中心度因子

研报复现结题报告

目录

[第一章 研报分析 2](#_Toc5325)

[1.1 思路 2](#_Toc10099)

[1.1.1 类比复杂网络建模 2](#_Toc16224)

[1.1.2 因子构建 2](#_Toc20776)

[1.1.3 因子计算细节 3](#_Toc29590)

[1.2 方法步骤 3](#_Toc6060)

[1.3 主要结论 4](#_Toc4407)

[第二章 研报复现 5](#_Toc7008)

[2.2 过程 5](#_Toc24566)

[2.3 小结 19](#_Toc21176)

[2.4 数据差异分析对比 20](#_Toc15318)

[第三章 研报评价 22](#_Toc29568)

[3.1 创新点 22](#_Toc3557)

[3.2 不足点 22](#_Toc27448)

# 第一章 研报分析

## 1.1 思路

### 1.1.1 类比复杂网络建模

复杂网络是一个主要具有以下三种特征的网络。第一是集群效应；第二是拥有核心；第三是“小世界”特性，即网络中任意两点总能经由至多六个点联结起来。股票市场中耳熟能详的行业联动效应、板块联动效应对应着复杂网络的集群效应特征；龙头股效应对应着复杂网络拥有核心的特点；系统性风险则对应着复杂网络的“小世界”特性。由此，作者自然地想到用股票间相关性去代替复杂网络中点与点之间的线段，构造出一个基于相关性的股票网络。

### 1.1.2 因子构建

首先定义股票间的相关性即为收益率之间的相关性。



对于某股票i, 定义因子为其收益率和其他股票收益率相关系数的加权平均。



设置权重为股票j在市场指数中的权重，若表示市场指数收益率，则



由于黄玮强(2008)研究得出股票的聚类结构和行业分类非常贴近，故作者需要因子还要能刻画与申万一级行业指数的相关性。于是扩写经典套利定价模型（APT），令

则该回归方程的即为大盘指数与行业指数对个股收益率的解释程度。为了使得因子正态分布，最终定义



### 1.1.3 因子计算细节

收益率：收益率都取对数收益率

回归：使用过去23个交易日的数据进行回归

停牌：停牌时收益率取0

剔除：剔除停牌10日股票

## 1.2 方法步骤

① 绘制个股的走势与收益率走势，查看对于收益率的预示性作用如何。

② 绘制中心度因子P与各大常见因子的相关系数热力图。

③ 单因子回测：回测区间是2007年1月到2016年10月。

根据因子值将全部A股分成10组，月度调仓。在全市场、沪深300和中证500三个不同的股票池查看第一组的分年选股表现。

④ 查看因子P各组的行业和市值分布。

⑤ 选股能力检验：得出因子P的rankIC。

⑥ 稳定性检验：得出因子P的自相关系数。

⑦ 正交性检验：用Fama MacBeth得出风格因子对因子P的解释能力。

⑧ 对因子P进行行业与市值中性化之后重复步骤④⑤⑥

## 1.3 主要结论

因子区分度：具有显著的收益区分能力。

市值和行业偏好：一般来说，股票流通市值越大，中心度越高；同时有色、煤炭、钢铁、非银、房地产，银行等市值与权重较大的行业的股票的中心度也一般较高。

回测表现：

未做市值和行业中性化前，第一组相对于中证 500 指数，年化超额收益为 13.6%，相对最大回撤为 16.2%，相对胜率为 64%，信息比率为 1.43，中心度因子 P 的 rankIC 均值为 0.064， IC\_IR 为 0.715。

做了市值和行业中性化后，中心度因子的收益区分能力显著提高。第一组年化收益有 33.8%，平均每年超额中证 500 指数 19.6%，超额胜率为83.9%，信息比率为 2.82；中性因子的 rankIC 的均值在 0.078 上下波动，IC\_IR 为 0.928。正显著比例为 61.5%，负显著比例为 6%。

