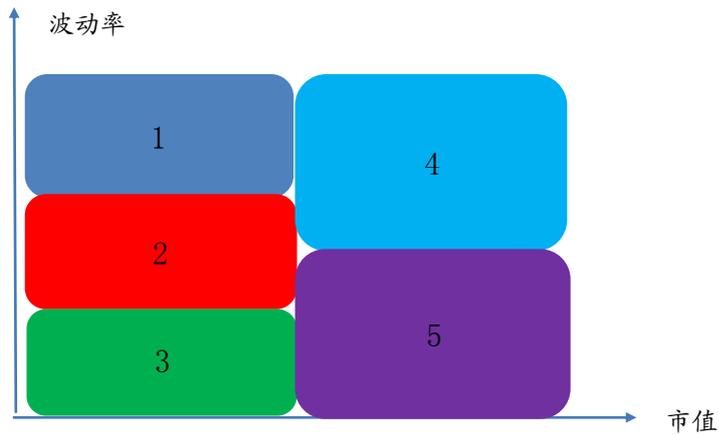
	1	2	3	4	5	全 A
ETOP	1.93%	0.38%	-0.39%	1.76%	2.71%	-0.07%
ETP5	0.12%	-1.54%	-2.43%	0.40%	-0.28%	-2.12%
EXTE	0.44%	-0.27%	-0.22%	-0.21%	-0.21%	-0.13%
VFLO	-0.09%	0.06%	-0.04%	-0.07%	0.14%	-0.15%
VERN	-0.26%	0.04%	0.51%	-0.89%	-0.41%	-0.20%
PAYO	0.69%	-0.19%	0.38%	-0.51%	0.35%	-0.09%
VCAP	-0.05%	-0.65%	-0.12%	-2.07%	-2.11%	-1.32%
AGRO	2.40%	1.36%	0.93%	1.07%	0.08%	-0.29%
EGRO	-0.54%	-0.70%	-0.69%	0.87%	1.51%	-0.88%
DELE	0.10%	-0.09%	-0.17%	-0.05%	0.75%	-0.36%
S_SalseG	2.12%	0.16%	0.86%	1.69%	1.58%	0.11%
C_SalseG	1.91%	0.52%	0.32%	-0.15%	-0.14%	-0.02%
T_SalseG	3.26%	1.08%	1.48%	1.66%	2.38%	0.76%
S_ProfitG	0.79%	-0.38%	0.49%	0.67%	0.46%	-0.11%
C_ProfitG	0.75%	0.03%	0.18%	0.10%	0.87%	-0.09%
T_ProfitG	0.37%	0.38%	1.31%	0.72%	1.71%	0.15%
S_CFOG	0.95%	0.13%	1.02%	0.13%	0.54%	0.18%
C_CFOG	0.63%	-0.15%	2.52%	0.33%	0.13%	0.02%
T_CFOG	0.05%	1.11%	-0.26%	-0.53%	0.78%	0.46%
S_ROEG	0.39%	-0.33%	0.07%	1.76%	0.98%	0.18%
C_ROEG	-0.04%	-0.22%	0.04%	1.71%	1.35%	0.29%
T_ROEG	0.36%	0.49%	1.24%	1.27%	1.70%	0.39%
S_ROAG	0.66%	-0.17%	0.44%	1.64%	0.98%	0.19%
C_ROAG	0.49%	0.18%	0.26%	0.85%	1.22%	0.44%
T_ROAG	0.04%	0.32%	1.18%	1.16%	1.61%	0.21%
MLEV	0.38%	-1.30%	-1.41%	-0.02%	-0.97%	-1.09%
BLEV	-0.09%	-1.22%	-1.55%	-1.25%	-2.51%	-1.63%
DTOA	-1.49%	-1.56%	-1.21%	-1.01%	-1.59%	-1.54%
STO_1M	-7.55%	-6.13%	-6.77%	-4.78%	-2.03%	-4.70%
STO_3M	-4.02%	-3.77%	-5.11%	-2.97%	-0.62%	-2.36%
STO_6M	-3.32%	-2.33%	-4.35%	-3.19%	-0.75%	-1.69%
STO_12M	-1.40%	-0.67%	-1.17%	-1.18%	-0.44%	-0.53%
STO_60M	-1.00%	0.09%	-0.28%	-0.85%	-0.63%	0.78%
HALPHA	-4.10%	-1.63%	-1.39%	-1.55%	-1.49%	-3.49%
RSTR_1M	-7.21%	-6.17%	-3.73%	-5.15%	-2.75%	-5.51%
RSTR_3M	-7.19%	-6.64%	-3.88%	-3.87%	-1.85%	-5.86%
RSTR_6M	-5.09%	-5.19%	-3.31%	-3.35%	-2.28%	-4.81%
RSTR_12M	-4.83%	-3.78%	-2.58%	-2.65%	-0.47%	-4.42%
LNCAP	-7.96%	-7.80%	-9.76%	-3.98%	-3.16%	-8.09%
LNCAPCB	-7.96%	-7.80%	-9.76%	-3.98%	-3.16%	-8.09%
BTOP	3.81%	0.89%	-0.65%	0.61%	1.09%	1.20%
STOP	0.71%	0.30%	0.63%	0.09%	0.93%	0.52%
EBITDAvsEV	0.28%	-0.48%	-1.28%	0.09%	-0.74%	-0.41%
HILO	-2.49%	-2.46%	-3.09%	-1.35%	-0.46%	-2.09%
BTSG	-1.25%	0.35%	1.47%	-1.49%	-1.60%	-1.06%
DASTD	-3.37%	-3.73%	-1.56%	-1.29%	-0.07%	-1.68%
LPRI	-1.50%	-0.70%	-0.31%	-0.26%	-1.06%	-2.40%
CMRA	0.35%	1.36%	1.65%	-1.42%	-1.42%	-0.58%
VOLBT	-0.91%	-1.14%	0.51%	-1.05%	-0.70%	-0.07%
SERDP	-1.49%	-0.92%	-2.64%	-0.34%	-1.09%	-1.48%
BETA	-1.27%	-0.96%	1.13%	-1.57%	-1.64%	-1.56%
SIGMA	-1.17%	1.27%	1.67%	-0.31%	-0.21%	-0.34%
YLD	1.13%	0.30%	-0.23%	0.87%	1.42%	-0.30%

