

【金融工程专题报告】

股票网络中心度因子异象研究

❖ 股票网络中心度因子

本文测试了全 A、沪深 300 和中证 1000 在不同时间跨度下（240 分钟、22 交易日、63 交易日、126 交易日）利用最小生成树构建的网络及三种中心度因子（度中心度、接近中心度和中介中心度）的表现。

结果表明，全 A 在 22 交易日下的接近中心度因子表现较好，其多空年化收益为 24.24%，IC 均值为 0.043。

❖ 接近中心度在 22 交易日网络中表现最好的原因

从结果来看，以全 A 作为分析样本，接近中心度的表现相对于中介中心度和度中心度要更好，时间跨度 22 交易日的表现最好，而日内 1 分钟股价变化对未来一期几乎没有预测作用。

接近中心度更接近几何上的中心位置，其更能反应节点在网络中的重要程度。

度中心度仅仅利用了节点的连接数量，即节点的连接数有多少，但连接数多，并不代表节点处于网络的核心位置，中介中心度主要强调了该节点在其他节点之间的调节能力。

从时间上来说，22 交易日的跨度较好。日内 1 分钟股价变化数据噪声较大，网络不够稳定。63 个交易日或者 126 个交易日的窗口可能包含了太多的历史数据，反而降低了对当前市场条件的敏感性。

从股票池来说，越大的股票池，IC 衰减越慢，其可能原因在于越多节点数量的网络结构更加稳定。

❖ 从风险因子角度理解接近中心度可以获得收益的原因

从统计结果上来，在 1% 的置信度水平下，原始的接近中心度水平和特质波动因子呈现反向关系；在 5% 的置信度水平下，原始的接近中心度水平和动量因子呈现反向关系，和 BETA 因子呈现正向关系，和 BTOP 因子呈现出正向关系，在 10% 的置信度水平下，和市值因子和流通性因子呈现反向关系。

从特质波动因子角度，特质波动因子与未来收益呈现反向关系得到了众多实证。可能的解释是高特质波动率的股票被视为类似于彩票的股票，意味着这些股票的高回报概率低，但仍然吸引了投资者。从网络的角度上，这些特质性波动大的股票往往位于结构的边缘。

从动量因子角度来说，中心度因子高的股票受到的关注更多，因此其短期的动量更大，BETA 也更大。

❖ 定性理解接近中心度可以获得收益的原因

股票网络中的中心度越高，意味着该股票与其他股票之间的关联性更强，市场参与者关注度更高。这种高中心度的股票通常会更快、更广泛地获取市场信息。因此，当有利好消息传递时，这些股票能够迅速反映出信息，投资者可以更早地捕捉到这些机会。另外的，高中心度的股票往往会影响整个网络的其他股票。这种表现可能会扩散到其他关联股票中，从而产生溢出效应，进一步提升这些高中心度股票的市场关注度和资金流入，推高其价格。

❖ 投资建议：

利用复杂网络计算股票中心度因子可能获得一定超额收益。

❖ 风险提示：

基于历史结果计算不代表未来收益。

相关研究报告

《宏观因子 FMP 组合》

2024-08-16

《量化复盘光伏行业与行业动态因子建模》

2024-08-03

《2024 年二季报公募基金十大重仓股持仓分析》

2024-07-23

《形态学研究之十：6 种新 K 线形态》

2024-07-11

《短期信号以中性或偏空为主，市场或仍处于筑底过程——2024 年二季度策略总结与未来行情预判》

2024-07-10

投资主题

报告亮点

本文将复杂系统网络引入股票市场获得个股的中心度因子，经过测试发现，三种股票中心度因子中，接近中心度因子表现最好。从时间跨度看利用 22 交易日股价变化构建的网络表现较好。股票池越大，IC 衰减越慢。此外，本文尝试从风险因子定量和定性角度理解这一异象。

投资逻辑

本研究利用复杂网络理论刻画股票市场的网络结构，在网络中，接近中心度越大的股票未来收益可能更好。

目 录

一、基于股价变动构建网络	6
(一) 常见基于股价变化构建网络的方式.....	6
(二) 衡量节点位于网络位置的指标.....	7
1、度中心度.....	7
2、接近中心度.....	7
3、中介中心度.....	8
二、因子计算流程	9
(一) 因子评价指标.....	9
1、回归法.....	9
2、IC 法.....	9
3、分层测试.....	10
(二) 度中心度因子表现.....	10
(三) 接近中心度因子表现.....	12
(四) 中介中心度因子表现.....	14
三、因子收益 讨论与总结	17
(一) 接近中心度在 22 交易日网络中表现最好的原因.....	17
(二) 接近中心度与风险因子的关系.....	18
(三) 接近中心度因子收益的定性理解.....	20
四、风险提示	21
五、参考文献	22
六、附录	23
(一) 沪深 300 度中心度因子表现.....	23
(二) 中证 1000 度中心度因子表现.....	23
(三) 沪深 300 接近中心度因子表现.....	24
(四) 中证 1000 接近中心度因子表现.....	25
(五) 沪深 300 中介中心度因子表现.....	25
(六) 中证 1000 中介中心度因子表现.....	26

