

方正证券研究所证券研究报告

“远山”量化选股系列（一）

金融工程研究

2018.03.06

分析师：韩振国

执业证书编号：S1220515040002

联系人：朱定豪

TEL：021-61375707

E-mail：zhudinghao@foundersc.com

相关研究

《抢跑者的脚步声：基于价量互动的选股因子》

请务必阅读最后特别声明与免责条款

投资要点

“不以规矩，不能成方圆”。没有系统的因子评价体系，难以筑起多因子研究的高楼。

本篇报告专注于单因子评价体系的介绍，挖掘能产生 α 的因子是量化选股的核心竞争力！主流的因子评价包括 IC 体系和分组测试两大块，两者共涉及到二十余项因子评价指标，共同构建了一套完整的体系。

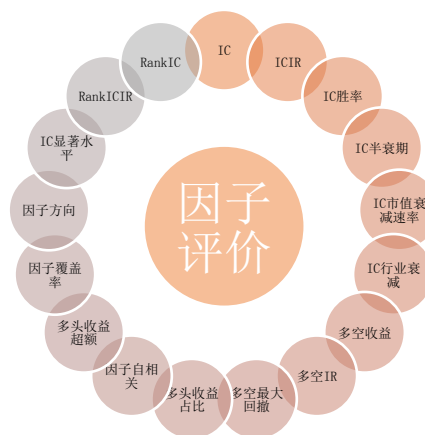
IC 评价体系的核心在于 IC，其基石是主动管理基本定律 $IR=IC*\sqrt{BR}$ 。IC 的衍生指标有 RankIC、IC 胜率等，IC 衰减可分为时间衰减、市值衰减和行业衰减，我们引入因子半衰期，市值衰减速率等指标来衡量。

分组测试是弱化版的回归，相较于 IC 体系，能更好展现指标收益非线性、交易费用对于收益的实际侵蚀、策略的日度波动特性、策略收益和回撤的具体大小等特性。常做的研究有因子多空组收益和回撤、多头和基准超额、多头收益占比、不同指数内测试、组合分层等。

因子相关性有两种衡量方式，既可以计算横截面相关性，也可以分组计算分位数的均值。**情景分析**是指有些因子上涨时表现更优，有些因子在下跌中更好。通过情景分析，能帮助我们更好地构建因子组合，对因子进行选择时。

风险提示

本报告基于历史数据进行评价，不构成任何投资建议。市场未来可能发生较大变化，本报告结论仅供参考。



目录

1	规矩：因子评价体系	4
2	量化选股概述	4
3	数据的获取	6
3.1	数据供应商	6
3.2	数据提供形式	6
3.3	数据类型	7
4	数据的处理	7
4.1	异常值修正	7
4.2	空缺值填充	9
4.3	分布调整	9
4.4	数据标准化	10
4.5	因子中性化处理	10
5	IC 评价体系	11
5.1	IC 的定义	11
5.2	主动管理基本定律	12
5.3	IC 衍生：RANKIC	12
5.4	IC 衍生：ICIR	13
5.5	IC 衍生：IC 胜率	14
5.6	IC 的时间衰减	15
5.7	IC 的市值衰减	16
5.8	IC 的行业衰减	17
6	因子评价之分组测试	18
6.1	数据的分组	19
6.2	组合的诞生	19
6.3	指标评价体系	20
6.4	多头、空头与多空组	21
6.5	指数内测试	23
6.6	基准的选取	24
6.7	组合分层测试	24
6.8	分组的本质	25
7	因子相关性检验	26
8	情景分析	27
9	风险提示	27