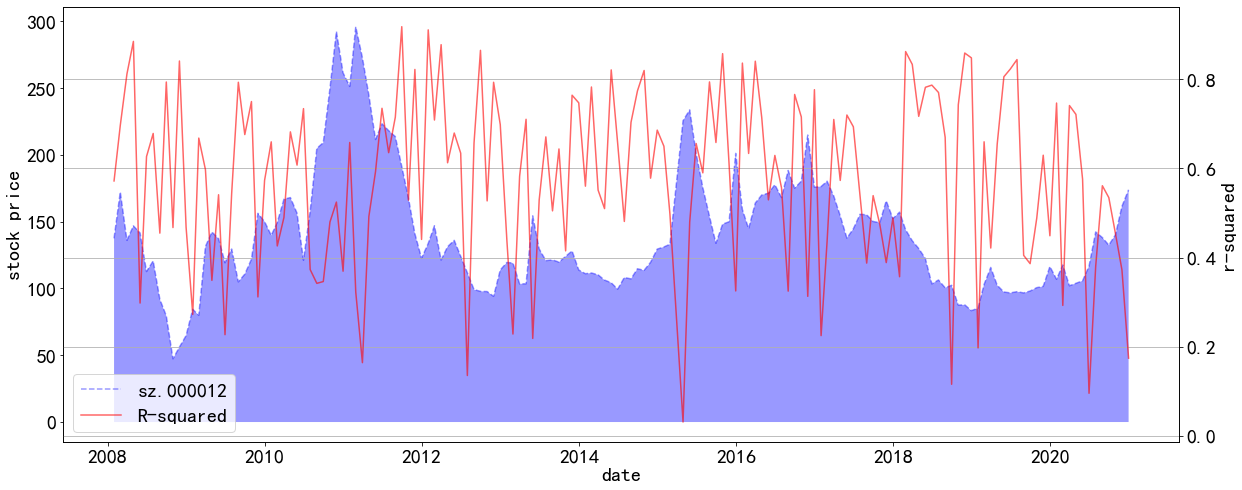
# 第二章 研报复现

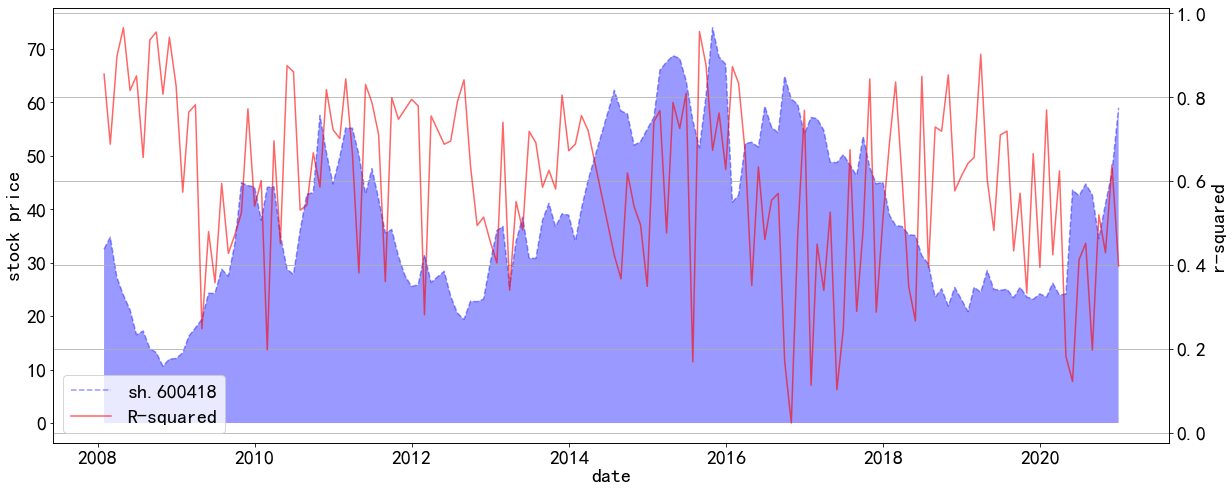
## 2.2 过程

回测时间取2008年1月至2020年12月底。

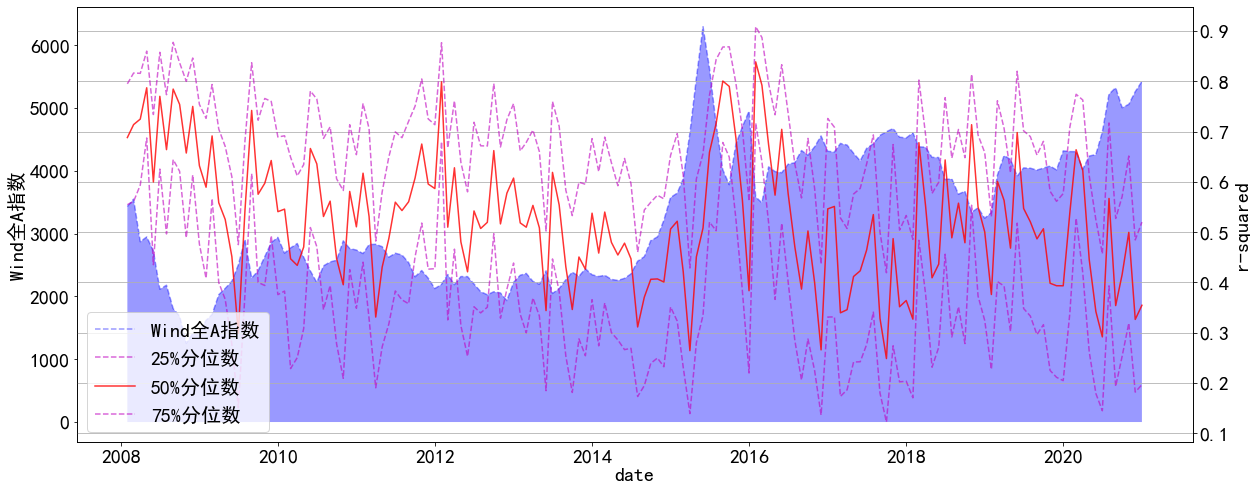
①走势与收益率走势图

例图：个股南坡A收益率走势及其走势

****例图：个股江淮汽车收益率及其走势

****

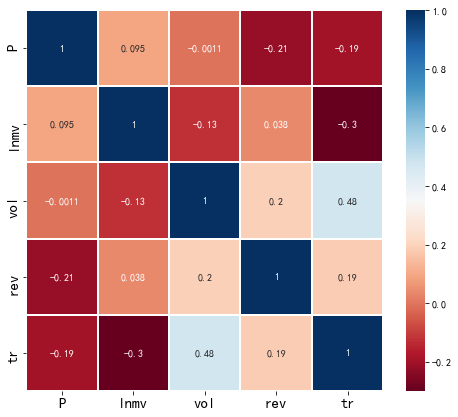
例图：Wind全A指数及全市场分位数走势

****

总体来说，当股票的处于一个历史低位的时候，一般对应着股价的阶段顶部，而处于历史高位的时候，一般对应着股价的阶段底部。

但通过仔细观察我们发现把作为买卖指标的准确性在不同行情下有显著不同。我们观察到18年年初到年中这段时间，先是在一个历史高位，可是股价并没有出现触底反弹，反而一路下跌；到了18年年中，已经处于历史低点而股价同样处于历史低点。倘若我们使用单因子策略，完全依靠作为买卖信号，即在取极大值时买入，取极小值时卖出，那么18年年初到年中会是损失惨重的一段时间。同样地，我们发现在11年的“跌跌不休”行情中，利用作为买卖信号同样会造成大额损失。相反地，我们容易发现：在上涨行情或是震荡行情中，使用作为买卖指标的准确性是十分值得信赖的。

②中心度因子P与常见因子的相关系数矩阵

我们计算了中心度因子 P 与一些常见的风格因子之间的相关系数。其中包括lnmv(规模因子)， rev(反转)，tr(过去一个月日均换手率)，vol(过去一个月波动率)。

