

花开股市，相似几何—— 弱关联关系下的特异性Alpha因子挖掘

首席分析师郑兆磊

兴业证券金融工程团队

2022年6月



- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进
- 四、降维方法改进
- 五、相似底层向量改进
- 六、相似方法论本质探讨和总结

1、引言—兴证金工新因子与相似度体系



新因子困局

- 新因子是一件可遇不可求的事情。
- 投资者更多的是在因子的选择和合成方式方面开始进行探索，这方面的收获亦颇为丰富。



兴证金工新因子研究

- 新因子合成三种方式：老数据新加工、另类数据、合成方式探索（机器学习）；
- **基于逻辑演绎的新因子始终是最受欢迎最易接受的**，兴证金工孜孜不倦的展开新因子的探索；



因子合成与相似度体系

- 专利动量因子以及财务动量因子均是从信号合成角度出发，利用余弦相似度衡量任意两个公司的信息相似程度。
- 兴证金工团队也在深入思考该方法论的逻辑，**希望能进一步利用相似程度体系。**

- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进
- 四、降维方法改进
- 五、相似底层向量改进
- 六、相似方法论本质探讨和总结

2.1、相似动量方法论及改进维度一方法论回顾

财务关联度

我们定义**财务关联度**为两个股票之间在十个财务因子构建的高维空间下的余弦相似度：

$$F-Link_{ijt} = \frac{\sum_{k=1}^{10} C_{itk} * C_{jtk}}{\sqrt{\sum_{k=1}^{10} C_{itk}^2} * \sqrt{\sum_{k=1}^{10} C_{jtk}^2}}$$

- k 代表十个财务因子、 i, j 代表两个不同的公司、 t 代表相应时点；
- C_{itk} 代表公司 i 在 t 时刻时 k 因子的值大小；



财务动量因子

我们定义**财务动量因子**为表示针对于公司 i ，所有关联公司月度收益率（也可以季度或者其他频次的收益率）的财务关联度加权结果：

$$F-Momen_{it} = \frac{\sum_{i \neq j} F-Link_{ijt} * (Ret_{jt} - R)}{\sum_{i \neq j} F-Link_{ijt}}$$

- 构建逻辑为：因子值越大表示和公司 i 财务关联度高的集群公司在考察月份（默认月度，也可以是其他时间频次）涨幅相对较高。考察公

因子流程图

四维度作为一个向量

CurrentRatio_LR
Revenue_SQ_YoY
ROE_TTM_YoY
ReceivablesTurnover

假设市场就3家公司，通过余弦相似得到财务相似矩阵



	a	b	c
a		0.1	0.2
b			0.5
c			



以10月份a公司，其财务动量因子即为：

$(0.1 * b \text{公司10月份收益率} + 0.2 * c \text{公司10月份收益率}) / (0.1 + 0.2)$

因子值越大，表示和公司i财务关联度高的集群公司在考察月份（默认月度，也可以是其他时间频次）涨幅相对较高。考察公

2.1、相似动量方法论及改进维度一方法论回顾

➤ 财务动量因子具有**较强的选股能力**：行业市值中性化后因子IC达到0.026、T值达到4.9，分位数各组别严格单调，具体参见下表与下图。

表、F-Momentum 以及行业市值中性化后因子 IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
F-Momentum	0.027	0.12	-0.24	0.32	0.23	2.7
中性化 F-Momentum	0.026	0.06	-0.10	0.21	0.41	4.9

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

备注：回测区间：2009 年 12 月 31 日-2021 年 12 月 31 日

图、行业市值中性化后 F-Momentum 因子分位数测试

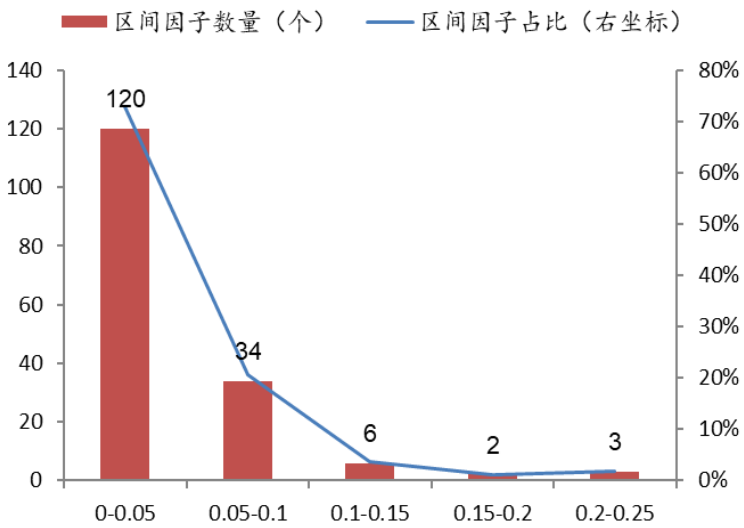
	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率	年化超额收益率	跟踪误差
Top	11.0%	29.6%	0.37	153.4%	51.5%	0.8%	15.7%
1	12.1%	28.9%	0.42	171.7%	52.0%	1.7%	46.6%
2	10.2%	28.4%	0.36	175.0%	49.8%	-0.2%	-6.5%
3	10.4%	28.2%	0.37	175.2%	50.2%	-0.1%	-3.0%
4	10.8%	28.2%	0.38	174.0%	47.3%	0.3%	10.6%
5	9.0%	27.4%	0.33	174.0%	49.5%	-1.5%	-43.7%
6	6.8%	27.7%	0.25	174.7%	52.6%	-3.5%	-101.9%
7	6.2%	28.0%	0.22	174.6%	53.9%	-4.0%	-128.5%
8	4.9%	28.5%	0.17	170.4%	58.1%	-5.0%	-117.0%
Bottom	-0.4%	29.6%	-0.02	151.0%	62.0%	-9.7%	-188.2%
市场	10.4%	28.6%	0.36		44.9%		
L-S	11.1%	7.7%	1.45		12.3%		

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

备注：回测区间：2009 年 12 月 31 日-2021 年 12 月 31 日

➤ 同时该因子具有**极强的特异性**。我们通过相关性分析、Fama-Mac回归测试、合成方式探索、对兴证金工指数增强的改进作用四个维度对该因子特异性进行了全方位的分析。结果显示该因子的特异性极强。

图、F-Momentum 与各选股因子相关性分布图



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

备注：回测区间：2009 年 12 月 31 日-2021 年 12 月 31 日

2.2、相似动量方法论及改进维度一改进维度介绍

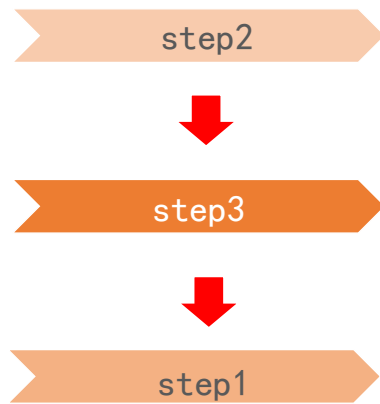
➤ 相似三部曲：

→ 底层向量（十个财务因子）；

→ 矩阵利用；

→ 降为方式

三步均可以尝试，尝试顺序(按照复杂度进行介绍)：



Step1:

得到任意一个公司十个财务，并进一步作为一个向量

Step3:

对矩阵进行归一化处理，对目标公司：计算除目标公司外的其他公司在财务关联度下的加权平均月收益率，进而得到财务动量因子

Step2:

计算任意两个公司财务分布相似的大小，得到一个 $N*N$ 矩阵（财务关联度矩阵）

- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进(step2)
- 四、降维方法改进(step3)
- 五、相似底层向量改进(step1)
- 六、相似方法论本质探讨和总结

3.1、矩阵应用改进一删除尾部噪音

- 我们知道风格因子的表现不仅仅取决于因子自身，还取决于股票池。除行业和特定板块外，主流的宽基指数包括沪深300、中证500等。理想情况下基于相应股票池进行研究得到的规律是更为可靠的（毕竟很少有策略会基于全市场选股）。
- 其中基于沪深300、中证500股票池计算得到因子我们称之为F-Momentum-300、F-Momentum-500
- 以沪深300为例，从一开始因子标准化到寻找财务相似度均是在沪深300股票池内进行；另外因子均是剥离了动量、行业市值风格。

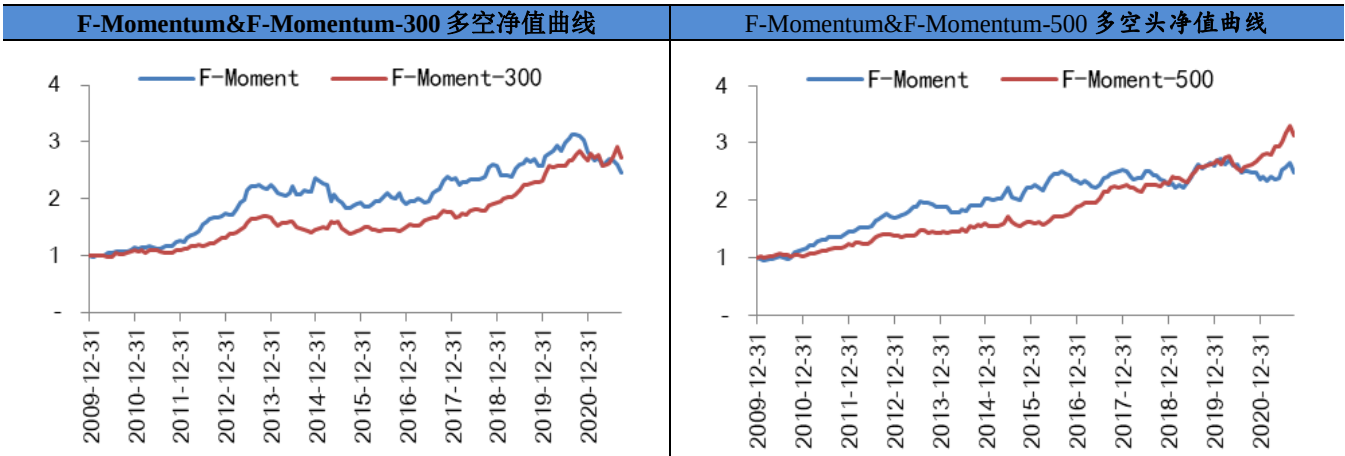
基于不同股票池的因子表现

股票池	因子	因子IC表现				因子分位数测试多空表现		
		均值	波动率	ICIR	T值	年化收益	夏普率	最大回撤
沪深300	F-Momentum-300	0.022	0.07	0.29	3.5	8.9%	0.95	18.4%
	F-Momentum	0.022	0.09	0.23	2.7	7.9%	0.67	22.8%
中证500	F-Momentum-500	0.032	0.07	0.47	5.6	10.2%	1.29	10.2%
	F-Momentum	0.027	0.08	0.32	3.8	8.0%	0.83	13.3%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

- 基于不同宽基指数范围构建的因子均有不错的选股能力。以中证500股票池为例，F-Momentum-500因子IC均值为0.032，T值为5.6，多空年化收益达到10.2%，夏普率为1.29，有效性较强；
- 基于全市场得到的因子在不同宽基指数范围内的选股有效性也非常明显。F-Momentum在中证500股票池内IC均值达到0.027、T值3.8、多空年化收益率8.0%，夏普率0.83；

图表、F-Momentum以及相应宽基训练的因子多空净值表现



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

3.1、矩阵应用改进一删除尾部噪音



改进方式1：降噪

可否通过一些手段删除与公司财务分布极其不相似的股票

1

删除尾部财务分布极不相似股票，可以得到新的财务动量。

3

得到的因子进一步进行动量、行业、市值中性化，之后得到的因子分别称之为F-Momentum-20Pec、F-Momentum-30Pec、F-Momentum-50Pec、F-Momentum-80Pec、F-Momentum-90Pec。

2

这里可以设置极不相似的水平，我们共计设置了20%、30%、50%、80%、90%的分档。

3.1、矩阵应用改进一删除尾部噪音

- 对于每只股票我们通过删除尾部财务分布极不相似股票，可以得到新的财务动量。这里可以设置极不相似的水平，我们**共计设置了20%、30%、50%、80%、90%的分档**，得到的因子进一步进行动量、行业、市值中性化。
- 从结果看，这种方式稳健的提升了因子表现：**随着删除水平的提高，因子IC、分位数等关键指标均得到提升**。以F-Momentum-80Pec为例：因子的IC均值达到0.029，T值达到5.78；因子分位数各组别严格单调，多空收益率达到11.6%，夏普比达到1.57，最大回撤7.2%，因子整体表现显著优于之前的F-Momentum因子。

表、删除不同噪音水平下的 F-Momentum IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
F-Momentum	0.025	0.06	-0.10	0.21	0.39	4.62
F-Momentum-20Pec	0.025	0.06	-0.11	0.19	0.41	4.89
F-Momentum-30Pec	0.026	0.06	-0.10	0.20	0.40	4.86
F-Momentum-50Pec	0.027	0.06	-0.11	0.18	0.43	5.13
F-Momentum-80Pec	0.029	0.06	-0.11	0.18	0.48	5.78
F-Momentum-90Pec	0.031	0.06	-0.11	0.17	0.53	6.39

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、删除不同噪音水平下的 F-Momentum 分位数测试对比