YLD3	1.37%	0.68%	-0.69%	0.92%	1.31%	-0.36%
S_GPM	1.68%	1.80%	0.90%	2.17%	2.48%	1.00%
C_GPM	1.77%	1.72%	1.30%	2.31%	2.22%	0.91%
T_GPM	1.59%	1.42%	0.77%	2.51%	2.04%	0.77%
S_NPM	1.90%	0.40%	0.59%	0.09%	1.17%	0.02%
C_NPM	-0.14%	0.18%	0.28%	-0.16%	1.07%	-0.01%
T_NPM	-0.36%	0.39%	-0.27%	0.40%	0.79%	-0.25%
S_CTP	-1.21%	-0.55%	1.23%	0.10%	0.08%	-0.14%
C_CTP	0.50%	-0.16%	1.92%	0.28%	0.56%	0.16%
T_CTP	-0.34%	-0.47%	0.12%	-1.02%	-0.08%	-0.21%
S_ROE	2.42%	0.43%	0.03%	3.24%	4.80%	0.19%
C_ROE	-0.16%	-0.01%	1.02%	2.09%	3.19%	-0.22%
T_ROE	-0.61%	0.11%	0.29%	1.64%	2.76%	-0.20%
S_ROA	2.08%	1.81%	0.23%	2.95%	4.11%	0.61%
C_ROA	0.60%	0.59%	1.50%	1.76%	2.50%	0.39%
T_ROA	0.37%	0.29%	-0.05%	1.55%	1.38%	-0.35%