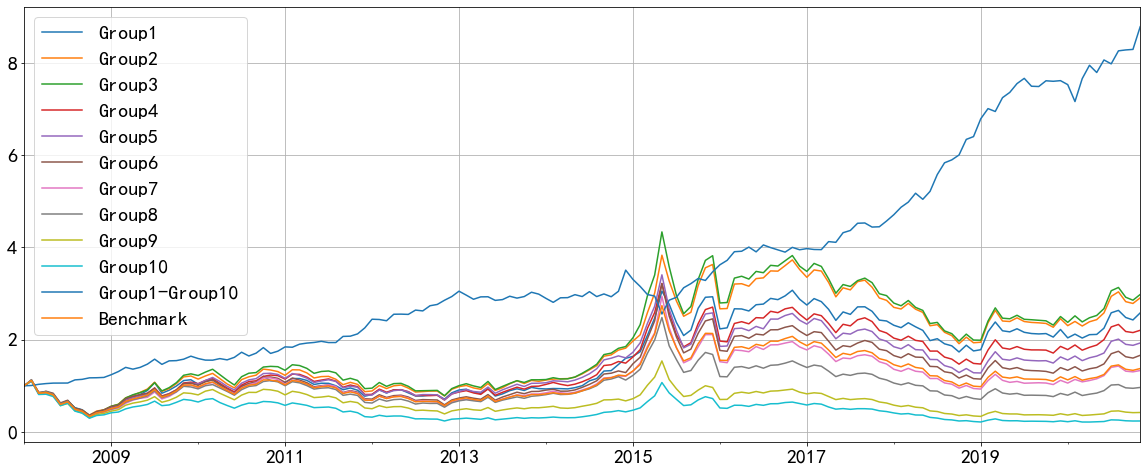
从上图来看，中心度因子 P 与规模因子有轻微的正相关关系，与 vol波动率因子呈负相关。

③单因子回测：回测区间是2008年1月到2020年12月底。

根据因子值将全部A股分成10组，月度调仓。基准指数为全A等权指数。

各组表现：

组间区分度：

****

因子组合回测结果如上。我们发现中心度因子 P 越大，其股票未来的收益越高，反之，股票的未来收益越低。中心度因子 P 从第 4 组到第 10 组的单调性表现较好，而第 1 组到第 3 组的单调性一般。表现最好的是第 3 组，年化超额全 A 等权指数 12.4%， 超额胜率为 69.0%，信息比率为 1.89；表现最差的是第 10 组，平均每年跑输全 A 等权指数8.8%，跑输概率为 76.2%，信息比率为-2.17。

从第 1 组-第 10 组的累计收益曲线来看，中心度因子 P 在 2015 年的表现和市场整体经历的大起大落有很高的相似性；同样，其2018年2月到2019年3月也和市场指数的下跌趋势有较强的相似性。这和中心度因子本身的计算方式有很大的关系，因为中心度因子可以被看作过去一段时间内个股股价对指数的影响力大小。虽然在上涨或震荡上涨行情中，追随中心度因子可以从指数中剥离出“有影响力”的个股，获得超额收益，它也同样会在大跌行情中挑选出有影响力的大跌个股。这幅图提醒我们在市场有下跌倾向时，不将单个的中心度因子纳入多因子库会是比较明智的选择。