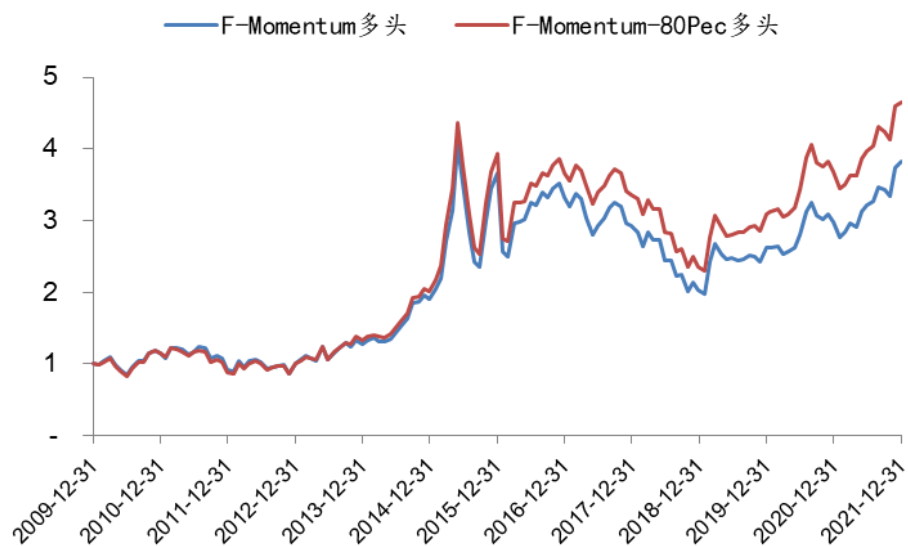
F-Momentum				F-Momentum-20Pec				F-Momentum-30Pec			
分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率
Top	11.8%	0.40	51.4%	Top	12.2%	0.42	49.2%	Top	12.6%	0.43	48.1%
1	13.1%	0.46	51.3%	1	12.4%	0.44	47.9%	1	11.9%	0.42	49.5%
2	10.8%	0.38	50.2%	2	12.0%	0.42	50.3%	2	11.4%	0.40	51.1%
3	11.5%	0.41	48.9%	3	11.4%	0.41	49.0%	3	12.0%	0.43	47.3%
4	11.4%	0.41	47.7%	4	10.6%	0.39	49.7%	4	10.7%	0.39	51.4%
5	9.9%	0.36	49.3%	5	9.3%	0.34	52.2%	5	9.4%	0.34	52.3%
6	8.0%	0.29	51.8%	6	8.6%	0.31	50.7%	6	8.0%	0.29	52.0%
7	6.9%	0.25	54.7%	7	6.9%	0.25	56.8%	7	8.3%	0.30	54.6%
8	6.0%	0.21	58.0%	8	5.3%	0.19	57.6%	8	4.5%	0.16	58.8%
Bottom	0.9%	0.03	61.0%	Bottom	1.6%	0.06	60.1%	Bottom	1.7%	0.06	58.8%
L-S	10.5%	1.39	11.5%	L-S	11.1%	1.40	11.2%	L-S	10.3%	1.49	11.0%
F-Momentum-50Pec				F-Momentum-80Pec				F-Momentum-90Pec			
Top	13.3%	0.46	48.6%	Top	13.7%	0.47	47.3%	Top	13.1%	0.45	48.5%
1	11.9%	0.42	50.6%	1	12.2%	0.43	52.8%	1	14.6%	0.51	48.2%
2	12.6%	0.44	47.7%	2	12.8%	0.45	44.8%	2	11.8%	0.42	50.7%
3	12.0%	0.43	47.4%	3	11.5%	0.41	47.0%	3	11.4%	0.41	47.5%
4	10.0%	0.36	50.7%	4	11.1%	0.41	48.4%	4	10.3%	0.38	48.5%
5	8.9%	0.33	52.3%	5	8.7%	0.32	49.1%	5	10.0%	0.37	45.7%
6	8.3%	0.30	52.0%	6	7.6%	0.27	55.3%	6	7.0%	0.25	55.0%
7	7.1%	0.25	55.8%	7	6.9%	0.25	58.0%	7	6.5%	0.23	56.6%
8	5.2%	0.18	59.6%	8	4.6%	0.16	58.3%	8	3.7%	0.13	62.0%
Bottom	1.2%	0.04	59.4%	Bottom	1.5%	0.05	61.1%	Bottom	2.0%	0.07	60.1%
L-S	11.5%	1.60	7.6%	L-S	11.6%	1.57	7.2%	L-S	10.6%	1.48	6.8%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

3.1、矩阵应用改进—删除尾部噪音

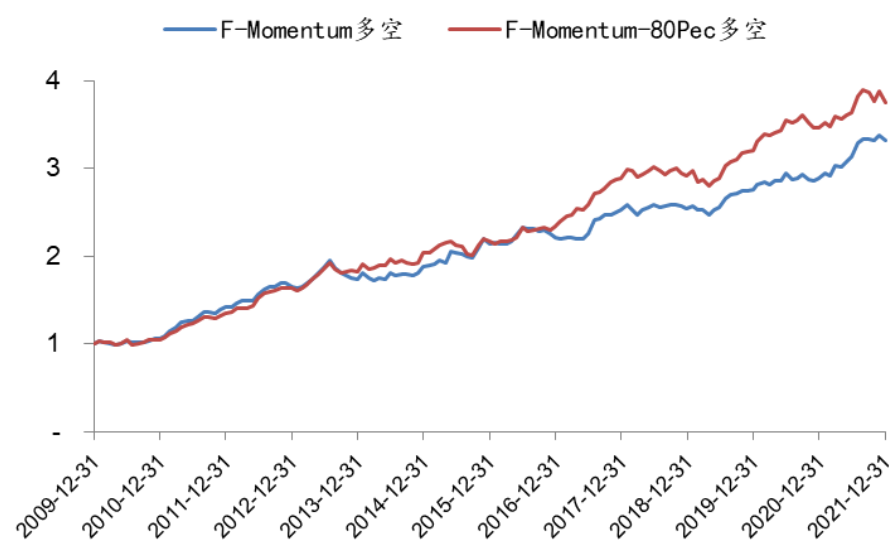
- 对于每只股票我们通过删除尾部财务分布极不相似股票，可以得到新的财务动量。这里可以设置极不相似的水平，我们**共计设置了20%、30%、50%、80%、90%的分档**，得到的因子进一步进行动量、行业、市值中性化。
- 从结果看，这种方式稳健的提升了因子表现：**随着删除水平的提高，因子IC、分位数等关键指标均得到提升**。以F-Momentum-80Pec为例：因子的IC均值达到0.029，T值达到5.78；因子分位数各组别严格单调，多空收益率达到11.6%，夏普比达到1.57，最大回撤7.2%，因子整体表现显著优于之前的F-Momentum因子。

图、F-Momentum Vs. F-Momentum-80Pec 因子多头净值



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、F-Momentum Vs. F-Momentum-80Pec 因子多空净值



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

3.1、矩阵应用改进一删除尾部噪音

➤ 即便将测试范围缩小至宽基范围内，效果依然如此：随着删除水平的提升，在宽基范围内表现也逐步提升。我们以因子在中证800内的测试为例，随着删除水平的提升，因子的IC从0.025提升至0.031，ICIR从0.32提升至0.44，T值从3.9提升至5.2，提升幅度显著。

表、删除不同噪音水平下的 F-Momentum 在宽基范围内 IC 测试

	沪深 300 内表现				中证 800 内表现	
	平均值	标准差	最小值	最大值	ICIR	T 统计量
F-Momentum	0.021	0.22	2.66	0.025	0.32	3.9
F-Momentum-20Pec	0.021	0.22	2.69	0.026	0.33	4.0
F-Momentum-30Pec	0.023	0.24	2.94	0.026	0.32	3.8
F-Momentum-50Pec	0.019	0.20	2.43	0.026	0.33	3.9
F-Momentum-80Pec	0.021	0.23	2.70	0.031	0.40	4.8
F-Momentum-90Pec	0.022	0.24	2.84	0.031	0.44	5.2

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、删除不同噪音水平下的 F-Momentum 在沪深 300 内分位数测试对比