资料来源：天软，wind，国信证券经济研究所整理

可以看到，在整体全 A 股的 IC 不显著的因子中，不同聚类簇的 IC 会体现出高于整体的显著水平。

图 20: 聚类划分图示

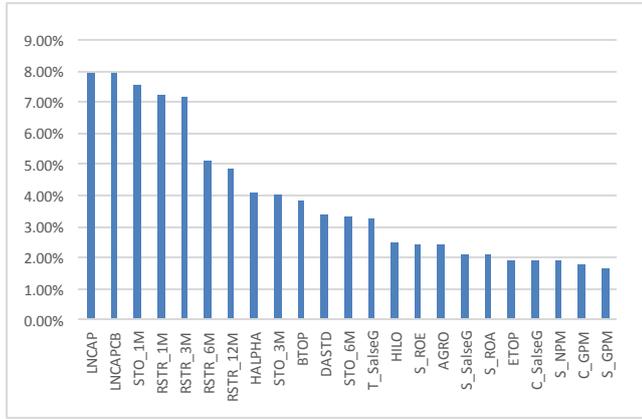


资料来源：国信证券经济研究所整理

EP、BP 两个因子在 1 号类别下的显著性十分明显，而这两个因子整体显著性较差。成长因子、财务质量因子更多的在 1、4、5 类的效果显著，而波动率因子在 1、2、3 类的效果更显著。动量、换手率、规模因子则在所有的类别中都具有显著效果，但相对而言第 5 类的效果较差。