图表目录

图表 1	中证白酒网络示例	7
图表 2	中证白酒三种中心度	8
图表 3	度中心度因子表现	10
图表 4	度中心度因子在不同股票集合和不同时间跨度下的 IC 表现	11
图表 5	IC 累计走势	11
图表 6	IC 半衰期曲线	11
图表 7	分组年化收益图	12
图表 8	分组累计收益图	12
图表 9	接近中心度因子表现	12
图表 10	接近中心度因子在不同股票集合和不同时间跨度下的 IC 表现	13
图表 11	IC 累计走势	13
图表 12	IC 半衰期曲线	13
图表 13	分组年化收益	14
图表 14	分组累计收益图	14
图表 15	中介中心度因子表现	14
图表 16	中介中心度因子在不同股票集合和不同时间跨度下的 IC 表现	15
图表 17	IC 累计走势	15
图表 18	IC 半衰期曲线	15
图表 19	分组年化收益	15
图表 20	分组累计收益图	15
图表 21	IC 均值表现	17
图表 22	t 均值表现	17
图表 23	接近中心度和风险因子的平均相关性和平均显著性	18
图表 24	接近中心度和风险因子相关性滚动 20 日平均	19
图表 25	沪深 300 度中心度因子 IC 累计图	23
图表 26	沪深 300 度中心度因子 IC 半衰期图	23
图表 27	沪深 300 度中心度因子分组年化收益	23
图表 28	沪深 300 度中心度因子分组累计收益	23
图表 29	中证 1000 度中心度因子 IC 累计图	23
图表 30	中证 1000 度中心度因子 IC 半衰期图	23
图表 31	中证 1000 度中心度因子分组年化收益	24
图表 32	中证 1000 度中心度因子分组累计收益图	24
图表 33	沪深 300 IC 接近中心度因子累计图	24

图表 34	沪深 300 接近中心度因子 IC 半衰期图	24
图表 35	沪深 300 度中心度因子分组年化收益	24
图表 36	沪深 300 接近中心度因子分组累计收益	24
图表 37	中证 1000 IC 接近中心度因子累计图	25
图表 38	中证 1000 IC 接近中心度因子 IC 半衰期图	25
图表 39	中证 1000 接近中心度因子分组年化收益	25
图表 40	中证 1000 接近中心度因子分组累计收益	25
图表 41	沪深 300 IC 中介中心度因子累计图	25
图表 42	沪深 300 IC 中介中心度因子 IC 半衰期图	25
图表 43	沪深 300 中介中心度因子分组年化收益	26
图表 44	沪深 300 中介中心度因子分组累计收益	26
图表 45	中证 1000 IC 中介中心度因子累计图	26
图表 46	中证 1000 IC 中介中心度因子 IC 半衰期图	26
图表 47	中证 1000 中介中心度因子分组年化收益	26
图表 48	中证 1000 中介中心度因子分组累计收益	26

一、基于股价变动构建网络

（一）常见基于股价变化构建网络的方式

梳理复杂网络方法在金融市场中的应用，Mantegna 最早将复杂网络研究方法引入股票市场，Barabasi 在 2003 年的研究中证实美国股票市场满足无标度网络性质。在国内股票市场的研究中，Huang 在 2009 年的研究中证实沪深 1080 只股票满足幂律分布，Mai 在 2014 年的复杂网络系统研究表明沪深 300 的核心行业为工业，谢凌峰和陈志 2016 年的研究表明在沪港通开通后网络的平均度和聚类系数都有了显著的上升。

在上述的研究中，构建网络的方法大致可以分为三类，第一类是类似牛晓健（2021）的做法，采用阈值截断法(Threshold Truncation Method, TTM)构建网络；第二类是类似于 Huang（2009）的做法，采用平面最大过滤图法(Planar Maximally Filtered Graph, PMFG)构建网络；第三类是类似于谢赤（2021）的做法利用最小生成树(Minimum Spanning Tree, MST)法构建网络。