分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率
F-Momentum				F-Momentum-20Pec				F-Momentum-30Pec			
Top	5.8%	0.23	47.7%	Top	6.7%	0.28	46.4%	Top	7.1%	0.29	48.3%
1	3.6%	0.15	44.5%	1	4.0%	0.17	42.9%	1	3.6%	0.15	41.4%
2	5.8%	0.24	41.0%	2	2.9%	0.12	45.2%	2	4.0%	0.17	39.9%
3	4.4%	0.19	44.7%	3	3.2%	0.14	43.2%	3	2.5%	0.11	45.6%
Bottom	-1.3%	-0.05	57.4%	Bottom	1.4%	0.06	53.9%	Bottom	1.0%	0.04	53.1%
L-S	7.0%	0.80	14.6%	L-S	3.8%	0.57	21.7%	L-S	5.4%	0.58	22.3%
F-Momentum-50Pec				F-Momentum-80Pec				F-Momentum-90Pec			
Top	7.5%	0.31	45.4%	Top	7.7%	0.31	43.5%	Top	7.5%	0.29	45.8%
1	2.8%	0.12	43.5%	1	3.3%	0.14	44.6%	1	4.0%	0.17	44.1%
2	2.5%	0.10	44.3%	2	5.1%	0.22	38.9%	2	3.9%	0.17	40.5%
3	4.1%	0.17	43.9%	3	0.9%	0.04	52.0%	3	2.1%	0.09	51.9%
Bottom	1.4%	0.06	52.4%	Bottom	1.3%	0.05	52.1%	Bottom	0.9%	0.04	50.4%
L-S	5.6%	0.62	21.6%	L-S	6.0%	0.63	16.6%	L-S	6.3%	0.62	17.8%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、删除不同噪音水平下的 F-Momentum 在中证 800 内分位数测试对比

分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率
F-Momentum				F-Momentum-20Pec				F-Momentum-30Pec			
Top	8.0%	0.31	50.9%	Top	8.3%	0.33	46.6%	Top	9.3%	0.36	47.7%
1	7.1%	0.29	45.9%	1	7.1%	0.29	47.9%	1	6.8%	0.27	45.5%
2	7.0%	0.29	43.1%	2	6.4%	0.26	42.5%	2	5.5%	0.22	46.0%
3	5.5%	0.22	47.3%	3	4.6%	0.19	50.8%	3	4.7%	0.19	49.9%
Bottom	-0.1%	-0.00	56.9%	Bottom	1.1%	0.04	55.1%	Bottom	1.3%	0.05	54.5%
L-S	7.8%	1.09	9.0%	L-S	3.8%	1.00	11.3%	L-S	7.5%	1.07	11.0%
F-Momentum-50Pec				F-Momentum-80Pec				F-Momentum-90Pec			
Top	8.6%	0.34	46.8%	Top	10.0%	0.39	44.4%	Top	9.9%	0.38	46.9%
1	7.7%	0.31	43.2%	1	7.4%	0.30	45.2%	1	6.9%	0.28	47.0%
2	4.9%	0.20	50.5%	2	5.6%	0.22	43.2%	2	6.5%	0.27	40.9%
3	5.3%	0.21	48.9%	3	3.7%	0.15	56.2%	3	3.6%	0.14	53.3%
Bottom	1.1%	0.04	54.0%	Bottom	0.9%	0.04	54.1%	Bottom	0.8%	0.03	54.9%
L-S	7.1%	0.97	10.8%	L-S	8.6%	1.14	10.9%	L-S	8.8%	1.20	11.6%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

3.2、矩阵应用改进—矩阵的另类利用

- 在之前的流程图中，我们经过Step2得到了财务关联矩阵。矩阵捕捉了任意两个公司的财务相似大小。以贵州茅台为例，我们是否可以直接将其他所有公司的相似度取均值，进而作为贵州茅台的因子值呢？通过这种方式我们得到全新的选股因子，称之为 Link_New。
- 从测试结果来看：Link_New因子IC达到0.042，T值高达8.18；因子分位数测试各组别严格单调，多空收益率达14.7%，夏普比达1.99，最大回撤9.1%。另外值得注意的是近几年来因子有效性非常强。

表、Link_New 因子 IC 测试结果

平均值	标准差	最小值	最大值	ICIR	T 统计量
0.042	0.06	-0.21	0.21	0.68	8.18

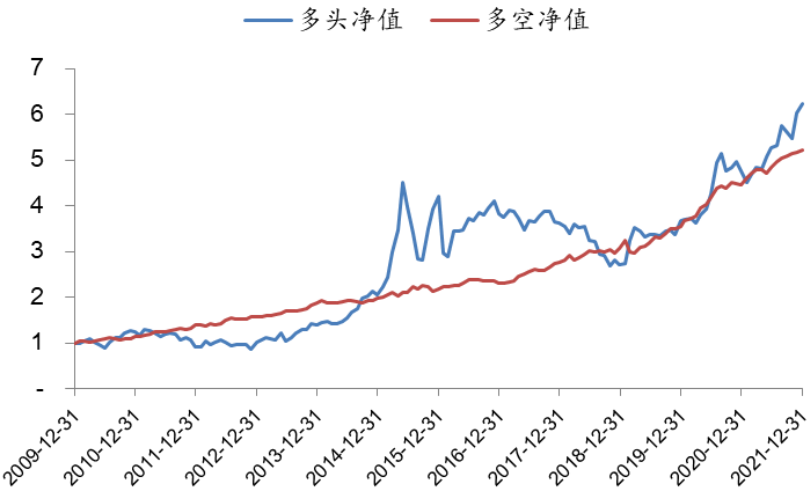
资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、Link_New 因子分位数测试结果

	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	16.4%	27.8%	0.59	52.3%	40.4%
1	12.6%	27.4%	0.46	80.3%	47.4%
2	11.6%	27.6%	0.42	91.3%	48.5%
3	10.2%	28.0%	0.36	97.2%	50.0%
4	9.7%	27.6%	0.35	100.3%	52.6%
5	9.6%	27.4%	0.35	101.3%	50.0%
6	6.4%	28.4%	0.22	100.9%	59.4%
7	6.8%	28.6%	0.24	96.4%	59.1%
8	4.7%	29.4%	0.16	85.6%	61.3%
Bottom	0.4%	30.5%	0.01	56.6%	63.3%
L-S	14.7%	7.4%	1.99		9.1%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、Link_New 因子分位数测试净值表现

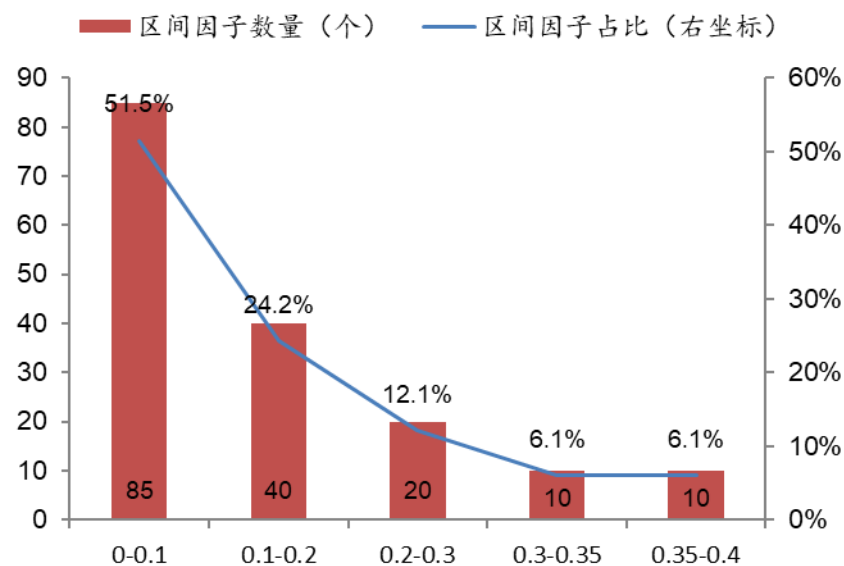


资料来源：Wind、聚源、兴业证券经济与金融研究院整理

3.2、矩阵应用改进一矩阵的另类利用

- 在之前的流程图中，我们经过Step2得到了财务关联矩阵。矩阵捕捉了任意两个公司的财务相似大小。以贵州茅台为例，我们是否可以直接将其他所有公司的相似度进行直接操作，进而作为贵州茅台的因子值呢？通过这种方式我们得到全新的选股因子，称之为Link_New。
- 针对于新构建的Link_New因子，我们进行了特异性分析：观察该因子和兴证金工底层165个选股因子的相关性，从因子相关性频率分布直方图来看，相关性高于0.3的占比为12.2%，且相关性最高仅为0.39。