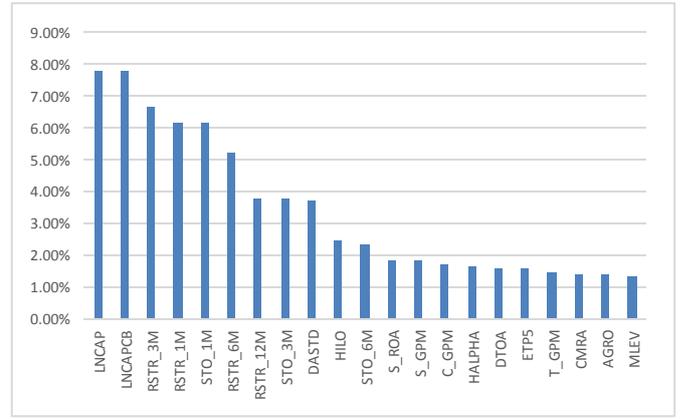
按照 IC 的绝对值排序，各类别因子的 IC 排列如下：

图 21: 1号簇 IC 绝对值排序



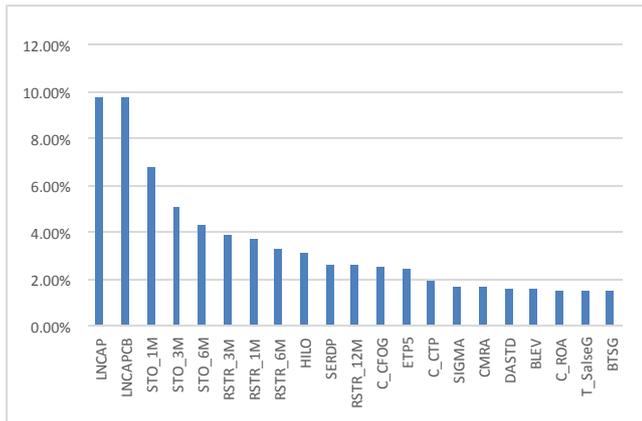
资料来源: 天软, wind, 国信证券经济研究所整理

图 22: 2号簇 IC 绝对值排序



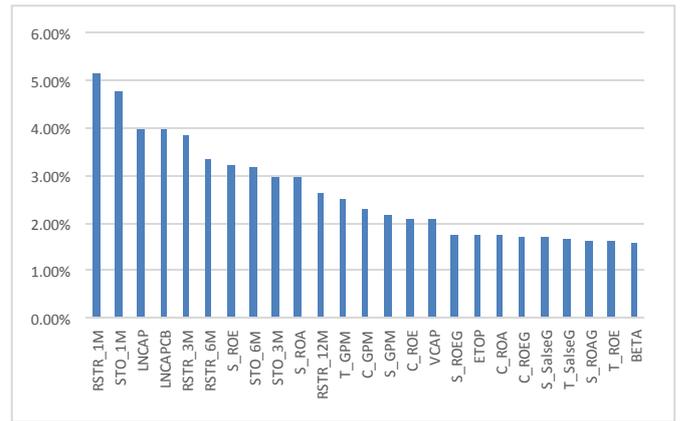
资料来源: 天软, wind, 国信证券经济研究所整理

图 23: 3号簇 IC 绝对值排序



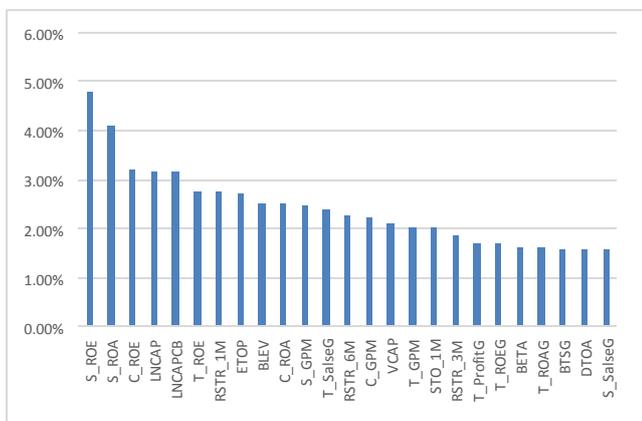
资料来源: 天软, wind, 国信证券经济研究所整理

图 24: 4号簇 IC 绝对值排序



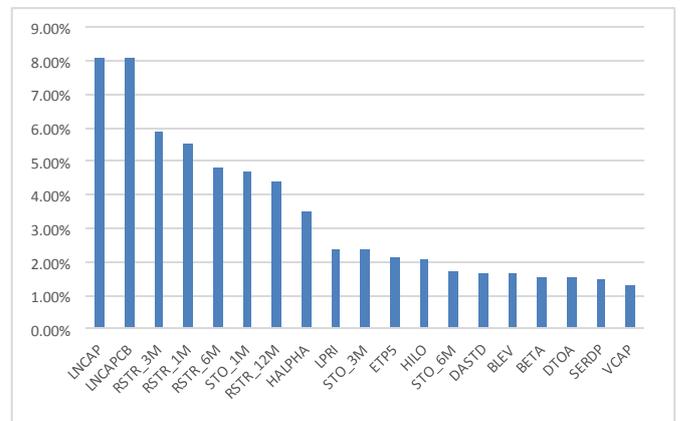
资料来源: 天软, wind, 国信证券经济研究所整理

图 25: 5号簇 IC 绝对值排序



资料来源: 天软, wind, 国信证券经济研究所整理

图 26: 全 A 股 IC 绝对值排序



资料来源: 天软, wind, 国信证券经济研究所整理

因子单向收益

更加直接的观察不同簇间因子的收益结果，我们将单因子按照排序打分的方式进行选择：

回测区间：2009-2016，

多空比例：前 40%对后 40%。

选择较为典型的多空收益曲线如下：

图 27: 因子 EP 单向多空收益净值



资料来源：天软，wind，国信证券经济研究所整理

图 28: 因子 BP 单向多空收益净值



资料来源：天软，wind，国信证券经济研究所整理

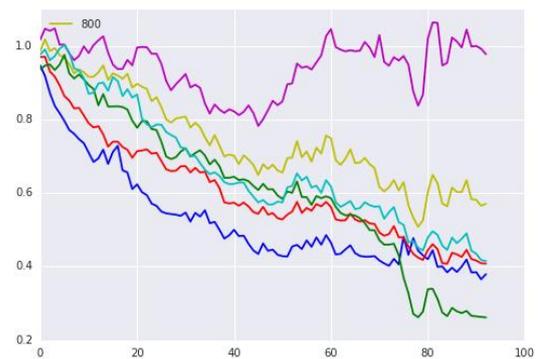
对于 EP、BP 两个因子，全 A 整体实际上并没有显著的超额收益，但 1 号类型的小市值高波动率股票组合却在这两个因子有显著的正向收益。

