

核聚类方法与寻找A股行业轮动的核

分析师 焦健

SAC执业证书编号：S0980210040012

2010，成都

内容目录

1 行业轮动方法的综述比较

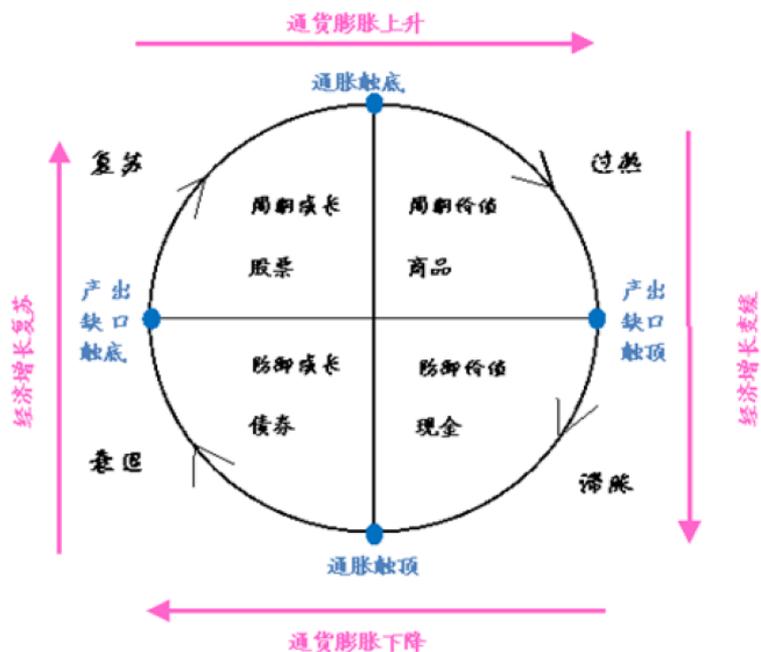
2 核聚类方法的介绍与应用

3 寻找A股行业轮动的两个核

4 轮动效果与模型稳定性检验

5 后续研究：从恒星到行星和流星

从美林投资时钟到国信投资时钟



美林投资时钟强调大类资产之间的轮动，而国信投资时钟强调资本市场中具有各类资产相似属性的行业之间的轮动。

宏观经济数据的滞后性及真实性、股市行业轮动的频度与经济周期轮动速率的不匹配，都可能影响最终的轮动效果。

复苏周期（周期成长指数）—房地产和汽车，国民经济的先导性行业；
过热周期（周期价值指数）—煤炭和有色，同时受益于经济增长和通货膨胀；
滞胀周期（防御价值指数）—医药生物和零售，典型的增长稳定和抗通胀行业；
衰退周期（防御成长指数）—机场和高速公路，现金流稳定的债券型股票。

穷举切分下的市场周期非周期轮动

对沪深300超过50万种切分，周期非周期分类轮动超额收益面积排名第一

排名	一类行业名称	平均超额收益差	切换次数
1	采掘, 金融, 交运, 金属非金属, 房地产	0.034553	78.00
2	采掘, 金融, 木材家具, 交运, 金属非金属, 房地产	0.034477	85.79
3	采掘, 金融, 交运, 电气水, 金属非金属, 房地产	0.034428	88.40
4	采掘, 金融, 交运, 社会服务, 金属非金属, 房地产	0.034313	90.49
5	采掘, 金融, 木材家具, 交运, 社会服务, 金属非金属, 房地产	0.034241	92.58
6	采掘, 金融, 交运, 传播文化产, 金属非金属, 房地产	0.034239	93.55
7	采掘, 金融, 木材家具, 交运, 传播文化, 金属非金属, 房地产	0.034161	94.58
8	采掘, 金融, 交运, 金属非金属, 造纸印刷, 房地产	0.034094	95.39
9	采掘, 金融, 交运, 金属非金属, 制造业, 房地产	0.034036	97.09
10	采掘, 金融, 木材家具, 交运, 金属非金属, 造纸印刷, 房地产	0.034017	96.52