图表目录

图表 1:	因子评价体系	4
图表 2:	量化投资体系	5
图表 3:	单因子测试流程	6
图表 4:	数据类型	7
图表 5:	原始净利润增长率因子	8
图表 6:	3 倍标准差修正法	8
图表 7:	固定比率修正法	8
图表 8:	中位数极值法	8
图表 9:	原始流通市值因子	9
图表 10:	调整后的市值（对数流通市值）	9
图表 11:	非流动性因子原始分布	10
图表 12:	非流动性因子调整后分布	10
图表 13:	中性化前后非流动因子的市值暴露	11
图表 14:	总市值因子的 IC 序列	12
图表 15:	因子 IC 和 RANKIC 统计	13
图表 16:	FR 因子的 IC 序列	14
图表 17:	总市值因子 IC 序列的变化	14
图表 18:	价格时滞因子 IC 时间衰减	15
图表 19:	价格时滞因子 IC 市值衰减	16
图表 20:	价格时滞因子 IC 市值衰减回归	17
图表 21:	换手率因子 IC 行业衰减	17
图表 22:	IC 体系的局限性	18
图表 23:	换手率因子全市场选股的持股数量	19
图表 24:	换手率因子的多头换手率	20
图表 25:	分组测试中的指标体系	20
图表 26:	市值调整换手率因子的统计指标	21
图表 27:	总市值因子多空组净值和回撤	21
图表 28:	反转因子各组收益	22
图表 29:	行业中性反转因子各组收益	22
图表 30:	多头收益占比	22
图表 31:	净利润同比增长率因子（全市场）	23
图表 32:	净利润同比增长率因子（中证 800）	23
图表 33:	净利润同比增长率因子（中证 500）	23
图表 34:	净利润同比增长率因子（沪深 300）	23
图表 35:	CAPM 残差波动率因子累计收益	24
图表 36:	CAPM 残差波动率因子年化收益	24
图表 37:	因子分层示意图	25
图表 38:	IC 检验与分组测试的关系	25
图表 39:	因子相关性矩阵	26
图表 40:	换手率因子在市值上的暴露	26
图表 41:	净利润增长率因子情景分析	27
图表 42:	CAPM 残差波动率因子情景分析	27

“不以规矩，不能成方圆”——《孟子》

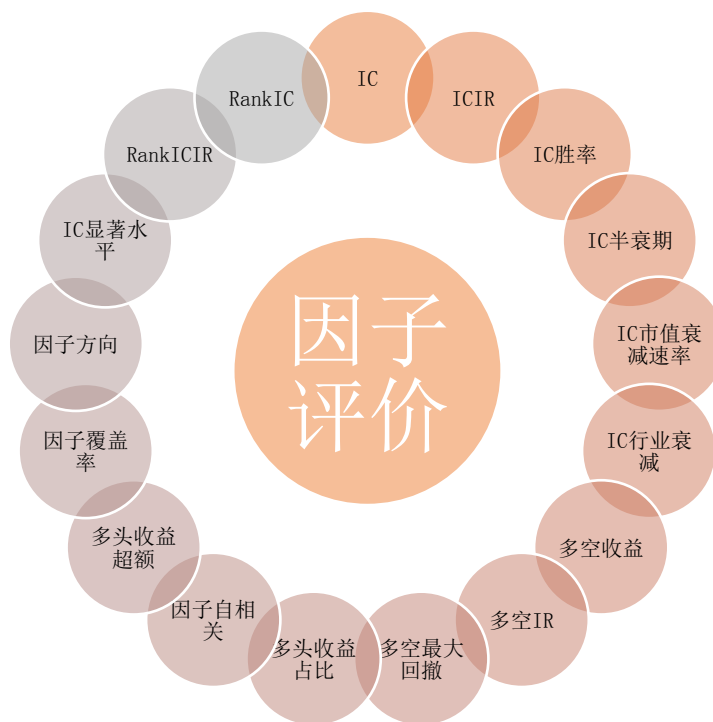
1 规矩：因子评价体系

挖掘 Alpha 因子、评价 Alpha 因子、改进 Alpha 因子是量化投资者一生为之奋斗的目标，而一套严密的因子评价体系是这一切的基石。不以规矩，不能成方圆；没有系统的因子评价体系，难以筑起多因子研究的高楼。

主流的因子评价包括 IC 体系和分组测试两大块，两者共涉及到二十余项因子评价指标，共同构建了一套完整的体系。我们可以用它来解答许多实际投资中的疑问：为什么高 IC 的因子收益不一定高？为什么因子在部分指数上效果欠佳？为什么跑赢大盘并不能代表因子具备 Alpha？

本篇报告带来方正多因子评价体系，市场最全面的因子评价系统。

图表1： 因子评价体系



资料来源：方正证券研究所

2 量化选股概述

量化投资体系可以被有效拆分为量化择时和量化选股，前者专注于在纵轴时间维度上进行判断，后者聚焦于横截面上比较不同股票的优劣。无数的理论研究和实践都指出，择时是比较困难的而且由于维度较少结果不稳定，选股带来的收益更为稳健，5.2 节的主动管理基本定律深刻揭示了这一点。