图 29: 一月换手率因子单向多空收益净值



资料来源：天软，wind，国信证券经济研究所整理

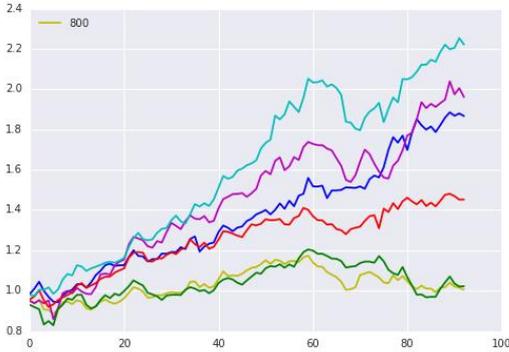
图 30: 一月动量因子单向多空收益净值



资料来源：天软，wind，国信证券经济研究所整理

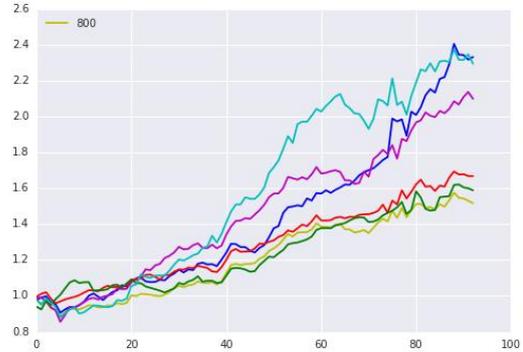
对于换手率、动量因子，5号类型的大市值低波动率的股票组合则稳定性较差，其它类别的组合因子收益率较为显著，且部分能否超过全 A 股整体的因子收益。

图 31: ROE 因子单向多空收益净值



资料来源: 天软, wind, 国信证券经济研究所整理

图 32: 营收增长率因子单向多空收益净值



资料来源: 天软, wind, 国信证券经济研究所整理

对于 ROE、营收增长率等因子, 则能够观察到 5 个类别的因子单向收益均高于整体, 而高波动率的 1、4, 以及大市值的 5 号类别因子收益率更高。

表 4: 聚类因子显著个数统计

	1	2	3	4	5	全 A
Earning Yield	1		1	1	1	1
Earning Variability						
Growth	4		1	6	6	1
Leverage					2	1
Liquidity	3	3	3	3	1	3
Momentum	5	5	4	4	3	5
Size	2	2	2	2	2	2
Value	1					
Volatility	2	2	3			3
Dividend Yield						
Financial Quality	5	3		7	7	

资料来源: 国信证券经济研究所整理

上述统计表并不能完全反应出各个类别的因子特征, 但整体的特点比较明确: 对于换手率、动量、波动率等价值因子, 大市值的因子显著性是低于小市值组合的, 其中以大市值、低波动的组合最为明显。而对于小市值组合中, 高波动率组合的因子显著性较好, 包括财务因子和价值因子都较为明显。对于财务因子而言, 显著性更多的体现在大市值的两个组合之中。

聚类-因子选股策略

总结前文我们观察到的因子特征, 聚类后的各个类别的因子收益与总体的因子收益存在 2 种关系:

1. 整体因子不显著, 但聚类中的某一类在该因子具有显著性, 如 EP、BP 因子, 在高波动, 小市值类别下的因子显著性较高, 但整体效果不明显。
2. 整体因子显著, 聚类后的各类组合同样具有显著效果, 且部分类别的显著性高于整体, 如营收增长率因子, 在大市值的类别下因子显著性高于其他类别。

这两种关系, 实际上都是能够使得聚类后的因子收益总和超过整体的因子收益的。

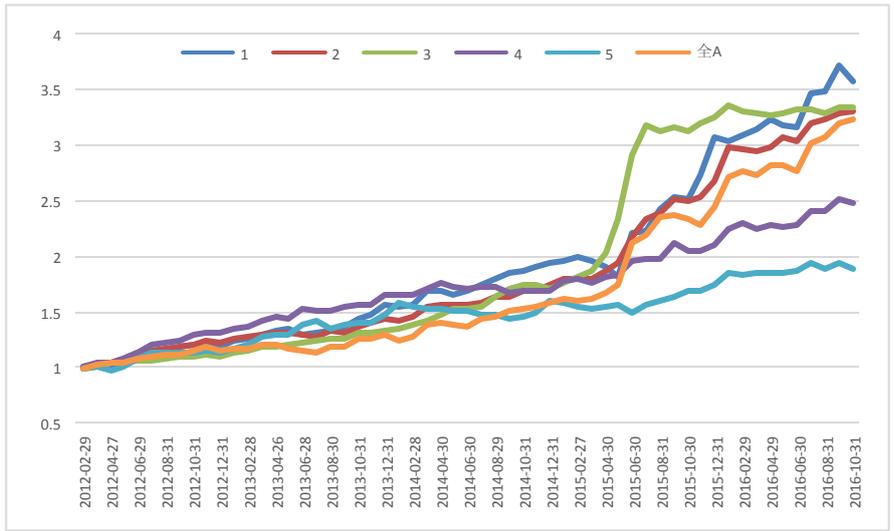
因此我们先尝试构建因子等权打分排序的选股策略:

1. 以 2009-2011 年, 三年的因子 IC 的均值, 多空组合的胜率作为标准, 选择各

个聚类组合的有效因子。

2. 根据每个月新的因子分数，按照大类因子等权相加得到总分数，选取分数的前后 30% 的股票最为多空组合。

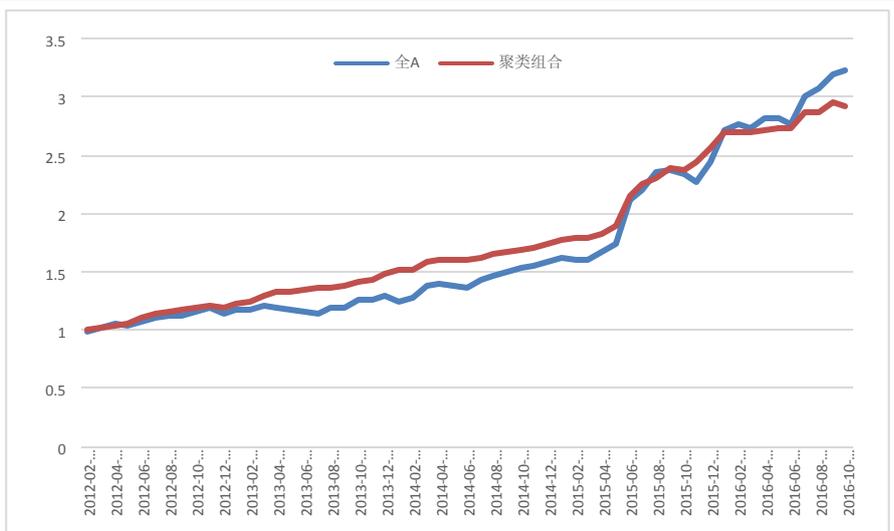
图 33: 等权打分多空收益



资料来源: 天软, 国信证券经济研究所整理

从多空组合的净值曲线来看，不同类别的市值效应比较明显，小市值的 1、2、3 号组合的多空收益高于全 A 股组合。由于不同组合选择的因子并不相同，直接的收益比较实际上是有了一定的局限性的，我们将 5 个组合的收益加总：