伪轮动还是过度拟合？多元线性方法

使用**LOGIT**模型，将是否切换的**0**与**1**事件转换为多元线性方程：

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

通过历史数据对多个因子进行线性拟合，得到轮动的预测方程式。

期望用简单多元线性模型对轮动进行左侧的预判，无法摆脱“看图说话” - “海选因子” - “刷参数”的三部曲，美好的效果基本都停留在样本内。

在多因子轮动预测中，建议使用**CART**决策树方法而非线性/非线性拟合。

相对强弱RS模型结合技术分析指标

两资产相对强弱（**relative strength**）方法上世纪90年代提出后，受到广泛应用，在国内资本市场的实证中，也出现了多种变形：

$$RS1 = \ln(A \text{ 价格}) - \ln(B \text{ 价格})$$

$$RS2 = \ln(A \text{ 累积收益率}) - \ln(B \text{ 累积收益率})$$

$$RS3 \text{ (GSRS)} = \ln(1+A \text{ 累积收益率}) - \ln(1+B \text{ 累积收益率})$$

对生成的RS序列，通过技术分析指标，寻找趋势和拐点，进行轮动。

常用技术指标：均线、MACD、布林线、KDJ

尽管有多种“变形”，我们必须强调相对强弱方法的某些基本特性。

内容目录

1 行业轮动方法的综述比较

2 核聚类方法的介绍与应用

3 寻找A股行业轮动的两个核

4 轮动效果与模型稳定性检验

5 后续研究：从恒星到行星和流星

传统的聚类分析方法

变量之间的聚类即R型聚类，常用相似系数衡量变量间的亲疏关系。
样本之间的聚类即Q型聚类，常用空间距离衡量样本间的亲疏关系。

对距离的定义，必须满足以下条件：

$d_{ij} \geq 0$ 对一切的i和j成立；

$d_{ij} = 0$ 当且仅当 $i = j$ 成立；

$d_{ij} = d_{ji}$ 对一切的i和j都成立；

$d_{ij} \leq d_{ik} + d_{kj}$ 对一切i和j成立；

常用的距离衡量方法：

欧式距离、杰式距离、兰式距离、马氏距离、斜交空间距离

核聚类的关键在从确定到不确定

在聚类方法中，没有所谓核样本的概念。Chaudhurid在1997年《*A novel multiseed nonhierarchical data clustering technique*》中提到多中心最小生成树的概念，进行聚类。与中心聚类不同，我们这里提出的是一种以空间中某些已知具有强特性或强区分度的样本作为核，根据其他点与核样本的距离亲疏程度，化繁为简，进行快速分类的方法。

在对样本空间 X 进行研究时，

若 X 内样本性质未知，而 Y 空间样本性质已知，则可采用灰色预测的方法；

若 X 内部分样本性质已知，则无需引入其他的样本，可直接用核聚类方法。

核聚类应用：不确定概念分类

问题提出：如何用量化的方法，构建新兴产业投资股票池？

传统解决方案：

用统计方法对主营业务收入、财务报表、公司公告等进行筛选

核聚类解决方案：

我们已知某些行业中的某些股票 X_i ，一定是新兴产业的代表。

我们已知某些行业中的某些股票 Y_i ，一定不是新兴产业范畴。

分别以 X_i 和 Y_i 作为样本空间中的核，计算空间其他点的亲疏距离。

快速排除大部分样本后，保留的样本，则可能具有潜在的目标属性。

其他应用：

行业指数上涨的噪音分析，行业内部的细分与归因。

内容目录

1 行业轮动方法的综述比较

2 核聚类方法的介绍与应用

3 寻找A股行业轮动的两个核

4 轮动效果与模型稳定性检验

5 后续研究：从恒星到行星和流星

行业轮动中核的判定因素



基于前述核聚类的思想，我们希望找到两个轮动效果最为明显的行业，作为观察A股行业轮动的核或称恒星，进而挖掘出围绕着恒星运动的行星，最后处理与核行业关系不稳定的流星，或本身质量极高却起着破坏星系轮动效果的“黑洞”。

对于核行业选取，我们从质量（行业流通权重）、距离（欧式空间距离）、区分度（与其他行业距离的分歧）和经济含义（具有鲜明的区别）出发。

从欧氏距离矩阵中寻找核

	金融保险	机械设备	采掘业	钢铁有色	交运仓储	医药生物	房地产业	食品饮料	石油化学	信息技术	批零贸易	电煤气水	电子	建筑业
金融保险		0.88	0.81	0.83	0.79	0.98	0.83	0.90	0.92	0.92	0.91	0.85	0.98	0.88
机械设备	0.88		0.68	0.49	0.46	0.41	0.68	0.50	0.34	0.49	0.36	0.48	0.46	0.46
采掘业	0.81	0.68		0.59	0.62	0.81	0.84	0.73	0.68	0.72	0.73	0.67	0.83	0.74
钢铁有色	0.83	0.49	0.59		0.48	0.67	0.73	0.65	0.51	0.66	0.59	0.53	0.67	0.60
交运仓储	0.79	0.46	0.62	0.48		0.58	0.70	0.55	0.46	0.58	0.51	0.46	0.62	0.55
医药生物	0.98	0.41	0.81	0.67	0.58		0.78	0.50	0.42	0.56	0.39	0.56	0.54	0.55
房地产业	0.83	0.68	0.84	0.73	0.70	0.78		0.79	0.69	0.78	0.69	0.76	0.77	0.65
食品饮料	0.90	0.50	0.73	0.65	0.55	0.50	0.79		0.52	0.60	0.45	0.60	0.65	0.63
石油化学	0.92	0.34	0.68	0.51	0.46	0.42	0.69	0.52		0.51	0.36	0.50	0.45	0.48
信息技术	0.92	0.49	0.72	0.66	0.58	0.56	0.78	0.60	0.51		0.51	0.59	0.47	0.58
批零贸易	0.91	0.36	0.73	0.59	0.51	0.39	0.69	0.45	0.36	0.51		0.55	0.49	0.49
电煤气水	0.85	0.48	0.67	0.53	0.46	0.56	0.76	0.60	0.50	0.59	0.55		0.63	0.52
电子	0.98	0.46	0.83	0.67	0.62	0.54	0.77	0.65	0.45	0.47	0.49	0.63		0.56
建筑业	0.88	0.46	0.74	0.60	0.55	0.55	0.65	0.63	0.48	0.58	0.49	0.52	0.56	