本篇报告专注于量化选股，将择时的问题放到后续报告再做讨论。换言之在纯选股的体系中，我们始终假设满仓位运行，我们致力于研究如何从在横截面选股上获取 alpha。

图表2： 量化投资体系



资料来源：方正证券研究所

量化选股可以大致分为以下几个步骤：

1、数据获取与清洗：在早期没有数据供应商介入前，该步骤往往要耗费投资者大量的精力。而随着国内数据供应商越来越多，获取数据的难度下降，数据质量提升，让我们有更多的精力专注于后续研究。

2、单因子评价：构建科学有效成体系的因子评价体系，能帮助我们全面了解一个因子的特性。量化投资者每天都在质问自己：我们所暴露的是 alpha 还是 beta，我们从中承担了什么样的风险？挖掘未知的 alpha 因子，带来超额收益正是量化投资最具魅力的地方。

3、组合优化：从单因子到投资组合，要历经多因子结合，换手率、风格、回撤约束等等限制条件。不同的产品会有不同的约束条件，例如部分资金可能对回撤要求极其严格，而可以适当降低预期收益。

4、交易执行：交易执行中会遇到滑点、价格冲击等问题，有效地减少交易带来的价格冲击，控制换手降低交易成本，确保订单的完成是这一块研究的核心。

5、绩效归因：组合的收益归因常用的是 Barra 模型，了解自身组合的收益来源和风险暴露，有助于我们对模型作进一步的完善。相关报告详见《“星火”多因子系列（一）：Barra 模型初探：A 股市场风格解析》和《“星火”多因子系列（二）：Barra 模型进阶：多因子风险预测》。

本篇报告专注于第二步单因子评价，挖掘能产生 alpha 的因子是量化选股的核心竞争力！一般认为因子评价有两套系统，分别为 IC 评价系统和分组测试评价系统，IC 评价体系简洁高效，分组测试繁琐但能展现更多细节，两者混合使用效率更高。我们在 6.8 节会阐述两者本质其实是一致的，分组是横截面回归的一种弱化形式。

单因子的测试步骤分为：

数据准备→交易条件限制→因子的异常处理与标准化→因子中性化

→因子分组→构建组合→策略评价→因子暴露分析

→因子 IC→IC 时间衰减/IC 市值衰减/IC 行业衰减/IC 情景分析

我们将分章节介绍上述步骤

图表3： 单因子测试流程



资料来源：方正证券研究所

3 数据的获取

构建多因子模型的第一步在于数据的获取与存储，万事开头难，如果没有靠谱的数据源，数据的抓取与清洗往往耗费我们大量的精力。在理想状态下，数据供应商能为我们提供及时、可靠、整洁的数据，使量化投资人不用再耗费精力在数据清洗上。实际投资中，我们也期望尽可能少的把时间用于下载和清洗数据，将更多的精力用于后期的因子挖掘与模型优化。

3.1 数据供应商

目前市场上金融终端普遍采用东方财富 Choice 以及 Wind 资讯金融终端等，而主流的量化数据库供应商有 Wind 资讯、恒生聚源、朝阳永续、天软、巨潮、天相、巨灵、国泰安、通联等，数据各有特色。

比如像 Wind 资讯数据全，在机构中使用最多，也是我们主要的数据来源。朝阳永续在分析师一致预期、盈利预测数据以及私募基金数据方面占据优势。天软科技在高频行情数据上比较突出。国泰安数据库主要为学术用途其中公司金融数据较为完善。

3.2 数据提供形式

目前主要的数据提供方式有三种：

一是采用 API 接口，如 Wind、Choice 等均有提供。该类方法往往会限制数据下载量，因此只适合小型数据的下载，批量下载面板数据往往会导致数据超限。部分供应商提供额外数据流量包，购买后可突破流量限制。

二是落地数据库，如 Wind，聚源等。购买落地数据库后，数据供应商每天会推送所有数据至本地数据库。这意味着每次访问的数据均在本地，自由度较高，不受外部限制，是量化机构投资者主要使用的方式。

三是基于第三方平台，如优矿、米矿等平台。近年来随着量化投资走入公众视野，第三方量化研究平台崭露头角。该类平台集成第三方授权数据、回测研究模块，对新人而言更容易上手，对私募而言成本更低，缺点就是模块化程序限制了使用者一定的自由度。