第一组在全市场选股的分年表现：



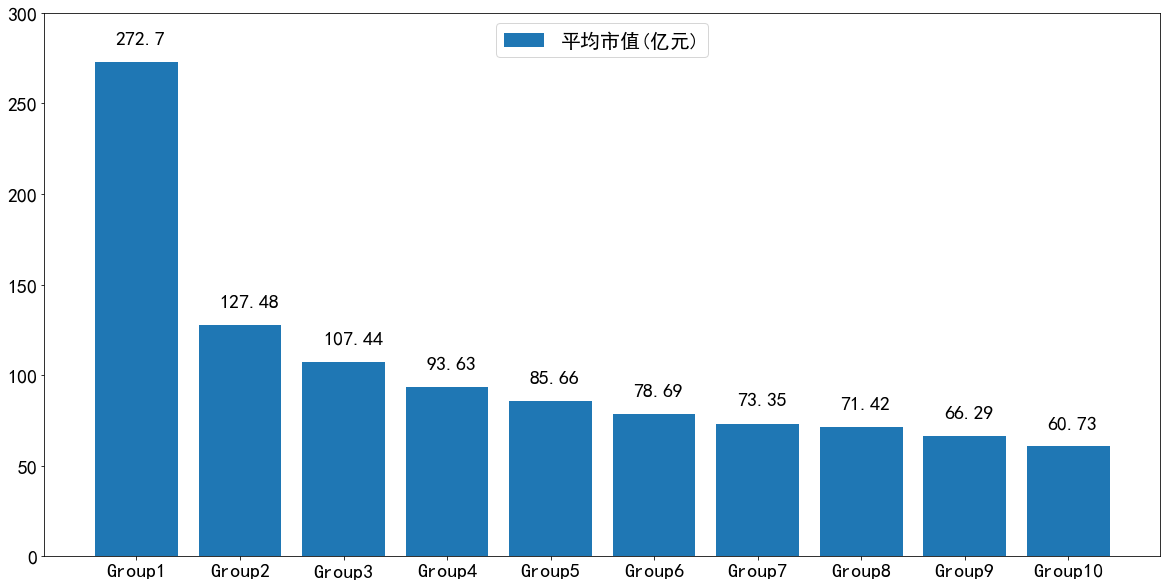
第一组分别在沪深300和中证500选股的分年表现：

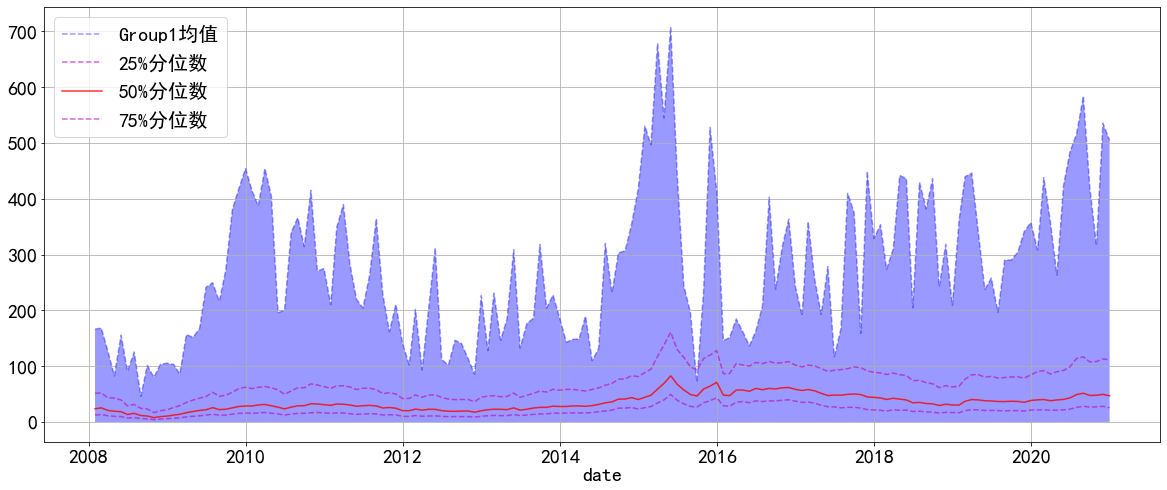


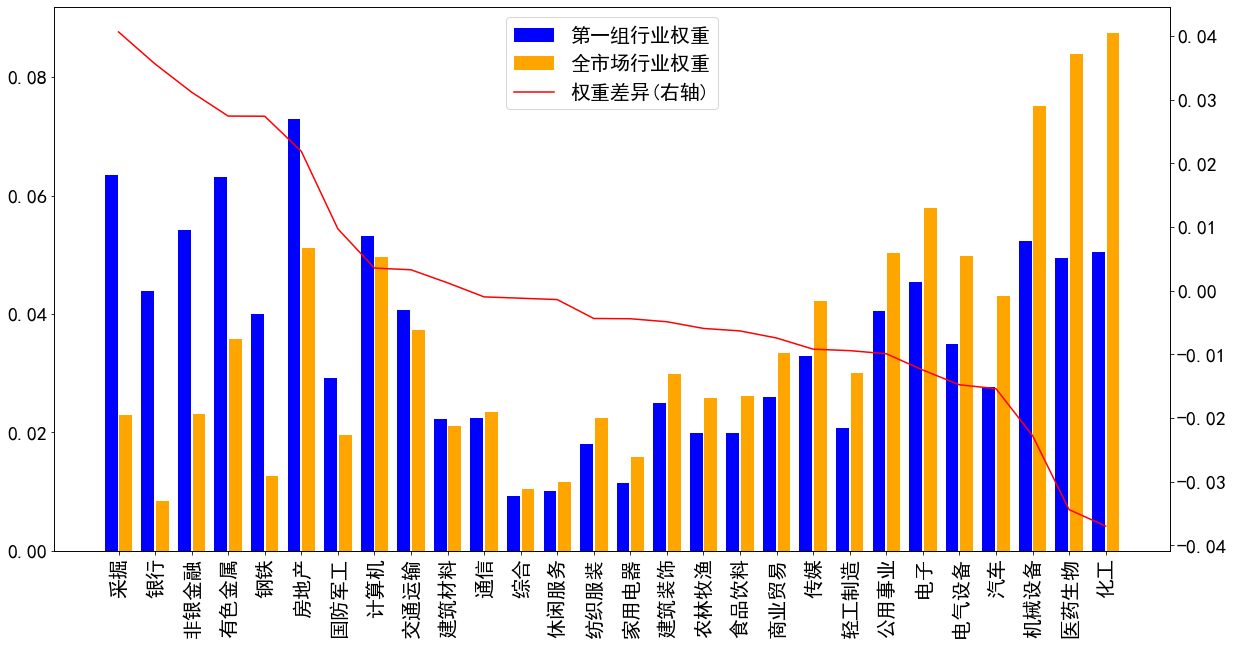
以上分别为中心度因子 P 第 1 组组合在不同股票池的分年表现情况。在全市场选股时，第 1 组年化收益为17.9%，最大回撤为 22%，相对于全 A 等权指数，年化超额收益为 4%，最大回撤为 4%，胜率为 53%，信息比率为 0.87，超额效果比较弱；在中证500中选股时，相对于中证 500 指数而言，年化超额收益为 6%，最大回撤为 5%，胜率为 55.4%，信息比率为 0.67，超额效果略胜全 A 等权；在沪深300中选股时，相对于沪深 300，年化超额收益为 9.8%，最大回撤为 8.5%， 相对胜率为 54%，信息比率为 0.69。