图、Link_New 与各选股因子相关性分布图



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进(step2)
- 四、降维方法改进(step3)
- 五、相似底层向量改进(step1)
- 六、相似方法论本质探讨和总结

4、降维方法改进



改进方式：降维

将月度收益率替换为其他数据进行降维

替换月度收益率

收益率降维只是其中一个方式，我们这里进行了全新的探索。

基金重仓

保持步骤一和步骤二不变，而是将收益率改变成其他度量指标，比如基金重仓。FundHold表示基金持仓的公司j在t时刻的比例。

持仓动量因子

针对于公司i，利用财务关联度对基金持仓比例进行加权的结果。从这个因子构建逻辑来讲，因子值越大实际上表示和公司i财务关联度高的集群公司在考察月份基金持仓越相似（配置幅度相同），我们认为这种情况下考察公司i接下来会有上涨的可能，简称为持仓动量因子（当然这里假设基金经理具有较强的选股能力）。

4、降维方法改进

➤ 行业市值中性化的Fund-Moment因子IC达到0.032、ICIR达到0.54、T值高达6.3，分位数各组别严格单调，有效性非常强。

表、Fund-Moment 因子 IC 测试结果

	平均值	标准差	最小值	最大值	ICIR	T 值
Fund-Moment	0.031	0.12	-0.37	0.27	0.26	3.18
行业市值中性化后因子	0.032	0.06	-0.11	0.15	0.54	6.53

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、Fund_Moment 因子分位数测试结果

	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	10.7%	29.2%	0.37	57.2%	50.7%
1	12.8%	29.3%	0.44	99.3%	51.6%
2	12.0%	28.3%	0.42	114.4%	52.6%
3	11.2%	28.0%	0.40	122.0%	54.4%
4	9.9%	27.4%	0.36	125.6%	49.8%
5	9.1%	27.8%	0.33	125.0%	51.3%
6	9.2%	27.9%	0.33	120.9%	51.2%
7	5.8%	28.1%	0.21	111.6%	57.5%
8	5.1%	28.9%	0.18	94.7%	59.6%
Bottom	1.3%	30.7%	0.04	59.9%	59.6%
L-S	7.6%	14.1%	0.54		19.1%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、行业市值中性化后 Fund_Moment 因子分位数测试结果

	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	12.8%	30.2%	0.43	72.0%	49.9%
1	12.9%	28.3%	0.46	116.1%	44.7%
2	11.8%	27.9%	0.42	131.9%	48.8%
3	11.4%	27.5%	0.41	138.2%	52.1%
4	10.2%	27.2%	0.38	141.3%	48.5%
5	9.3%	26.9%	0.35	140.9%	49.7%
6	8.4%	27.3%	0.31	137.1%	52.2%
7	6.4%	27.6%	0.23	129.1%	56.4%
8	4.5%	28.6%	0.16	112.8%	61.3%
Bottom	2.5%	30.0%	0.08	70.2%	60.0%
L-S	9.7%	7.2%	1.35		10.1%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

4、降维方法改进

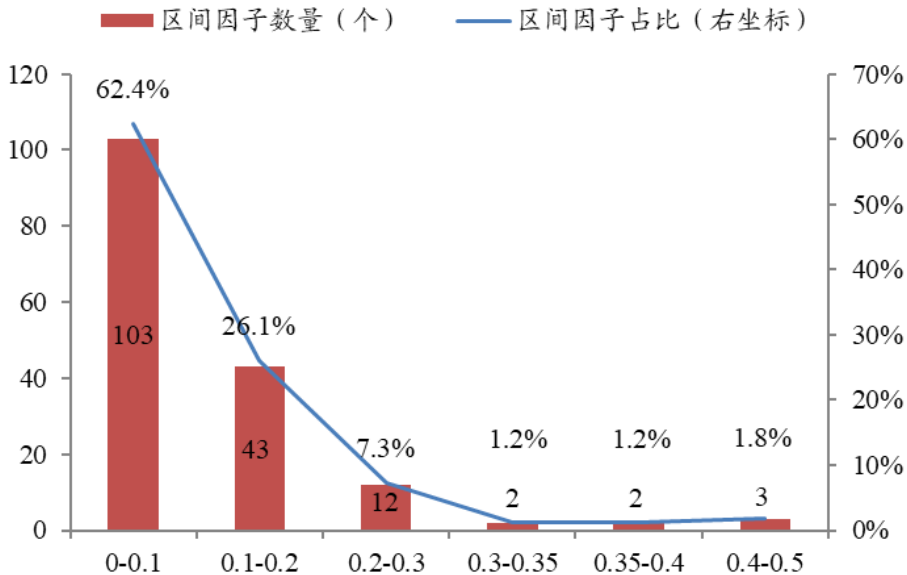
➤ 针对于新构建的Fund-Moment因子，我们分析其特异性分析：观察该因子和兴证金工底层165个选股因子的相关性。从因子相关性频率分布直方图来看，相关性高于0.3的共计7个因子，占比为4.2%（相关性最高为0.48）。**即便剥离相关性最高的5个因子，其有效性依然很强，具体参见右表。**

表、剥离相关性最高的因子后 Fund-Moment 因子 IC 测试结果

平均值	标准差	最小值	最大值	ICIR	T 统计量
0.019	0.04	-0.10	0.13	0.42	4.23

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、行业市值中性化后 Fund_Moment 因子和底层所有因子相关性分析



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进(step2)
- 四、降维方法改进(step3)
- 五、相似底层向量改进(step1)
- 六、相似方法论本质探讨和总结

5.1、相似向量尝试改进一专利动量因子

- 兴证金工团队使用的专利数据由深圳德高行知识产权数据技术有限公司提供。我们将所有上市公司过去5年的有效授权发明专利映射到IPC二级分类。
- 我们计算任意两个公司的专利分布相似度，该指标本质上代表着两家公司的专利在类别上的相似程度：指标值越大，代表两个公司在申请的专利类别上布局越相似、创新研发方向也越趋于一致。这种相似性的衡量方式，已经跨域了传统财报以及行业分类所能提供的信息，在一定程度上反映了公司未来的发展战略方向。

图表、IPC 一级分类示意图

类别代码	类别释义
A	电学
B	物理
C	机械工程、照明、加热、武器、爆破
D	固定建筑物
E	纺织、造纸
F	化学、冶金
G	作业、运输
H	人类生活需要（农、轻、医）

资料来源：德高行、兴业证券经济与金融研究院整理

[19]中华人民共和国国家知识产权局

[51]Int. Cl.³
C02F 9/04
C02F 9/02 A47J 31/00
///(C02F9/04,1: 2
8,1: 68)