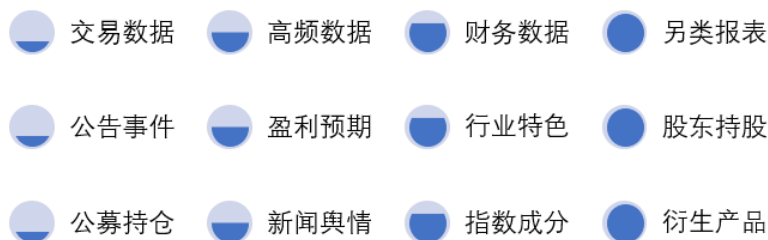
方正金工因子回测平台主要数据来源是 Wind 落地数据库，辅以及其他数据供应商的一些特色数据。

3.3 数据类型

常见的选股数据类型一般有以下 12 种，分别为交易数据、高频数据、财务数据、另类报表、公告事件、盈利预期、行业特色、股东持股、公募持仓、新闻舆情、指数成分和衍生产品。

方正金工的高频系列选股就是在分钟数据上挖掘出新的 Alpha，基金停牌套利是利用个股停牌和基金持股的数据做出的事件类套利策略，在此不再赘述。

图表4： 数据类型



资料来源：方正证券研究所

4 数据的处理

4.1 异常值修正

在处理因子的过程中，往往会遇到一些异常值，如净利润增长率因子可能会出现极端大值（上一期净利润接近于 0），在进行后续的回测前，我们要对该类数据做预处理。如果不进行异常值的修正，异常值会干扰因子中性化（回归）的结果，对 IC 等指标的计算也会带来误差。

异常值修正的方法大致有以下几种：

1、均值标准差修正法：将偏离均值 3 倍标准差的数据拉回

3 倍标准差修正法也称为 3σ 法，基于因子的平均值，标准差 σ 可以体现因子的离散程度，所以可以通过 3σ 法来衡量因子与平均值的距离从而修正极端异常值。

$$\tilde{x}_i = \begin{cases} \mu + 3 * \sigma, & \text{if } x_i > \mu + 3 * \sigma \\ \mu - 3 * \sigma, & \text{if } x_i < \mu - 3 * \sigma \\ x_i, & \mu + 3 * \sigma \leq x_i \leq \mu - 3 * \sigma \end{cases}$$

2、固定比率修正法：将最大的 2% 或最小的 2% 数据拉回

固定比率修正法也称为百分位法，方法是将因子值进行升序排序，对排位高于 98% 或排位低于 2% 的异常因子值进行调整。

$$\tilde{x}_i = \begin{cases} p_{98}, & \text{if } x_i > p_{98} \\ p_2, & \text{if } x_i < p_2 \\ x_i, & p_2 < x_i < p_{98} \end{cases}$$

3、中位数极值法：将偏离中位数 5 倍去均值中位数的数据拉回

绝对中位差 (MAD) 较标准差而言对离群值 (outlier) 也就是异常值更加的稳健。在标准差的计算中，数据点到其均值的距离要求平方，因此对偏离较为严重的点偏离的影响得以加重，也就是说异常值严重影响着标准差的求解，因此使用中位数极值法能更好的度量修正异常值。

$$md = \text{median}(\{x_i, i = 1, 2 \dots n\})$$

$$MAD = \text{median}(\{|x_i - md|, i = 1, 2 \dots n\})$$

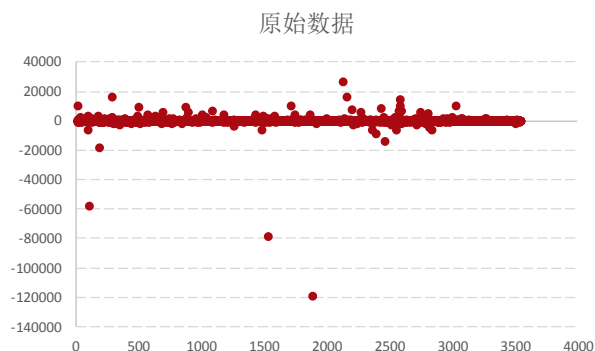
$$\tilde{x}_i$$

$$= \begin{cases} md + n * MAD, & \text{if } x_i > md + n * MAD \\ md - n * MAD, & \text{if } x_i < md - n * MAD \\ x_i, & md - n * MAD \leq x_i \leq md + n * MAD \end{cases}$$