④ 因子P各组的市值、行业分布

例图：分组市值

****例图： 第一组持仓的平均市值时序图

****例图： 第一组持仓的行业权重图

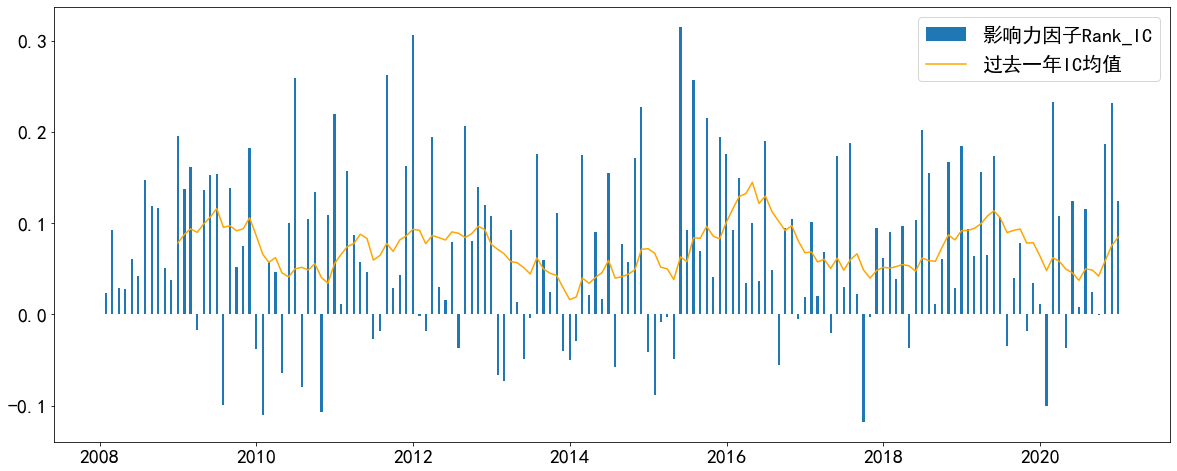
****

从中心度因子 P 的分组市值分布来看，中心度越大，市值越大。

从持仓市值的时间序列来看，第一组的平均市值远高于全市场的市值 75%分位数， 但是在行情暴跌或者长时间下跌的市场里面，第一组的平均市值与全市场的市值 75% 分位数的差异会比较小。

从持仓行业的权重图来看，中心度因子 P 第一组的持仓主要集中在采掘、银行、非银、有色、钢铁、房地产等市值与权重较大的行业，在基础化工、医药、机械和汽车等行业分布较少。