图 34: 聚类组合收益对比



资料来源: 天软, 国信证券经济研究所整理

可以看到，在同样的选股数量（均为总数的 30%）下，聚类之后的股票多空组合收益显著比全 A 组合稳定。月度的夏普率接近全 A 组合的 2 倍。

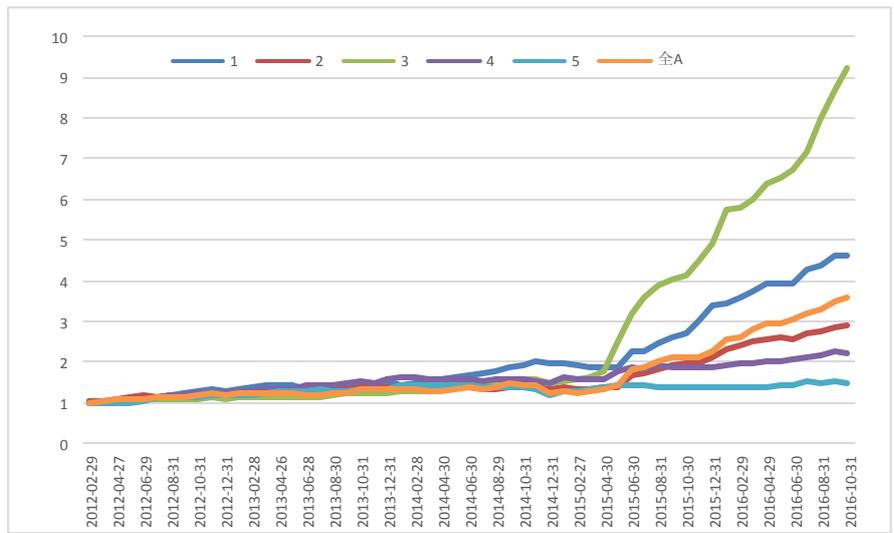
表 5: 组合收益统计

	月均收益	月波动率	夏普率
全 A	2.15%	4.11%	1.81
聚类加权	1.92%	2.28%	2.91

资料来源：天软，国信证券经济研究所整理

由于等权忽略了不同因子的显著性水平，进一步，我们按照年度滚动的方式，以过去 12 个月的因子平均 IC 值作为因子分数的权重，对聚类组合进行多空收益回测：

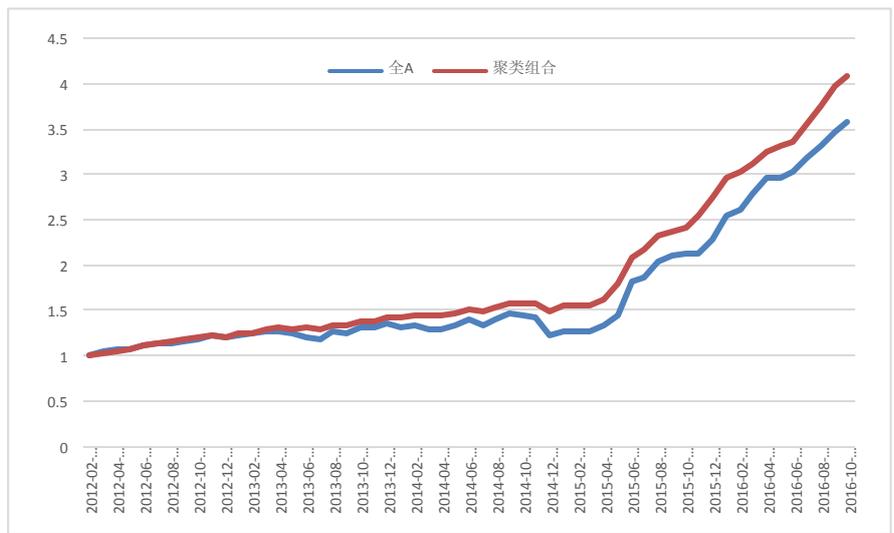
图 35: IC 加权打分多空收益



资料来源：天软，国信证券经济研究所整理

可以看到通过 IC 加权的多空收益本身有所提高，而小市值的 3 个组合的收益提升最为明显。同样将聚类的 5 个组合加权，与全 A 组合对比：

图 36: 聚类组合收益对比



资料来源：天软，国信证券经济研究所整理

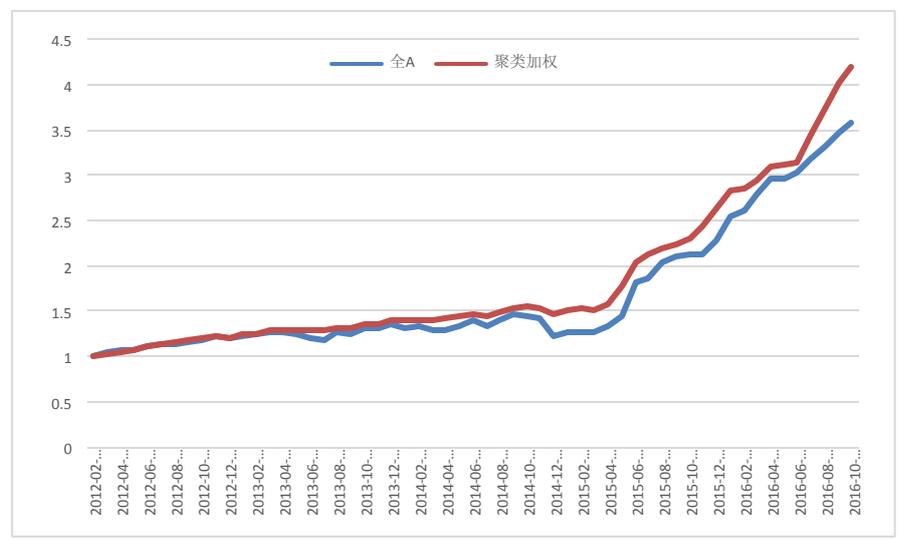
表 6: IC 加权组合收益统计

	月均收益	月波动率	夏普率
全 A	2.38%	5.01%	1.65
聚类加权	2.55%	3.19%	2.77

资料来源：天软，国信证券经济研究所整理

再进一步考虑，由于股票的多空收益是等权计算的，聚类加权的组合收益中，小市值部分的权重在等权相加的情况下实际上是被稀释了，股票个数较少的大市值组合被赋予了更多的资金权重。因此，我们以聚类组合的平均股票个数作为权重，多 5 个组合的多空收益进行加权：