最小生成树法对网络信息的过滤最为严格，只提取网络中各节点间相关性最大的主要信息；平面最大过滤图方法最大限度上保留了网络中与有效信息相关的连边，但有效信息分析相对复杂、网络计算速度相对较慢；阈值截断法基于不同阈值构建的网络结构差异较大。

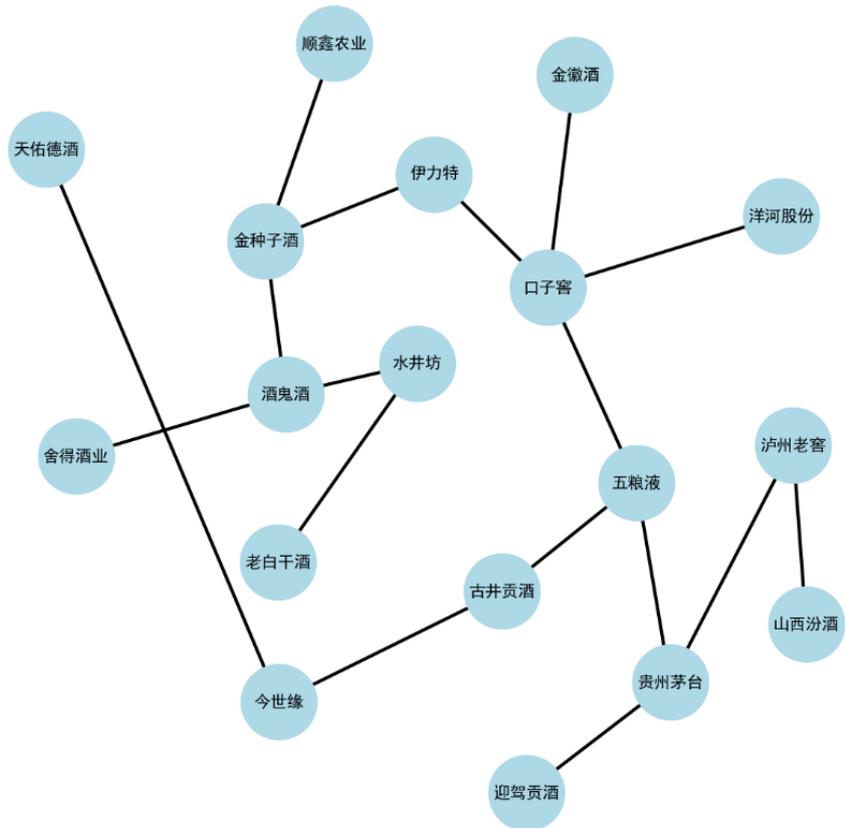
本文产生因子的方法主要是利用最小生成树法构建网络。构建的方法如下：

- 1 计算过去一段时间股票 i 和股票 j 股价变动的相关性，计为 ρ_{ij} ；
- 2 将相关性转化为股票 i 和股票 j 之间的距离 d_{ij} ， $d_{ij} = \sqrt{2(1 - \rho_{ij})}$ ；
- 3 当获得不同股票之间的距离以后，可以通过 Prim 算法和 Kruskal 算法构建网络。

最终构建的网络应当满足，所有边的权重之和最小，包含图中的所有顶点，并且没有环。

例如 2024 年 8 月 30 日采用过去 22 交易日利用最小生成树法构建中证白酒网络示例如下：

图表 1 中证白酒网络示例



资料：wind，华创证券

注：网络图根据算法于2024年8月30日采用过去22交易日股价变动绘制

在中证白酒网络中，可以观察到不同股票（节点）在网络中的位置不同，例如顺鑫农业位于网络的外层，其仅与金种子酒连接，而五粮液位于内层，其与口子窖、古井贡酒和贵州茅台都有连接。