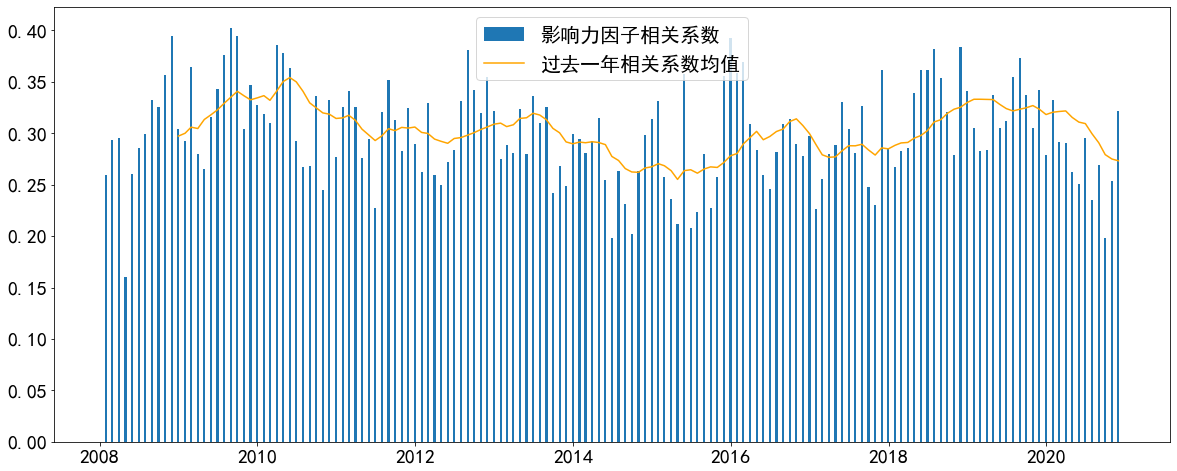
④ 选股能力检测：因子的rankIC



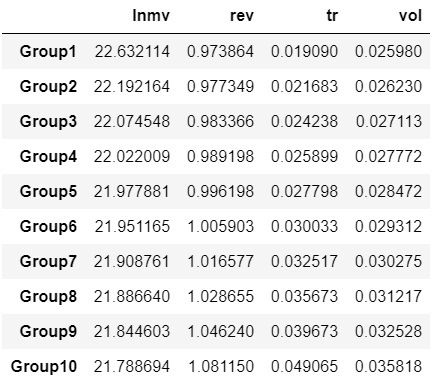
通过计算当期因子值与对应的股票的下期收益的 spearman 相关系数，我们得到了中心度因子 P 的 rankIC 大小。

长期来说，中心度因子 P 的 rankIC 的均值在 0.072 上下波动，属于因子里面中游的水平，IC\_IR 为 0.798。正显著比例为 69.2%，负显著比例为 30.6%。

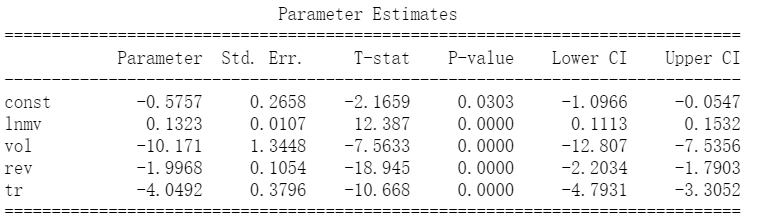
⑤ 稳定性检测：因子的自相关系数

****对于相邻两期的中心度因子 P 的值，我们计算着两期因子值的相关系数，来代表这个因子本身的稳定性。从上图来看，中心度因子 P 的自相关系数的长期均值在 0.30 上下波动，因此中心度因子是属于稳定性较强的因子。

⑥ 正交性检测：Fama MacBeth得出风格因子对因子P的解释能力

与风格因子的关系：

Fama MacBeth回归方程结果：

中心度因子 P 的第一组组合选出来的股票是市值大，反转小，换手率低，波动率较低的股票。

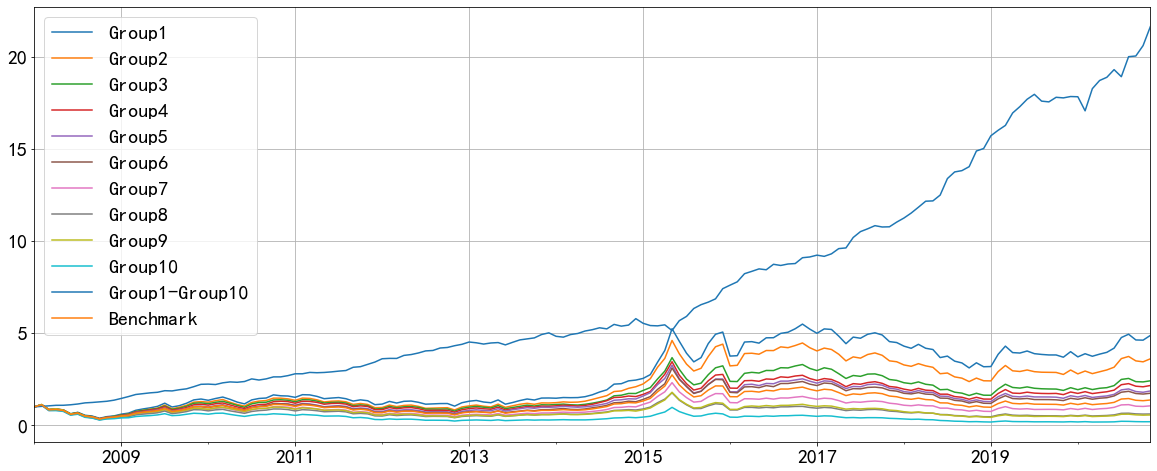
我们把中心度因子 P 与其他风格因子一起放进回归方程，以便观察风格因子对中心度因子 P 的解释解释能力。

从回归结果来看，由上表中的P-value一列，我们发现中心度因子与市值、波动率、反转、换手存在明显线性关系，相应的线性关系正负以及大小由表中的Parameter一列给出。可以看出，波动率因子对中心度因子的解释能力较强。

⑦ 行业市值中性化后的因子表现

例图：中性因子在沪深300的分组表现：

例图：中性因子在沪深300的区分度表现：

例图：中性因子第一组在沪深300的分年表现：



进行了市值和行业中性后，中心度因子 P 的单调性明显提高，从第一组到第十组表现出比较稳定的单调性，表现最好的是第一组，年化收益有 30%，平均每年超额沪深 300 指数 27%，超额胜率为 76%，信息比率为 2。表现最差为第十组，年化收益为-6%，平均每年跑赢沪深 300 指数 -9 %，信息比率为 -2.66。

从第 1 组-沪深 300 的累计超额曲线来看，我们发现虽然累计超额收益较高，但是净值曲线却表现出明显的凹凸不平，这表明经过粗糙的市值中性和行业中性虽然相对于沪深 300 指数有较高的收益，但是由于沪深 300 的市值和行业非常集中，所以中性后的中心度因子依旧可能导致第一组暗含着许多不经意的风险因子的暴露。