金融和地产，尤其是金融是行业空间中的另类，距离所有行业都很远。
医药生物的分度最高，可作为一个固定的核，与其距离最远的是采掘。

核行业累计收益率差的动量反转特性

2000-2010不同采样期/持有期收益相关性统计

	1周	2周	3周	4周	5周	6周	7周	8周	9周	10周	11周	12周	4月	5月	6月
1个月	3.49%	6.40%	6.17%	6.34%	8.15%	10.15%	12.47%	14.05%	13.70%	13.23%	11.99%	10.59%	5.34%	-4.52%	-9.47%
2个月	7.04%	11.06%	12.99%	14.07%	14.73%	14.81%	15.07%	14.71%	13.62%	13.27%	11.12%	8.15%	-2.67%	-13.13%	-18.06%
3个月	6.84%	9.79%	10.53%	10.54%	10.53%	10.73%	9.84%	7.84%	5.47%	3.63%	0.26%	-2.35%	-12.77%	-21.65%	-24.73%
4个月	5.54%	8.49%	7.60%	5.74%	4.10%	2.79%	0.49%	-1.42%	-3.98%	-6.10%	-8.65%	-11.29%	-20.34%	-27.13%	-28.13%
5个月	1.72%	2.29%	0.27%	-1.03%	-2.89%	-4.64%	-6.48%	-8.68%	-11.26%	-12.98%	-15.46%	-17.39%	-24.74%	-29.38%	-27.29%
6个月	0.06%	-0.36%	-2.05%	-4.16%	-6.59%	-8.51%	-10.86%	-12.60%	-14.82%	-16.42%	-18.22%	-20.03%	-25.94%	-27.37%	-21.95%

2005-2010不同采样期/持有期收益相关性统计

	1周	2周	3周	4周	5周	6周	7周	8周	9周	10周	11周	12周	4月	5月	6月
1个月	4.60%	4.76%	9.20%	8.07%	8.34%	11.51%	12.04%	14.06%	13.45%	12.43%	12.42%	11.58%	1.79%	-8.43%	-16.15%
2个月	6.10%	9.70%	13.17%	14.37%	13.63%	14.43%	14.70%	15.24%	12.78%	11.86%	9.39%	5.81%	-8.36%	-20.88%	-28.86%
3个月	5.52%	7.83%	11.28%	11.96%	9.78%	10.28%	8.11%	5.72%	2.69%	0.21%	-3.34%	-6.45%	-20.70%	-31.78%	-36.39%
4个月	2.55%	5.33%	5.24%	3.13%	0.50%	-0.61%	-3.62%	-5.85%	-9.72%	-12.23%	-15.07%	-18.15%	-30.51%	-38.46%	-40.90%
5个月	-0.07%	-0.31%	-1.58%	-3.38%	-7.39%	-9.13%	-11.72%	-14.32%	-18.05%	-20.16%	-23.04%	-25.55%	-35.12%	-40.89%	-39.57%
6个月	-4.44%	-5.09%	-6.28%	-8.78%	-13.04%	-14.92%	-18.01%	-20.35%	-23.25%	-25.27%	-27.02%	-28.91%	-36.83%	-38.75%	-32.66%