[12] 实用新型专利说明书

[21] ZL 专利号 01246490.2

[45]授权公告日 2002 年 5 月 22 日

[11]授权公告号 CN 2492553Y

[22]申请日 2001.7.11

[73]专利权人 蔡忠春

地址 315411 浙江省余姚市丈亭镇宁波中联电
器有限公司

[72]设计人 蔡忠春

[21]申请号 01246490.2

[74]专利代理机构 浙江翔隆专利事务所

代理人 白洪长

权利要求书 1 页 说明书 4 页 附图页数 3 页

[54]实用新型名称 饮水机净水矿化装置

5.1、相似向量尝试改进一专利动量因子

- 由于专利动量因子构建于2019年6月份，其样本外已经2年6个月有余。由于客观原因，我们期初并没有对该因子进行实时更新。近期将因子更新至最新发现，其样本外表现着实优异。
- 从因子样本外来看，其IC达到0.023,T值达到2.8。分位数各组别相对严格单调，多头收益高达29.3%，多空收益达到14.9%，夏普比高达2.7。无论是多头亦或是多空最大回撤均非常小（多头最大回撤15.8%、多空最大回撤5.8%）。

表、IE -Momentum 因子 IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
IE-Momentum	0.023	0.05	-0.06	0.13	0.50	2.79

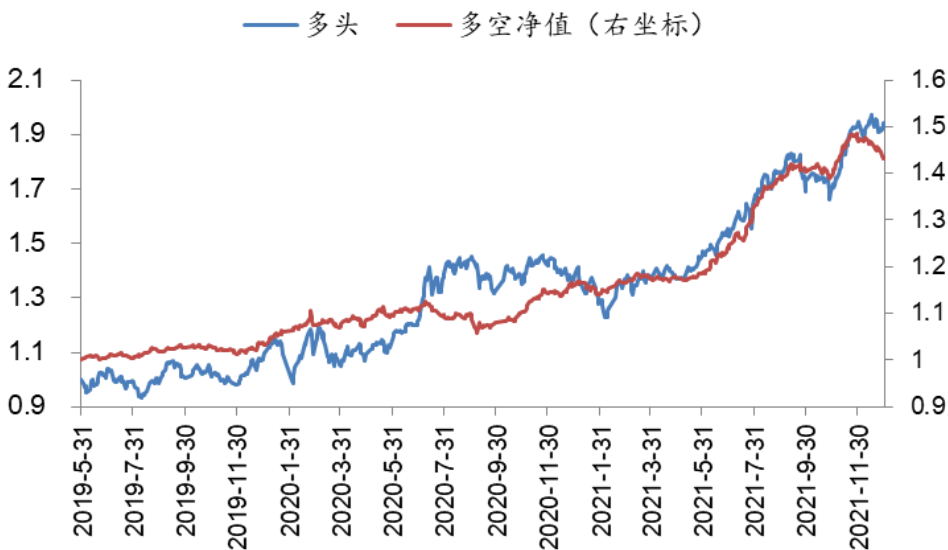
资料来源：德高行、Wind、聚源、兴业证券经济与金融研究院整理

图、IE-Momentum 因子分位数测试

	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	29.3%	21.3%	1.37	7.8%	15.6%
1	23.3%	20.9%	1.11	8.7%	15.5%
2	21.5%	20.8%	1.03	8.9%	15.9%
3	24.4%	20.6%	1.18	8.8%	16.5%
4	19.5%	20.5%	0.95	8.7%	16.6%
5	20.3%	20.2%	1.00	8.7%	15.6%
6	19.2%	20.6%	0.93	8.8%	18.3%
7	17.8%	20.7%	0.86	8.9%	20.2%
8	16.7%	20.6%	0.81	8.8%	20.2%
Bottom	13.1%	20.4%	0.64	7.8%	19.9%
L-S	14.9%	5.5%	2.71		5.8%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、IE-Momentum 因子多头&多空净值曲线

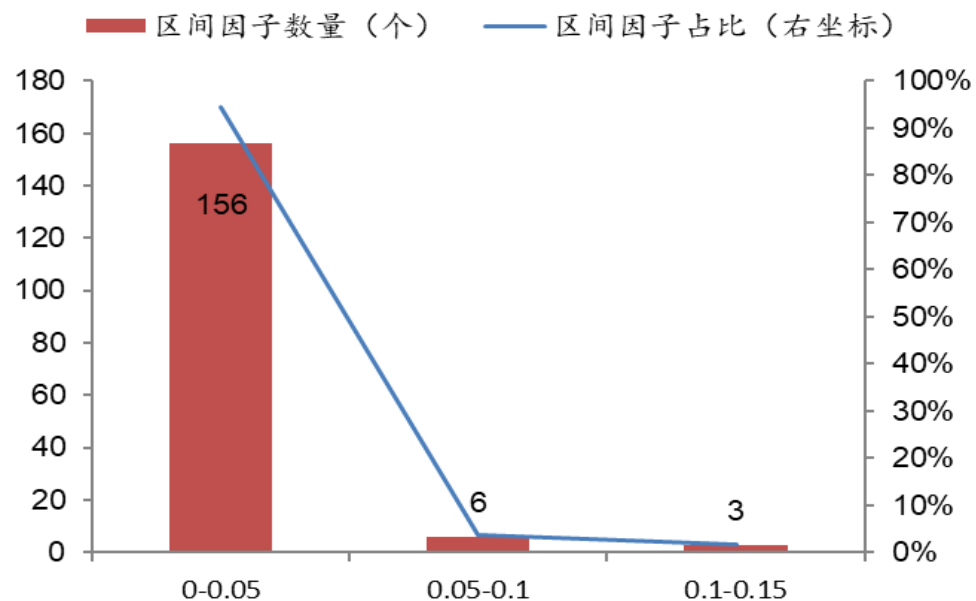


资料来源：Wind、聚源、兴业证券经济与金融研究院整理

5、相似向量尝试改进一专利动量因子

- 我们简单论证了该因子的特异性：通过测试该因子与兴证金工底层所有因子的相关性（兴证金融工程团队选股因子库共计165个因子，细分为价值、成长、质量、分析师情绪、动量反转、另类这六大类指标），我们发现相关性最高仅仅为**0.13**，这充分论证了该因子的特异性。

图、IE-Momentum 因子与各选股因子相关性分布图



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

5.2、相似向量尝试改进—另类动量因子

- 除了前面的各种改进之外，我们甚至可以调整初始的向量构成，在初始的财物动量因子中，我们从企业盈利、偿债、营运、发展四个维度出发选择一些代表性指标，以此作为基准构建了财物动量因子并以此作为基本盘对其改进进行了全方位的探讨，效果颇佳。那么在本章中，我们将打破这个最底层的约束，而是选择其他因子作为输入（添加SUE+SUR），并将删除尾部不相似的90%样本，进一步构建相似度因子。

$$Sig-link_{ijt} = \frac{\sum_{i=1}^n C_{itk} \cdot C_{jtk}}{\sqrt{\sum_{i=1}^n C_{itk}^2} \cdot \sqrt{\sum_{i=1}^n C_{jtk}^2}}$$

- 其中k代表新因子、i代表公司、t代表相应时点； C_{itk} 代表公司i在t时刻时因子k的值大小； $Sig-link_{ijk}$ 代表公司i和j在t时刻的余弦相似度，简称为关联度指标。



- 关联度是任意两个公司维度相似性的衡量方式，在一个时点实际上有3000*3000对数值（假设市场截面共计3000个在上市公司）。
- 在关联度的基础上，我们构建相似动量因子，具体参见公式

$$Sig-Momen_{it} = \frac{\sum_{i \neq j} F-link_{ijt} \cdot Ret_{jt}}{\sum_{i \neq j} F-link_{ijt}}$$

$$F-link_{ijt} > 90\% \text{相似度}$$

5.2、相似向量尝试改进一另类动量因子

- 我们测试该因子的有效性和特异性。从有效性结果来看：因子IC达到0.035、ICIR高达0.57、T值达到6.78，因子分位数测试各组别严格单调，多头收益达到15.0%，多空收益达到13.6%。