中性因子第 1 组的表现如下，平均年化收益为 24.7%， 平均绝对胜率为 58%，平均最大回撤为 22%，最大回撤发生在 2008 年和 2015 年 ; 相对于沪深 300 指数而言，平均年化超额收益为 16%，相对沪深 300 的胜率为 57%，信息比率为 0.9。

分年来看的话，中性因子第一组在2017年表现较差，超额收益只有-29%，信息比率为 -2.2，在去年表现也较为一般，超额收益为 0.3%（至 2020 年 12 月底）。

例图：中性因子第一组在中证500的分年表现：

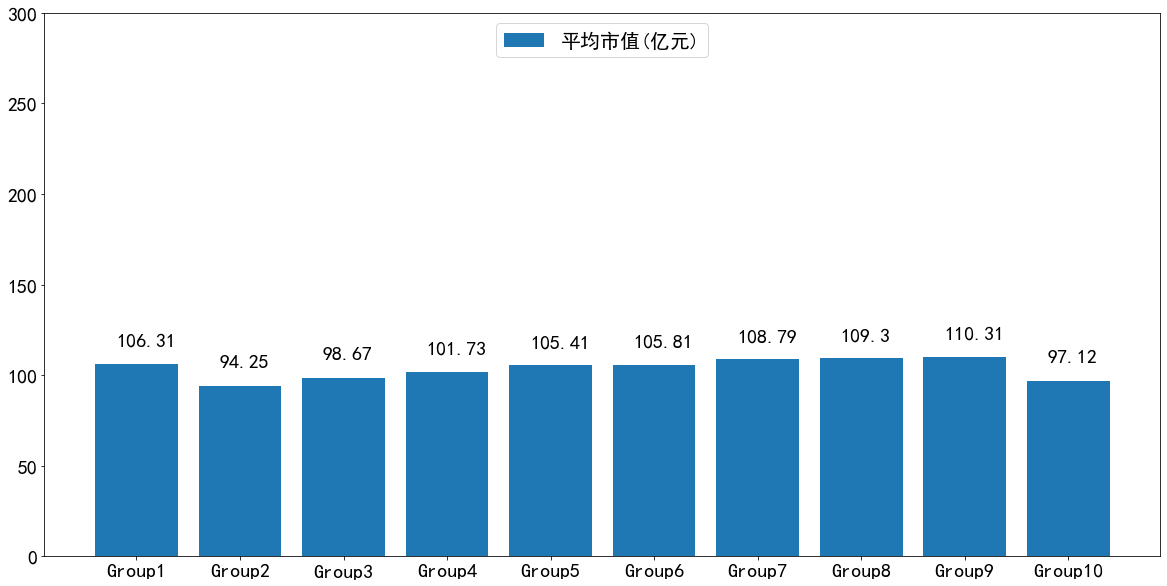


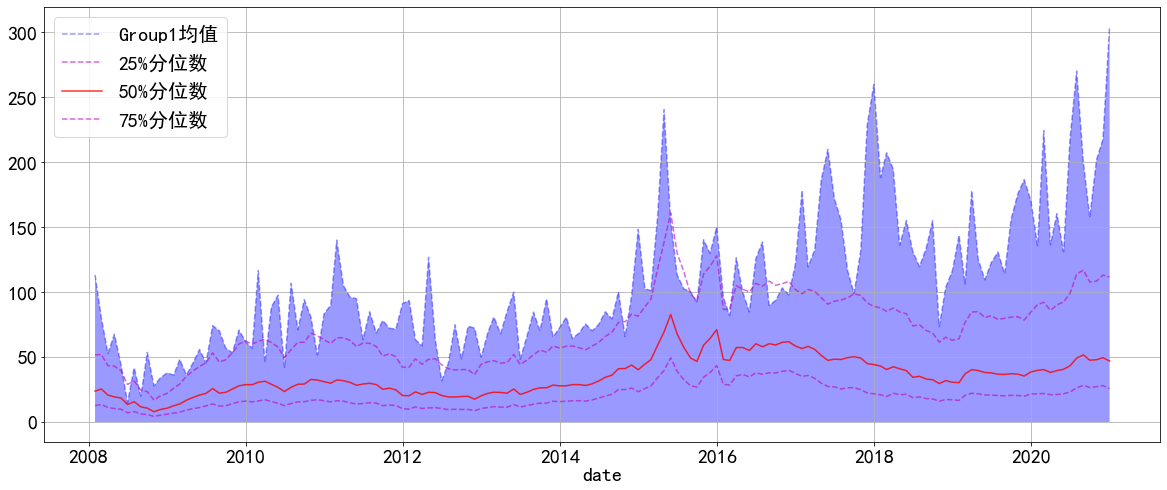
中性因子第 1 组组合的分年表现情况如下，相对于中证 500 指数而言，平均每年超额收益 11.6%，超额胜率为63.9%，信息比率为 1.48。

分年来看的话，中性因子第一组在 2017 年和 2019 年的表现稍弱，超额收益为福，表现最好的是 2009 年和 2016 年，信息比率高达 4.7 和 3.8。