长期看，3个月以内采样均呈现动量效果，而5个月以上基本体现为反转。
05年以来该特征尤其是反转特征更加明显，且动量有效的采样期在缩短。

内容目录

1 行业轮动方法的综述比较

2 核聚类方法的介绍与应用

3 寻找A股行业轮动的两个核

4 轮动效果与模型稳定性检验

5 后续研究：从恒星到行星和流星

医药、采掘行业的轮动规则

我们无法证明最完美的轮动方法，因此把精力放在了寻找最远的核。对于收益差异极大的资产，我们倾向选用最简单的“追逐动量法”。

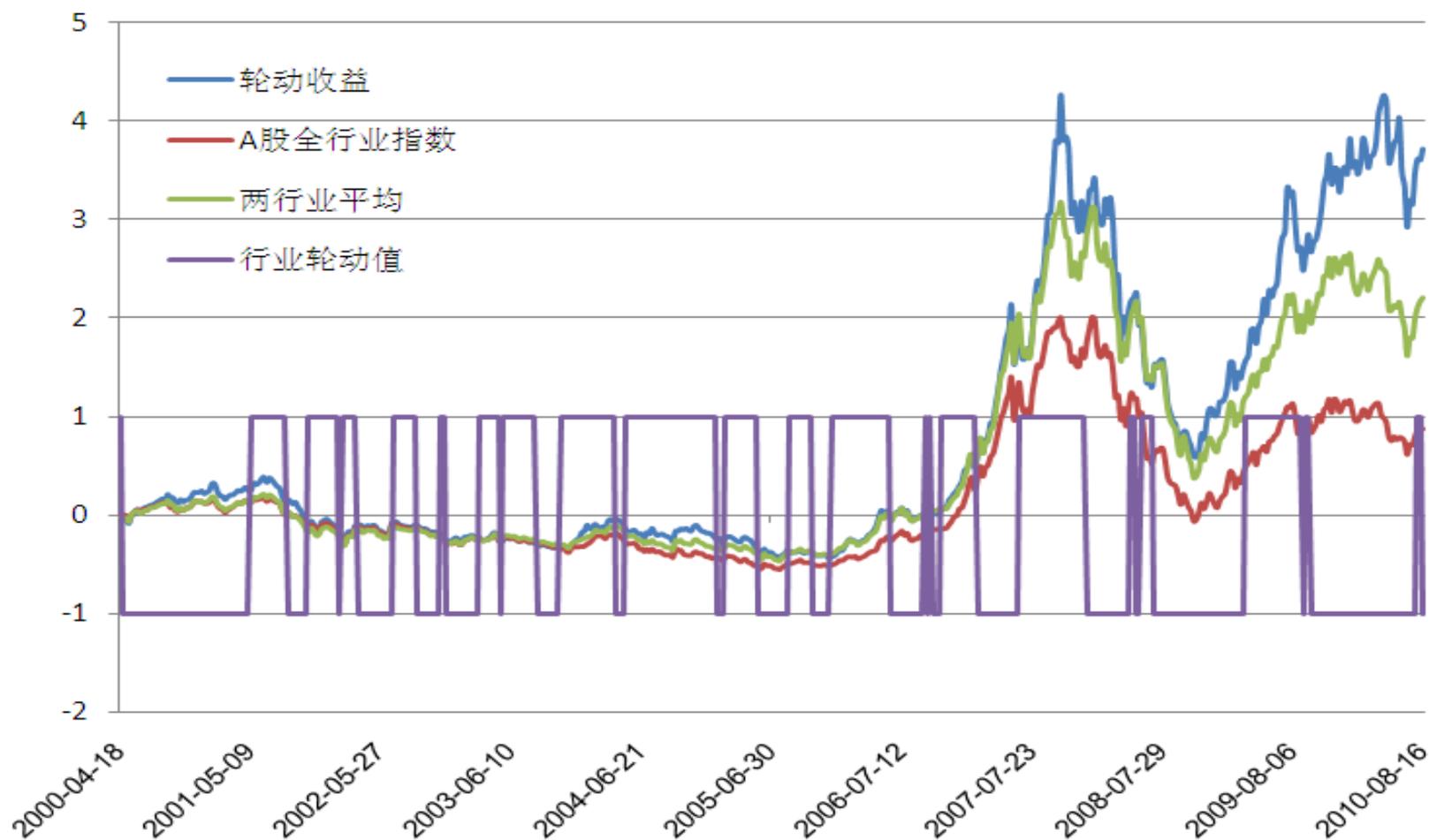
追逐动量法：

对于收益率序列差异极大的资产A和B，其相对收益应体现为动量效应。我们无法预测其动量效应的持续周期，因此会随时根据市况进行切换。以周为单位，观察过去若干周A与B的收益情况，并选择买入强的一方。对收益的度量有：区间累积收益、区间平均收益、指数移动加权收益。