图 37: 聚类组合收益对比



资料来源：天软，国信证券经济研究所整理

对比来看，净值的波动率反而有所增加，而收益的增加幅度并不显著。但聚类组合整体的多空收益比全 A 组合更高，且更加稳定。

结论

本篇报告尝试了以波动率和市值两个因子作为聚类分析的特征，以此观察不同的聚类簇中的多因子收益特征。

在欧氏距离的基础上，聚类对于市值和波动率的分割很清晰，结合簇数以及分割的大小关系，我们最终选择将全 A 股分为 5 簇。虽然划分的依据是市值和波动率，但是划分后的 5 个类别的盈利、增长、换手率等因子的数值特征也是不同的。波动率对于营收增长具有显著的相关性，高波动率股票的整体营收增长率更高。而对于 ROE 因子，则可以看到显著的市值特征，大市值股票的整体 ROE 更高。而对于 EP 和换收益，市值与波动率特征同时显著。

以 IC 作为聚类检验的标准，在整体全 A 股的 IC 不显著的因子中，不同聚类簇的 IC 会体现出高于整体的显著水平。对于换手率、动量、波动率等价值因子，大市值的因子显著性是低于小市值组合的，其中以大市值、低波动的组合最为明显。而对于小市值组合中，高波动率组合的因子显著性较好，包括财务因子和价值因子都较为明显。对于财务因子而言，显著性更多的体现在大市值的两个组合之中。

构建因子打分选股策略，在同样的选股数量（均为总数的 30%）下，聚类之后的股票多空组合收益显著比全 A 组合稳定。月度的夏普率接近全 A 组合的 2 倍。

从回测的角度来看，通过聚类将全 A 股划分为不同的类别之后，因子的显著性在各类别内部相较整体提高，并以此带来了整体选股收益的增强。

国信证券投资评级

类别	级别	定义
股票 投资评级	买入	预计 6 个月内，股价表现优于市场指数 20%以上
	增持	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数 $\pm 10\%$ 之间
	卖出	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	超配	预计 6 个月内，行业指数表现优于市场指数 10%以上
	中性	预计 6 个月内，行业指数表现介于市场指数 $\pm 10\%$ 之间
	低配	预计 6 个月内，行业指数表现弱于市场指数 10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

风险提示

本报告版权归国信证券股份有限公司（以下简称“我公司”）所有，仅供我公司客户使用。未经书面许可任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。我公司不保证本报告所含信息及资料处于最新状态；我公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券机构销售团队

华北区 (机构销售一部)	华东区 (机构销售二部)	华南区 (机构销售三部)	海外销售交易部
李文英 010-88005334 13910793700 liwying@guosen.com.cn	汤静文 021-60875164 13636399097 tangjingwen@guosen.com.cn	赵晓曦 0755-82134356 15999667170 zhaoxxi@guosen.com.cn	赵冰童 0755-82134282 13693633573 zhaobt@guosen.com.cn
王玮 13726685252	吴国 15800476582	邵燕芳 0755-82133148 13480668226	梁佳 0755-25472670 13602596740
许婧 18600319171	梁轶聪 021-60873149 18601679992 liangyc@guosen.com.cn	颜小燕 0755-82133147 13590436977 shaoyf@guosen.com.cn	程可欣 886-0975503529(台湾) yanxy@guosen.com.cn
边祎维 13726685252	倪婧 18616741177	黄明燕 18507558226	chengkx@guosen.com.cn
王艺汀 13726685252	林若 13726685252	刘紫微 13828854899	夏雪 18682071096
陈雪庆 18150530525	张南威 13726685252	简洁 13726685252	吴翰文 13726685252
杨云崧 18150530525	周鑫 13726685252	欧子炜 18150530525	
赵海英 010-66025249 13810917275 zhaohy@guosen.com.cn	张欣慰 13726685252		