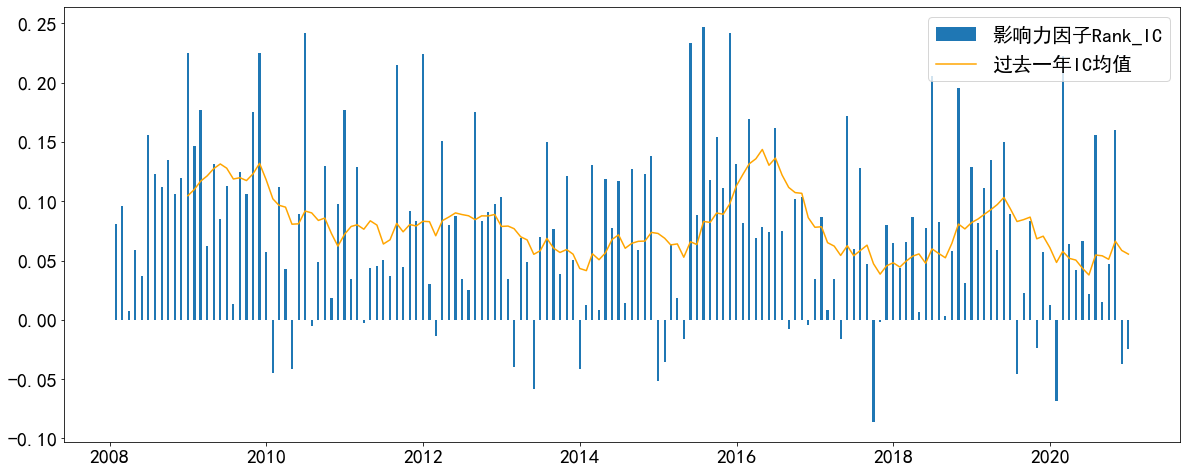
⑧ 中性因子的市值分布：

例图：中性因子分组市值

例图：中性因子第一组平均持仓市值的时序图



⑨ 中性因子的选股能力检测：rankIC



通过横截面对市值因子回归取残差的市值中性处理后，中心度因子 P 收益预测能力有了明显的提高。由上图，中性因子的 rankIC 的均值在 0.078 上下波动， IC\_IR 为 1.105，正显著比例为 72%，负显著比例为 20%。而经过市值中性处理后，中心度因子 P 与市值因子之间 rankIC 的相关性从0.377降低到了-0.006。

## 2.3 小结

本文以股票间的相关系数矩阵作为数据基础，以复杂网络理论刻画出股票市场的网络结构。在这个网络结构的基础上，我们仔细研究了股票在网络结构中的位置对股票的未来收益是否有显著影响。其主要结论如下：

1. 可把个股的作为参考性的买卖指标，高点对应股价底部，低点对应股价顶部。但其准确性在不同行情下有一定差异，从历史数据上来看，上涨或震荡行情中的准确性更强。

2、 中心度因子越大，股票未来的超额收益越明显；中心度因子具有显著的收益区分度。

3、 中心度因子有较明显的市值和行业偏好，一般来说，股票流通市值越大，中心度越高；同时采掘、银行、非银、有色、钢铁、房地产等市值与权重较大的行业的股票的中心度也一般较高；

4、 基于中心度因子有明显的市值和行业偏好，处理后的中心度因子的收益区分能力显著提高，第一组年化收益有 24.7%，平均每年超额中证 500 指数 11.6%，超额胜率为63.9%，信息比率为 1.48；中性因子的 rankIC 的均值在 0.078 上下波动，IC\_IR为 1.105。正显著比例为 71.8%，负显著比例为 19%。

## 2.4 数据差异分析对比

主要差异1：未经市值和行业中性化之前，和原研报相比，同以中证500为基准的情况下，年化超额和最大相对回撤和相对胜率和信息比率都降低，而rankIC和IC\_IR却不降反升。

主要差异2：经过市值和行业中性化以后，和原研报相比，同以中证500为基准的情况下，年化超额和最大相对回撤和相对胜率和信息比率都降低，而rankIC则无明显变化，IC\_IR不降反升，正显著比例和负显著比例均上升。

分析:

因为变小，在rankIC变大或者几乎不变的情况下，我们得出变大了。这说明中心度因子的选股能力在近年有所提高，但它选股能力的稳定性有所下降。

正负相关显著比例的同时升高也说明了上述分析结论，即选股能力提高但选股能力稳定性下降。

主要差异3：

通过对比热力图的颜色深浅，我们发现最显著的不同是中心度因子P和规模因子的相关性由正转为负。

主要差异4：

与原研报相比，中心度因子第一组和第二组市值的比值现已增大到2。

分析:

我们知道近两年规模因子的选股表现不佳，小市值偏好消失，甚至出现“规模因子已死”的说法，故衡量“影响力”的中心度因子与规模因子的相关性由正转负且第一组的小市值偏好愈发不显著。

# 第三章 研报评价

## 3.1 创新点

将复杂网络知识迁移到股票网络上，并扩充经典套利模型构建了中心度因子。体现了研报作者扎实的数学功底、经济学知识和对股票市场的经验。

## 3.2 不足点

在归纳作为买卖指标时，只是粗略地根据曲线就得出了：高点对应股价低点，低点对应股价高点的结论，没有注意到该规律在大跌行情下准确性出现大幅降低的情况。进一步地，就没有为后续因子择时做好准备工作。再者，作者没有从直观经济学角度或是模型构建的角度指出中心度因子在大跌行情下表现不佳的原因。