表、删除不同噪音水平下的 Sig-Momentum IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
全市场测试结果	0.035	0.06	-0.12	0.21	0.57	6.78
中证 800 内测试结果	0.036	0.07	-0.16	0.23	0.51	6.10

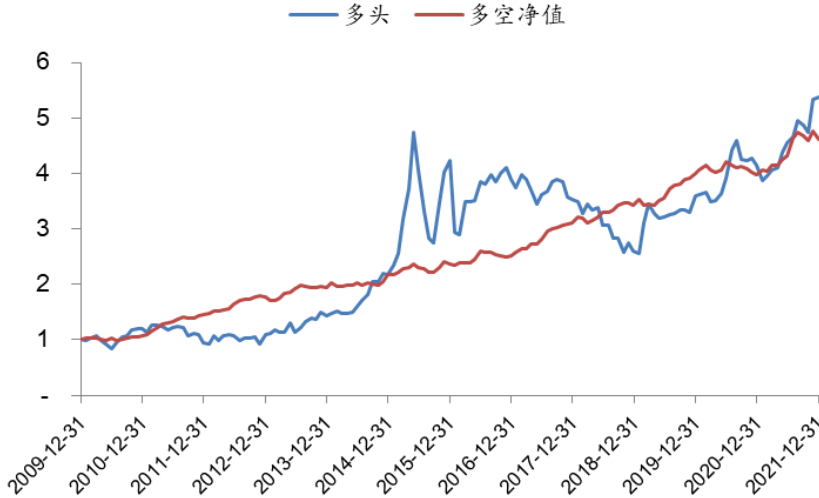
资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、行业市值中性化后 Sig-Momentum 因子分位数测试结果

分组	年化收益率	Sharpe比率	最大回撤率	分组	年化收益率	Sharpe比率	最大回撤率
全市场				中证800内			
Top	15.0%	0.52	46.2%	Top	10.7%	0.41	42.0%
1	13.2%	0.47	48.6%	1	9.5%	0.37	45.0%
2	13.2%	0.46	46.7%	2	8.8%	0.36	47.6%
3	10.9%	0.39	48.6%	3	8.1%	0.33	41.3%
4	10.8%	0.39	48.4%	4	4.7%	0.19	47.9%
5	8.7%	0.32	50.3%	5	6.2%	0.25	48.7%
6	7.4%	0.26	55.3%	6	4.7%	0.19	51.2%
7	6.6%	0.23	55.5%	7	2.4%	0.10	53.3%
8	3.8%	0.14	60.1%	8	0.4%	0.02	54.4%
Bottom	1.0%	0.03	63.4%	Bottom	-0.5%	-0.02	57.1%
L-S	13.6%	1.79	6.6%	L-S	10.9%	1.15	17.8%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、行业市值中性化后 Sig-Momentum 因子分位数净值



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

5.2、相似向量尝试改进一另类动量因子

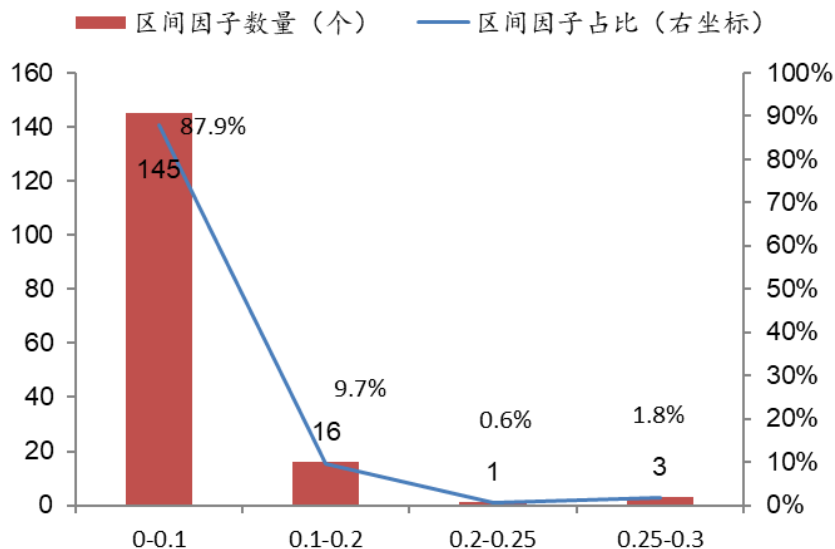
- 我们测试该因子的有效性和特异性。
从有效性结果来看：因子IC达到0.035、ICIR高达0.57、T值达到6.78，因子分位数测试各组别严格单调，多头收益达到15.0%，多空收益达到13.6%。
- 从特异性结果来看：该因子和兴证金工底层所有因子相关性最高的仅仅**0.29**（净利润同比增速因子）。

表、删除不同噪音水平下的 Sig-Momentum IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
全市场测试结果	0.035	0.06	-0.12	0.21	0.57	6.78
中证 800 内测试结果	0.036	0.07	-0.16	0.23	0.51	6.10

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、行业市值中性化后 Sig_Moment 因子分位数净值



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

- 一、引言
- 二、相似动量方法论及改进维度
- 三、矩阵应用改进
- 四、降维方法改进
- 五、相似底层向量改进
- 六、相似方法论本质探讨和总结

6、相似方法论本质探讨和总结

表、各因子逻辑简介

改进维度	简介
矩阵操作 -- 删除尾部噪音	由于市场一些股票和考察标的确实不相似，所以通过删除能提纯信息，进而提升表现（具体参见章节 3.1）→ F-Momentum-80Pec
矩阵操作 — 信息再利用	由于相似度矩阵不太适合选股，所以期初我们通过收益率进行降维，但对矩阵本身可以直接利用并产生新的因子（具体参见章节 3.2）→ Link_New
降维方式 -- 基金持仓	基于相似度矩阵，我们利用月度收益率进行降维构建财务动量因子。将月度收益率替换成基金持仓，全新的因子，效果颇佳（具体参见第四章）→ Fund-Moment
改变输入向量 — 专利	前面的所有的分析都是基于财务进行分析。我们多次强调，相似是一种信息提取的方式。这里我们将财务动量换成专利的分布，进而构建全新的专利动量，效果颇佳（具体参见 5.1）→ IE-Momentum
改变输入向量 — 其他	除了利用专利之外，我们亦可以利用其它维度信息进行研究以形成向量基础，并进一步构建相似度动量因子，同时在构建该因子时，我们借鉴了删除尾部噪音（具体参见 5.2）→ Sig-Momentum。 这里需要注意的是，由于 Sig-Momentum 构建的基础向量是在 F-Momentum-80Pec 之上，同时也利用了删除尾部噪音。所以后续复合因子时，我们不在考虑 F-Momentum-80Pec

资料来源：兴业证券经济与金融研究院整理

表、各因子相关性测试

	Sig-Momentum	Fund-Moment	Link_New	IE-Momentum
Sig-Momentum		0.11	0.09	0.05
Fund-Moment			0.32	0.01
Link_New				-0.00
IE-Momentum				

资料来源：德高行、Wind、聚源、兴业证券经济与金融研究院整理

- 至此，我们完成了相似动量目前所有的改进。百尺竿头更进一步，我们对上述各个改进的步骤和结果进行简单的归纳总结，为进一步构建复合因子打下基础。
- 总结完毕后，我们简单测试了各个因子的相关性，从结果来看，各因子相关性非常低。