（二）衡量节点位于网络位置的指标

为了衡量节点在网络中的位置和重要性，定义以下节点位置衡量指标。

1、度中心度

节点度中心度，计算方法为节点的连接的边的数量除以可能的最大连接数，对于 n 个节点的网络，最大连接数为 n-1，即：

$$d_v = \text{deg}(v)/m$$

其中 deg 为节点 v 的连接边数量，m 为最大连接数 (n-1)。

2、接近中心度

接近中心度衡量了节点与其他节点的距离，计算方法为加总节点到其他所有节点的最短

距离，接近中心度也通过最大连接距离进行归一化，即：

$$c_v = (n-1)/d$$

其中 d 为节点与其他节点连接距离的加总。

3、中介中心度

中介中心度衡量了，节点位于其他节点的最短路径上的程度，即：

$$b_v = \sum_{s,t} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

$\sigma(s,t)$ 为节点 s, t 的最短路径的数量， $\sigma(s,t|v)$ 为 s 和 t 最短路径经过节点 v 的数量。

图表 2 中证白酒三种中心度

	度中心度	接近中心度	中介中心度
泸州老窖	0.118	0.236	0.118
古井贡酒	0.118	0.283	0.221
酒鬼酒	0.176	0.258	0.324
五粮液	0.176	0.354	0.603
顺鑫农业	0.059	0.236	0.000
洋河股份	0.059	0.274	0.000
天佑德酒	0.059	0.189	0.000
伊力特	0.118	0.340	0.485
金种子酒	0.176	0.304	0.471
贵州茅台	0.176	0.293	0.324
老白干酒	0.059	0.177	0.000
舍得酒业	0.059	0.207	0.000
水井坊	0.118	0.213	0.118
山西汾酒	0.059	0.193	0.000
迎驾贡酒	0.059	0.230	0.000
今世缘	0.118	0.230	0.118
口子窖	0.235	0.370	0.640
金徽酒	0.059	0.274	0.000

资料 : wind, 华创证券

我们将三种中心度作为因子，计算其在不同时间跨度和样本中的表现。由于股票之间的网络由股票价格变动的相关性计算而来，而股票之间的相关性与其选取的时间跨度相关。我们考察了不同选取时间段和不同样本下的网络及其对应的中心度因子。

我们分别计算了在日内（240min）、22 交易日（22d）、63 交易日（63d）和 126 交易日（126d）时间跨度下的中心度，样本分别考察了全 A、沪深 300 和中证 1000 样本下的表现，调仓周期为 5 个交易日。

二、因子计算流程

(一) 因子评价指标

为了阅读的流畅性，这里简要介绍常用的因子处理方法和评价标准。

1、回归法

即在 T 期截面上，计算个股因子暴露对 T+1 期收益的解释力度。

$$r^{T+1} = X^T a^T + \sum_j Ind_j^T b_j^T + \ln(MV)^T b^T + \varepsilon^T$$

其中， r^{T+1} 所有个股在第 T+1 期的收益率向量， X^T 为个股在第 T 期的单因子的暴露度向量， Ind_j^T 为个股的行业哑变量， b_j^T 为行业收益， $\ln(MV)^T$ 为流通市值取对向量， b^T 为市值收益， ε^T 为残差向量。

这里个股的第 T 期的单因子的暴露度向量，首先对原始因子进行中位数去除极值，对于大于阈值的因子值直接替换为阈值，并进行标准化处理，若有缺失值，缺失值直接剔除。

为了求解上述方程，本文参考 barra cne6 的处理方法，对上述方程进行加权最小二乘回归，回归系数为流通市值取平方根。

回归可以获得因子的因子收益率序列(即所有截面期回归系数构成的序列)，对于每一期的回归，可以获得假设检验： $a^T = 0$ 和备择假设： $a^T \neq 0$ 。假设检验对应的 t 统计量为：

$$t = \frac{a^T}{SE(a^T)}$$

其中 $SE(a^T)$ 为回归的标准误差。比较 t 统计量的绝对值是否大于临界值 (1.96 对应 5% 显著性水平，实际使用中通常取 2)，若大于则认为系数显著。

对于回归法，通常关注：

- 1) a 的均值，其反应因子收益大小；
- 2) t 值序列的绝对值的均值，其反映了因子是不是显著；
- 3) t 值序列大于 2 的占比，其反应因子显著性是不是稳定；
- 4) t 值均值，其反应因子方向上是不是稳定。

2、IC 法

即计算当期因子暴露与下一期收益的 Spearman 相关系数 (Rank IC)。

$$IC_T = \text{corr}(r^{T+1}, X^T)$$

先对因子进行中位数去除极值，并进行行业 and 市值的中性化处理。

对于 IC 序列，通常关注：

- 1) IC 序列均值，其反应因子的显著性；
- 2) IC 序列标准差，其反应因子的稳定性；
- 3) ICIR，即 IC 序列均值除以 IC 序列标准差，其兼顾了因子的稳定性和显著性；

4) IC 大于 0 的比例, 反应因子方向是不是稳定。

5) IC 的半衰期, 即计算当期因子暴露与接下来第 N 期收益的 IC, 观察 IC 序列均值的下降速率。

3、分层测试

对因子进行中位数去除极值并进行行业和市值中性化以后, 将每一期因子按照从小到大分成 5 层, 每层组内股票等权, 测试其组合的收益等表现情况。

通常关注:

1) 分组的年化收益和累计收益, 反应因子分层单调性;

2) 分组的多空 (第五组减去第一组的组合收益) 收益表现, 包括多空年化收益、多空回撤情况、多空的胜率表现等, 反映了因子收益的稳健性和单调性;

3) 分组组合的换手率, 反应实际交易的换手水平。

(二) 度中心度因子表现

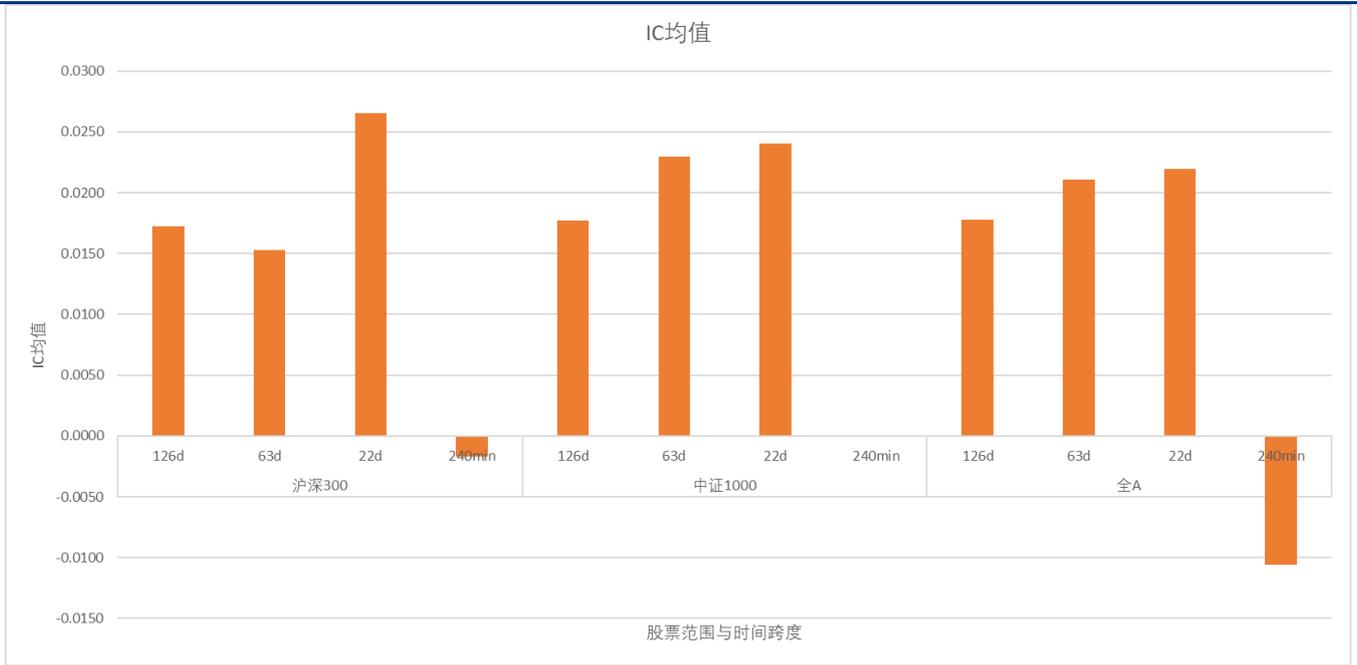
图表 3 度中心度因子表现

股票集	时间跨度	t 绝对值平均	t 大于 2 占比	t 均值	因子收益率均值	IC 均值	IC 标准差	ICIR	IC 大于 0 比例	多空年化 (%)	多空夏普比例	多空最大回撤	多空胜率	最大因子组合换手率	最小因子组合换手率
沪深 300	126d	0.91	6.5%	0.2213	0.0790	0.0172	0.0598	0.2882	61%	14.87	1.70	-12.22%	57.20%	28.64%	33.26%
	63d	0.96	6.8%	0.2295	0.0751	0.0153	0.0624	0.2446	60%	12.46	1.43	-11.08%	57.77%	38.40%	41.82%
	22d	1.04	9.6%	0.3998	0.1282	0.0266	0.0640	0.4149	64%	20.93	2.30	-6.19%	59.32%	60.02%	61.49%
	240min	1.23	9.1%	-0.1133	-0.1451	-0.0017	0.0696	-0.0245	48%	1.65	0.22	-19.15%	50.79%	67.52%	64.30%
中证 1000	126d	1.04	10.0%	0.3877	0.0684	0.0177	0.0538	0.3292	64%	10.46	1.78	-5.94%	60.31%	26.42%	32.17%
	63d	1.16	13.8%	0.5064	0.0855	0.0230	0.0518	0.4433	69%	11.76	2.05	-3.82%	62.07%	37.64%	42.19%