不可避免的，这是一个对采样期参数依赖的策略。

我们希望通过医药和采掘自身极强的差异，来弥补对外围参数的依赖。这也是我们首先寻找两个核进行轮动的原因。

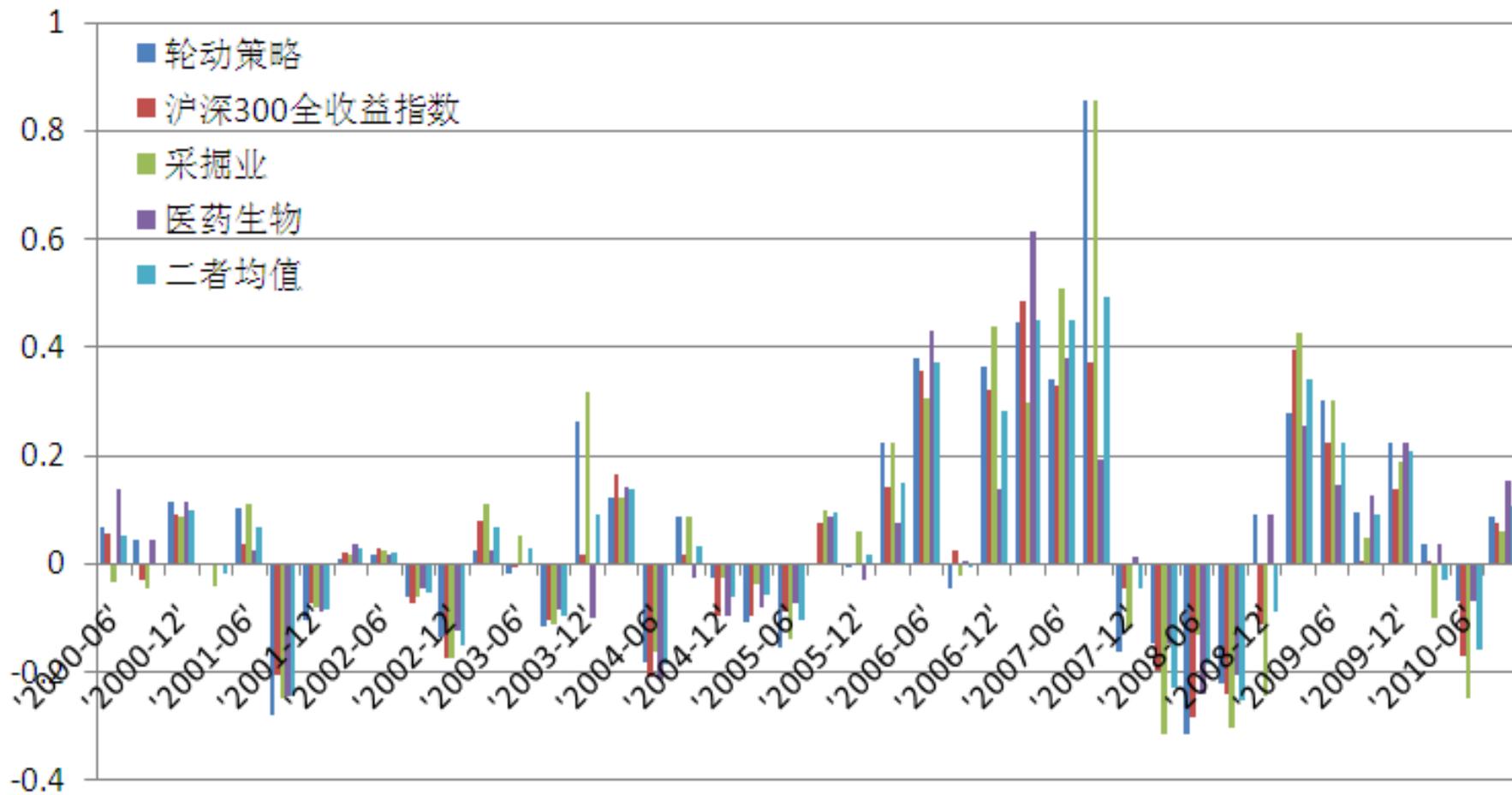
三个月动量驱动下的医药采掘轮动



轮动组合的绩效观测

	均值	方差	累计超额收益	年化累计超额收益	指数超额收益均值	指数超额收益方差	信息率(年化)	Rf超额收益均值	Rf超额收益方差	夏普比(年化)
医药vs采掘轮动	0.42%	4.71%	284.78%	10.77%	0.20%	2.26%	64.86%	0.37%	4.71%	57.19%
食品饮料	0.34%	4.26%	155.83%	6.94%	0.12%	2.05%	41.66%	0.29%	4.26%	48.78%
医药生物	0.32%	4.54%	113.29%	5.38%	0.11%	2.00%	37.97%	0.28%	4.54%	43.67%
采掘业	0.33%	4.79%	106.05%	5.10%	0.11%	2.35%	33.84%	0.28%	4.79%	42.18%
机械设备	0.31%	4.78%	83.71%	4.18%	0.10%	1.35%	50.97%	0.27%	4.78%	40.02%
批零贸易	0.31%	4.72%	80.42%	4.04%	0.09%	1.56%	41.52%	0.26%	4.72%	39.70%
金融、保险业	0.28%	4.75%	39.28%	2.14%	0.06%	2.95%	14.13%	0.23%	4.75%	34.56%
金属非金属	0.29%	5.04%	35.19%	1.93%	0.07%	1.76%	28.33%	0.24%	5.04%	34.23%
石油、化学	0.26%	4.80%	19.76%	1.12%	0.04%	1.53%	20.22%	0.21%	4.80%	32.00%
社会服务业	0.25%	4.83%	3.26%	0.19%	0.03%	1.80%	11.19%	0.20%	4.83%	29.52%
A股全行业指数	0.22%	4.32%	0.00%	0.00%	0.00%	0.00%	N/A	0.17%	4.32%	28.38%
交运、仓储	0.22%	4.42%	-5.03%	-0.30%	0.00%	1.51%	-0.32%	0.17%	4.42%	27.62%
房地产业	0.23%	5.14%	-20.72%	-1.29%	0.02%	2.43%	4.54%	0.19%	5.14%	25.99%
农林牧渔	0.24%	5.37%	-28.25%	-1.80%	0.02%	2.51%	5.90%	0.19%	5.37%	25.58%
纺织服装	0.23%	5.17%	-26.86%	-1.71%	0.01%	1.95%	3.89%	0.18%	5.17%	25.15%
传播与文化	0.22%	5.39%	-39.21%	-2.58%	0.01%	2.92%	1.49%	0.18%	5.39%	23.53%
电力煤气水	0.18%	4.24%	-33.81%	-2.19%	-0.04%	1.79%	-17.04%	0.13%	4.24%	21.67%
其他制造业	0.21%	5.39%	-50.88%	-3.47%	-0.01%	2.28%	-2.93%	0.16%	5.39%	21.49%
建筑业	0.19%	4.81%	-43.46%	-2.89%	-0.03%	1.75%	-12.55%	0.14%	4.81%	20.92%
造纸、印刷	0.19%	5.14%	-52.41%	-3.59%	-0.02%	1.96%	-9.15%	0.15%	5.14%	20.35%
综合类	0.19%	5.15%	-56.46%	-3.91%	-0.03%	1.73%	-12.52%	0.14%	5.15%	19.56%
信息技术	0.15%	4.63%	-64.37%	-4.58%	-0.07%	1.86%	-26.97%	0.10%	4.63%	15.61%
木材、家具	0.15%	5.70%	-91.47%	-7.20%	-0.06%	3.05%	-15.21%	0.11%	5.70%	13.35%
电子	0.11%	5.26%	-101.36%	-8.33%	-0.11%	2.13%	-37.18%	0.06%	5.26%	8.21%