6、相似方法论本质探讨和总结

- 进一步我们将Link_New、Fund-Moment、IE-Momentum、Sig-Momentum等权合成，构建复合因子，简称为Link-Comb-Sig。
- 从表现来看，**复合因子在全市场IC高达0.053、ICIR达到0.79、T值高达9.5。**因子分位数测试各组别严格单调，以在全市场测试为例：多头收益高达21.4%、多空收益高达21.9%、多空夏普比达到2.64、最大回撤4.5%。因子整体有效性非常强。
- 尤为难能可贵的是**复合因子在宽基范围内表现并没有出现衰减（IC维度）**，其中在中证800内表现与全市场几乎持平。

表、Link-Comb-Sig IC 测试

	平均值	标准差	最小值	最大值	ICIR	T 统计量
全市场测试结果	0.053	0.07	-0.11	0.22	0.79	9.50
中证 800 内测试结果	0.049	0.07	-0.11	0.25	0.67	8.08

资料来源：德高行、Wind、聚源、兴业证券经济与金融研究院整理

图、Link-Comb-Sig 因子分位数测试结果—全市场

	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	21.4%	29.3%	0.73	147.5%	37.8%
1	15.1%	28.3%	0.53	166.2%	43.7%
2	11.9%	27.6%	0.43	171.9%	48.1%
3	11.3%	27.7%	0.41	174.4%	48.4%
4	10.2%	28.3%	0.36	175.3%	50.5%
5	7.7%	27.8%	0.28	175.2%	58.2%
6	4.6%	27.6%	0.17	174.9%	60.2%
7	5.3%	28.5%	0.19	171.4%	60.5%
8	2.4%	28.4%	0.08	164.4%	61.3%
Bottom	-0.9%	29.4%	-0.03	134.4%	64.1%
市场	8.8%	28.1%	0.31		53.5%
L-S	21.9%	8.3%	2.64		4.5%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、Link-Comb-Sig 因子分位数测试结果—中证 800 内

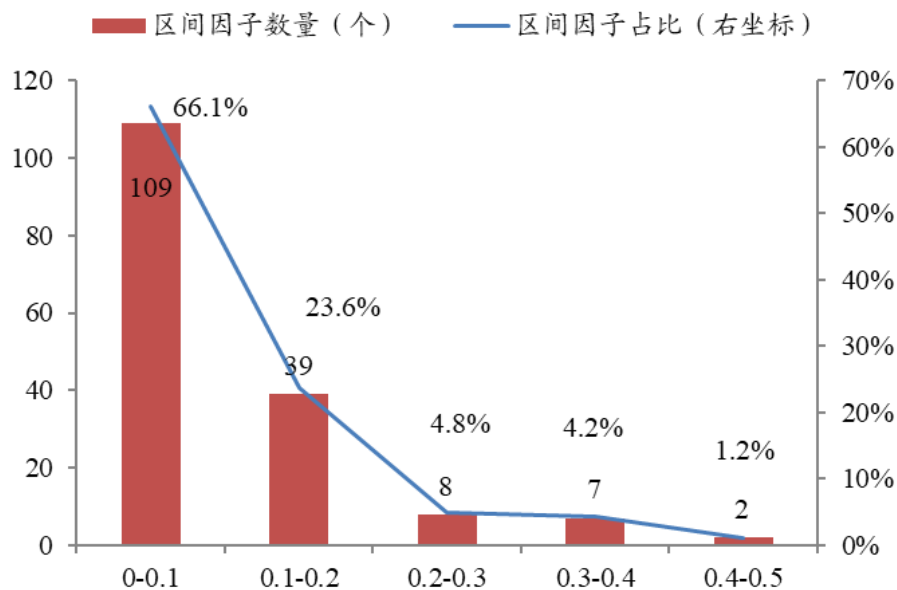
	年化收益率	波动率	Sharpe比率	平均换手率	最大回撤率
Top	13.9%	25.6%	0.54	146.3%	39.9%
1	10.6%	24.7%	0.43	166.6%	39.0%
2	7.5%	24.7%	0.31	171.5%	41.5%
3	6.9%	24.9%	0.28	173.5%	50.1%
4	7.3%	24.6%	0.30	173.3%	42.9%
5	5.4%	24.8%	0.22	175.0%	55.2%
6	2.9%	25.0%	0.12	173.8%	51.7%
7	1.9%	25.5%	0.07	171.1%	58.8%
8	0.5%	25.6%	0.02	163.7%	55.5%
Bottom	-4.1%	26.9%	-0.15	133.0%	66.0%
市场	5.3%	24.9%	0.21		50.0%
L-S	17.8%	9.4%	1.89		10.7%

资料来源：Wind、聚源、兴业证券经济与金融研究院整理

6、相似方法论本质探讨和总结

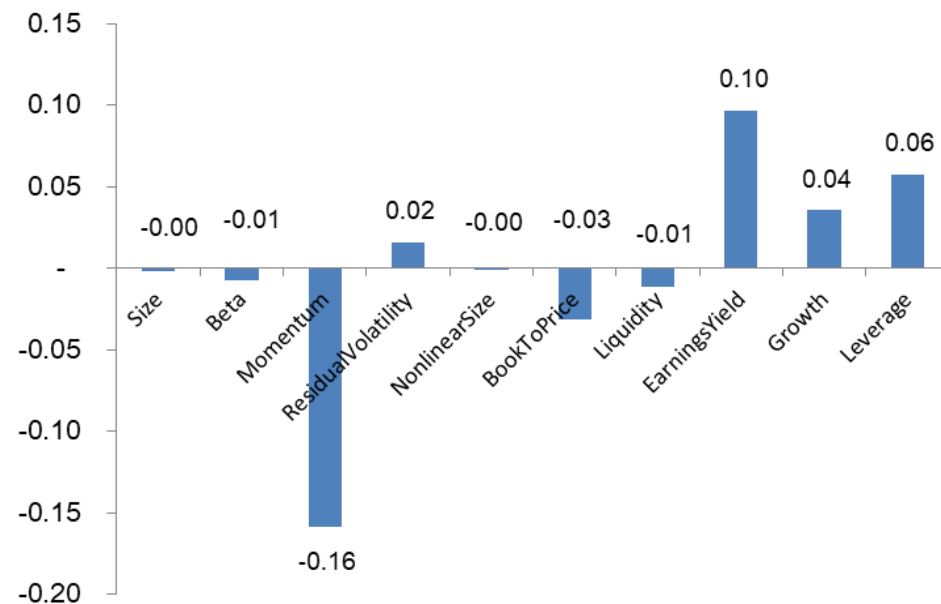
- 进一步观察因子特异性，我们通过三种方式来观察因子的特异性：直接测试该因子和底层所有因子相关性（这种方式 and 之前相同）；观察因子和Barra风格因子的相关性；将该因子放入到我们的复合模型中去看增量效应。
- 我们仍然通过极为严格的和底层所有因子相关性测试为基准进行分析，结果发现：底层所有因子和复合因子相关性极低，相关性高于0.4的仅仅只有2个。如果观察该因子和Barra的十个风格因子，我们发现其相关性也非常低，具体参见下图。

图、Link-Comb-Sig 因子相关性测试



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

图、Link-Comb-Sig 与 Barra 风格因子相关性测试



资料来源：Wind、聚源、兴业证券经济与金融研究院整理

感谢观看！