	22d	1.16	14.0%	0.4910	0.0870	0.0240	0.0461	0.5211	70%	11.03	2.06	-6.05%	59.75%	60.02%	62.66%
	240min	1.35	9.6%	-0.1494	0.0931	0.0001	0.0585	0.0011	46%	3.93	0.58	-25.72%	51.17%	72.29%	68.74%
全 A	126d	1.69	24.4%	0.5813	0.0503	0.0178	0.0446	0.3976	65%	7.57	1.50	-7.81%	59.46%	22.42%	26.77%
	63d	1.82	28.0%	0.7352	0.0647	0.0211	0.0431	0.4889	69%	9.93	2.04	-5.13%	62.43%	33.53%	37.49%
	22d	1.78	28.5%	0.7813	0.0667	0.0220	0.0388	0.5660	72%	10.79	2.36	-4.32%	61.86%	56.97%	59.07%
	240min	2.18	16.4%	-0.4702	0.0757	-0.0106	0.1007	-0.1051	43%	-5.36	-0.48	-57.88%	43.48%	70.53%	71.12%

资料 : wind, 华创证券

考虑 IC 在不同分组中的表现。

图表 4 度中心度因子在不同股票集合和不同时间跨度下的 IC 表现

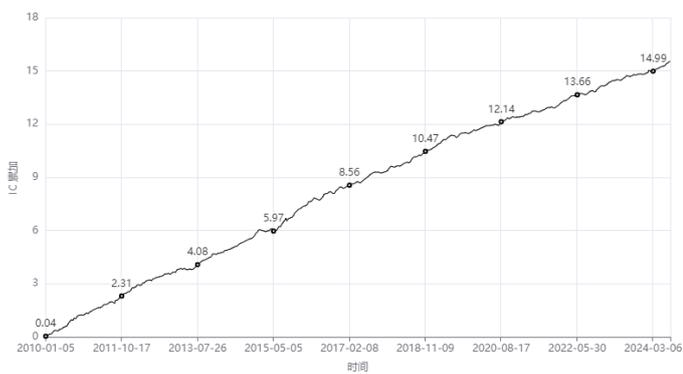


资料 : wind, 华创证券

从 IC 序列上来看, 在不同的股票集中, 日内 (240 分钟) 因子表现较差, 而 22 日中心度因子表现较好。

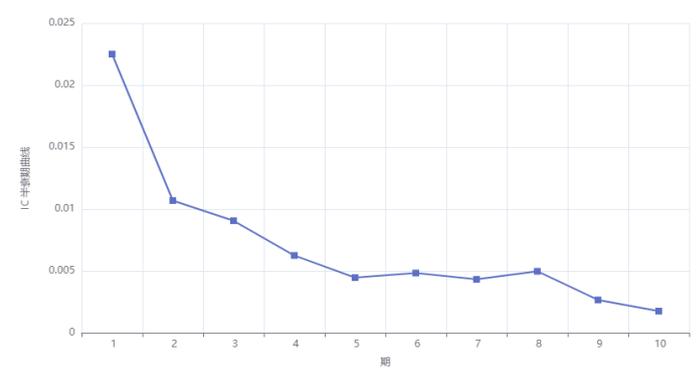
以全 A 为例, 22 交易日跨度的因子的表现如下图所示:

图表 5 IC 累计走势



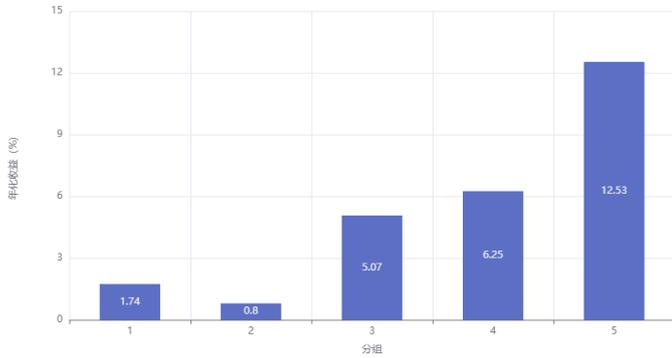
资料 : wind, 华创证券

图表 6 IC 半衰期曲线



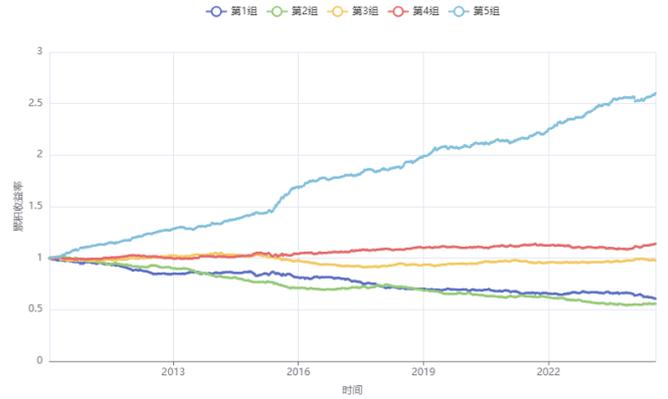
资料 : wind, 华创证券

图表 7 分组年化收益图



资料 : wind, 华创证券

图表 8 分组累计收益图



资料 : wind, 华创证券

从结果上看，度中心度因子的 IC 表现稳定，从 IC 衰减来看，其 IC 衰减较快，分层效果不佳。

沪深 300 和中证 1000 的 22 交易日度中心度结果见附录。沪深 300、中证 1000 和全 A 对比来看，度中心度因子表现出了因子非线性收益表现，第五组收益显著跑赢剩余四组，且剩余四组的收益分层不明显。