轮动组合的绩效分解

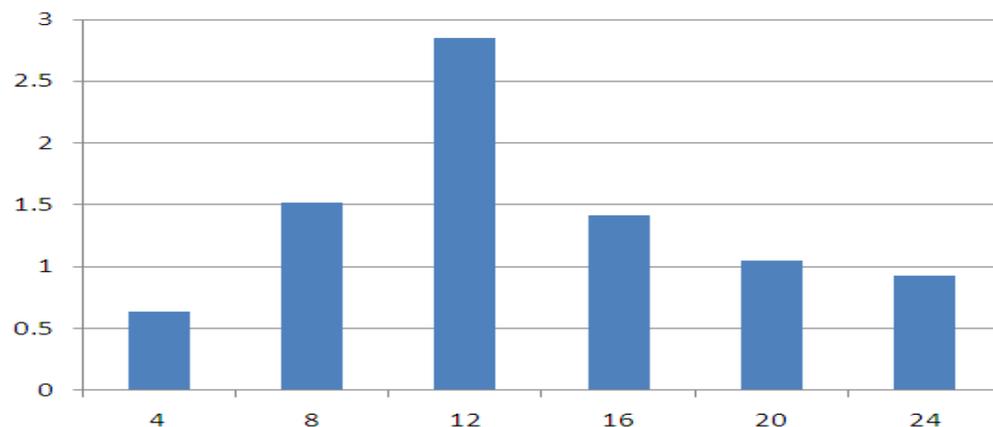


轮动组合的绩效分解

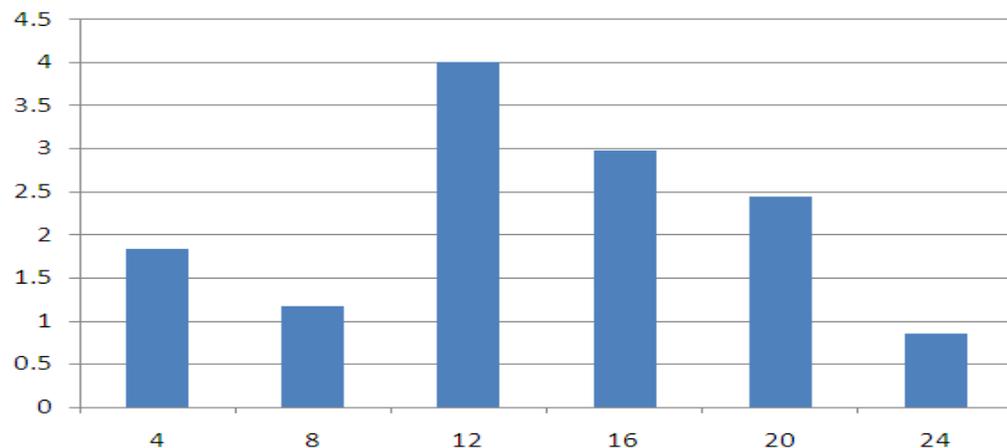


轮动组合超额收益对采样期的依赖

2000-2010不同采样期轮动超额收益的比较



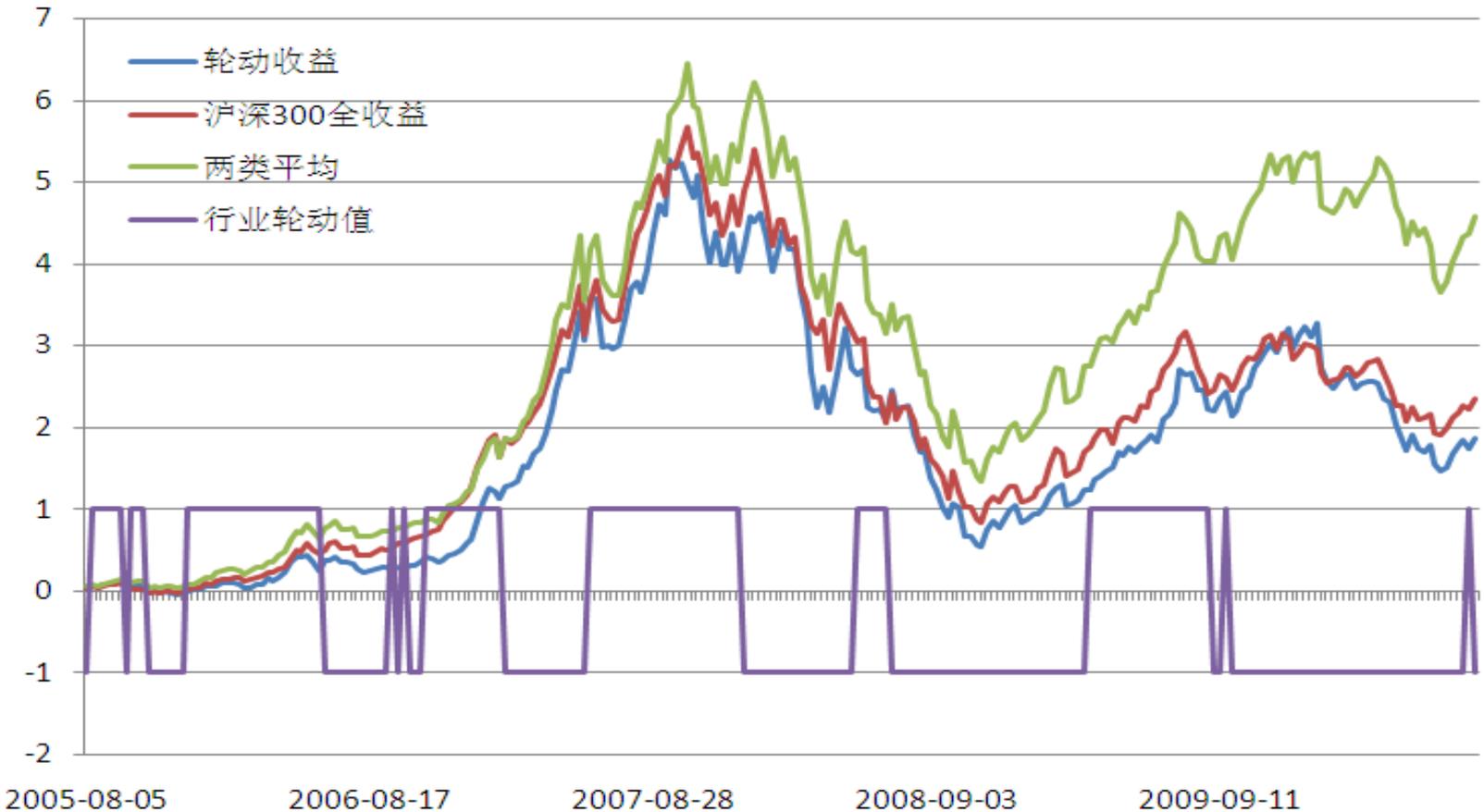
2005-2010不同采样期轮动超额收益的比较



从参数检验结果来看，三个月（12周）作为采样期进行动量追逐轮动的策略效果较好，且该结果适用于历史上几乎任意阶段。

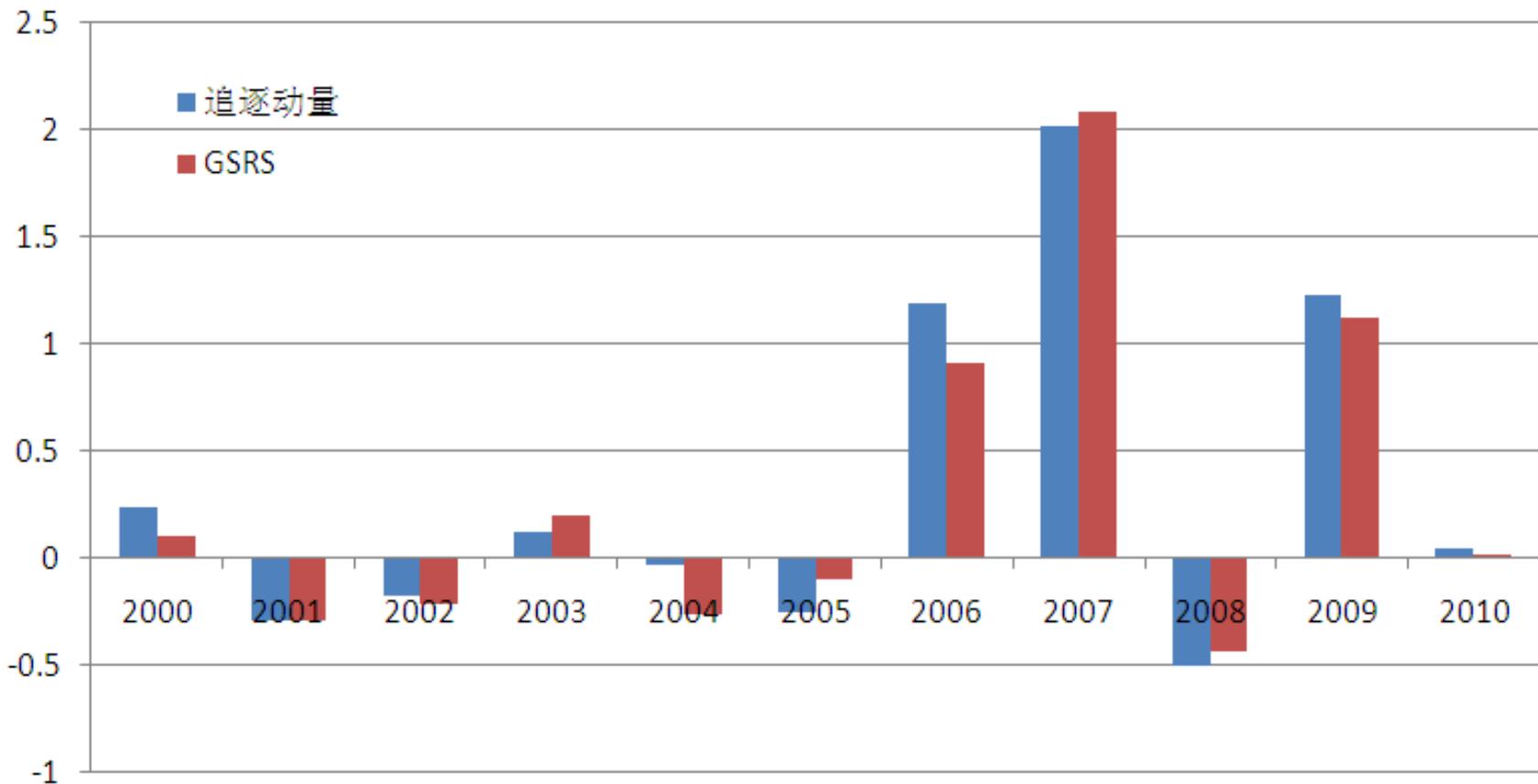
采样期指数移动加权未能提升轮动效果

对采样期各周收益率序列进行指数移动加权处理，未能提升轮动效果。



与GSRs+经典参数MACD轮动的比较

在大多数年份中，医药VS采掘的追逐动量轮动跑赢了经典参数GSRs策略。



内容目录

1 行业轮动方法的综述比较

2 核聚类方法的介绍与应用

3 寻找A股行业轮动的两个核

4 轮动效果与模型稳定性检验

5 后续研究：从恒星到行星和流星

从区分度的角度，衡量恒星周围的行星

确定了两颗恒星以后，我们试图将剩下的行业逐个添加进这个体系。

定义一个行业k对于两个核行业i、j的区分度为它到两个核所在星系内所有行业的平均距离的差异性：

$$\text{区分度} = \max(\text{mean_dist}_{ik} / \text{mean_dist}_{jk}, \text{mean_dist}_{jk} / \text{mean_dist}_{ik})$$

其中 $mean_dist_{ik}$ 指得是到第i个核行业及其行星行业的距离的均值。

我们选出剩余未分类行业中区分度最大者，加入到某一个核的周围。

再对剩余未分类行业重新计算区分度，以此类推。

从区分度的角度，衡量恒星周围的行星

	新加入	新加入	一类权重	二类权重	超额收益率	胜率	两边均值距离
初始恒星	采掘业	医药生物	8.86%	4.54%	284.78%	81.82%	0.8099
第1步		批零贸易	8.86%	8.34%	182.18%	63.64%	0.7461
第2步		机械设备	8.86%	19.36%	144.20%	63.64%	0.7086
第3步		石油化学	8.86%	23.73%	144.67%	54.55%	0.6893
第4步		综合类	8.86%	25.83%	114.94%	54.55%	0.6891
第5步		纺织服装	8.86%	27.21%	107.23%	54.55%	0.6935
第6步		电子	8.86%	29.41%	74.27%	54.55%	0.7007
第7步		其他制造业	8.86%	30.01%	67.13%	54.55%	0.7045
第8步		造纸印刷	8.86%	30.64%	72.42%	54.55%	0.7015
第9步		社会服务业	8.86%	32.00%	70.95%	54.55%	0.6987
第10步		农林牧渔业	8.86%	33.16%	89.04%	54.55%	0.7014
第11步		传播与文化	8.86%	33.60%	101.27%	54.55%	0.7062