从因子的半衰期来看，全 A 的 IC 衰减要慢于沪深 300 和中证 1000。可能原因是股票池越大，节点数量越多，其网络稳定性越好。

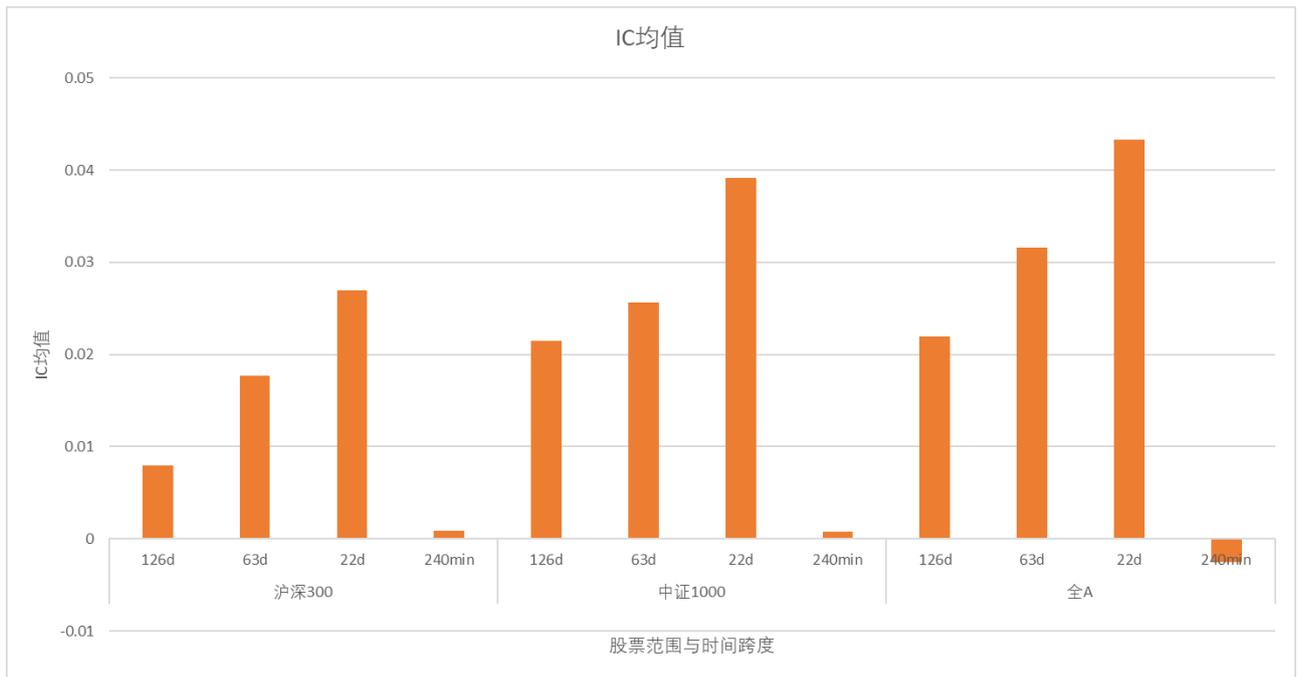
(三) 接近中心度因子表现

图表 9 接近中心度因子表现

中心度衡量指标	时间跨度	t 绝对值平均	t 大于 2 占比	t 均值	因子收益率均值	IC 均值	IC 标准差	ICIR	IC 大于 0 比例	多空年化 (%)	多空夏普比例	多空最大回撤	多空胜率	最大因子组合换手率	最小因子组合换手率
沪深 300	126d	1.34	12.6%	0.0892	0.0597	0.0079	0.0872	0.0910	55%	7.45	0.69	-20.44%	52.12%	42.10%	46.99%
	63d	1.45	16.4%	0.2637	0.1133	0.0177	0.0895	0.1984	57%	15.53	1.38	-13.90%	57.06%	51.79%	56.92%
	22d	1.52	18.1%	0.3802	0.1477	0.0270	0.0957	0.2822	62%	16.89	1.36	-19.53%	59.32%	65.36%	68.48%
	240min	1.26	9.1%	0.0264	-0.2189	0.0009	0.0751	0.0118	52%	1.90	0.26	-22.68%	52.17%	73.79%	74.27%
中证 1000	126d	1.68	22.0%	0.4581	0.0863	0.0215	0.0693	0.3102	63%	11.92	1.69	-7.78%	59.87%	46.28%	51.19%
	63d	1.58	22.2%	0.5268	0.0975	0.0256	0.0640	0.4004	65%	13.82	2.04	-4.14%	60.13%	55.00%	60.66%
	22d	1.70	28.6%	0.8835	0.1617	0.0391	0.0639	0.6127	73%	21.95	3.21	-5.56%	68.43%	65.10%	69.78%
	240min	1.44	11.0%	-0.0098	0.0237	0.0007	0.0573	0.0131	48%	1.43	0.22	-14.02%	52.64%	78.29%	78.65%
全 A	126d	2.54	31.2%	0.7274	0.0672	0.0220	0.0564	0.3892	66%	10.94	1.92	-9.87%	60.03%	44.15%	47.90%
	63d	2.65	35.9%	1.0262	0.1001	0.0316	0.0578	0.5466	72%	16.44	2.74	-5.51%	66.38%	53.68%	57.92%
	22d	2.78	39.8%	1.5197	0.1375	0.0433	0.0594	0.7296	77%	24.24	3.82	-2.68%	73.45%	63.32%	67.03%
	240min	2.81	18.0%	-0.8521	-0.0676	-0.0026	0.0946	-0.0271	43%	-0.90	-0.09	-18.24%	46.28%	75.18%	78.51%

资料 : wind, 华创证券

图表 10 接近中心度因子在不同股票集合和不同时间跨度下的 IC 表现

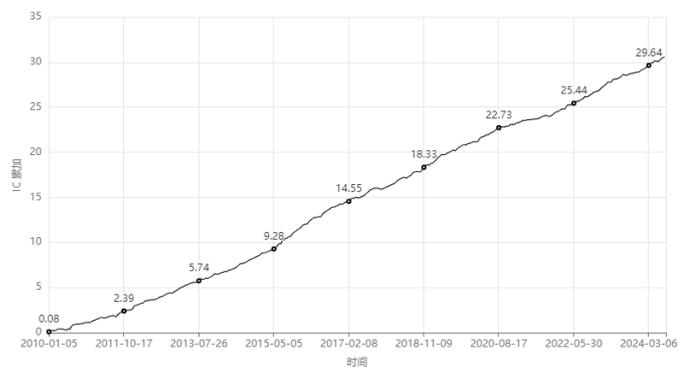


资料 : wind, 华创证券

从IC序列上来看,在不同的股票集中,日内(240分钟)因子表现较差,而22日中心度因子表现较好。

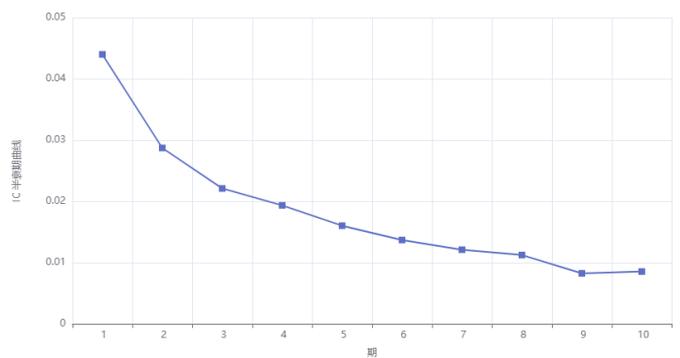
以全A为例,22交易日接近中心度的表现如下图所示:

图表 11 IC 累计走势



资料 : wind, 华创证券

图表 12 IC 半衰期曲线



资料 : wind, 华创证券

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/327060132153006150>