第12步		建筑业	8.86%	35.32%	99.07%	54.55%	0.6994
第13步		木材、家具	8.86%	35.45%	99.32%	54.55%	0.7047
第14步		信息技术业	8.86%	39.61%	63.41%	63.64%	0.6973
第15步		食品、饮料	8.86%	44.05%	79.74%	63.64%	0.6899
第16步	金融保险		36.82%	44.05%	80.48%	54.55%	0.685
第17步		电力煤气水	36.82%	47.25%	15.92%	45.45%	0.6734
第18步		交运仓储业	36.82%	52.01%	30.14%	54.55%	0.6613
第19步		金属非金属	36.82%	58.71%	-14.32%	45.45%	0.6485
第20步		房地产业	36.82%	63.18%	-12.21%	45.45%	0.6377

在轮动体系中增加新的行业，不会增加轮动收益，反而均造成收益的快速下降。

小行业间同质化严重，而占沪深300指数权重最高的几个行业则呈现出显著的特异性。

在下一步的研究中，我们有必要对这些特异性的行业进行单独分析。



国信证券经济研究所

GUOSEN Guosen Securities Economic Research Institute

全球视野 本土智慧
GLOBAL VIEW LOCAL WISDOM

分析师

焦健

E-mail: jiaojian1@guosen.com.cn

SAC执业证书编号: S0980210040012