我们以 2017/12/29 净利润增长率因子为例，原始的净利润增长率因子充斥着大量异常值，最大的 26962，最小的 -118733，如不进行修正严重干扰回归的结果。

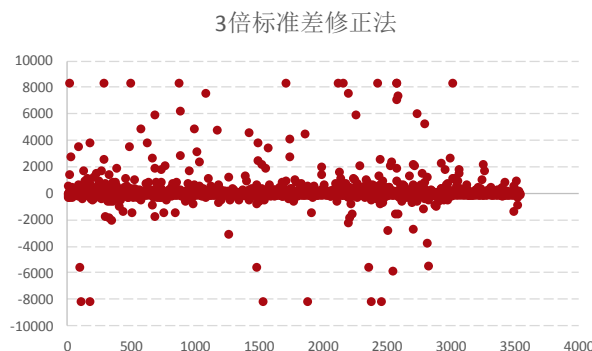
我们分别展示了 3 倍标准差修正、2% 固定比率拉回和中位数极值法。结果表明中位数极值法效果更好，成长类因子稳定性差，异常值波动大，中位数极值法对异常值处理效果更佳。

图表5： 原始净利润增长率因子



资料来源：Wind 资讯，方正证券研究所

图表6： 3 倍标准差修正法

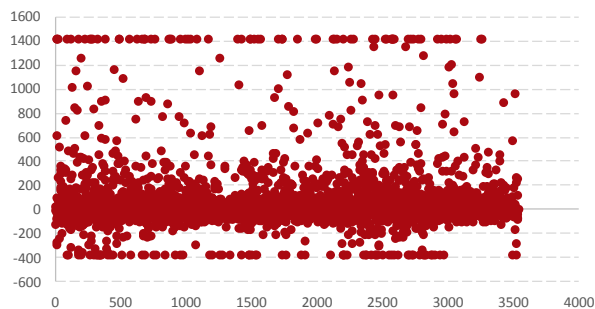


资料来源：Wind 资讯，方正证券研究所

图表7： 固定比率修正法

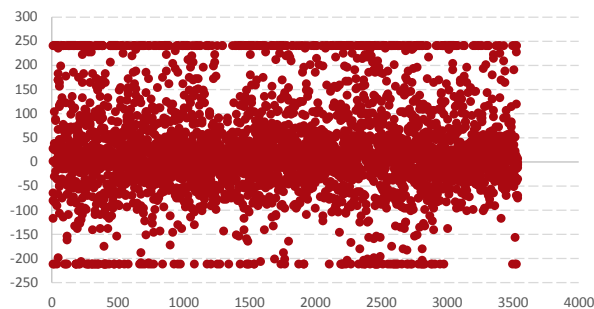
图表8： 中位数极值法

固定比率修正法



资料来源: Wind 资讯, 方正证券研究所

中位数极值法



资料来源: Wind 资讯, 方正证券研究所

4.2 空缺值填充

另一类常见的问题是空缺值, 一般而言, 我们期望因子的空缺值越少越好, 用因子覆盖率来衡量空缺值的数量, 则因子覆盖率越大越好。

$$\text{因子覆盖率} = \frac{\text{因子空缺数量}}{\text{因子总数量}}$$

标准化的价量交易型数据空缺值较少, 因子覆盖率基本都在 95% 以上, 但部分基本面因子、分析师一致预期、大数据文本挖掘等另类因子因子覆盖率可能较低, 在一定场合下需要填充, 比方用于后续回归模型的准备。

空缺值的处理的方法一般为中位数替代法、可比公司替代法、历史数据替代法等, 在部分情境下直接剔除该股票也是一个不错的选择, 一般根据实际需求进行选取相应的方法。

- 1、中位数替代法: 以市场中位数或者行业中位数替代
- 2、可比公司替代法: 寻找主营业务相近的可比公司指标替代
- 3、历史数据插值法: 以历史数据填充, 或者根据趋势进行插值

4.3 分布调整

理想状态的因子应呈现出近似正态分布, 如果有严重偏离, 我们常采用对数、开根号等方式对数原始数据的分布进行调整。

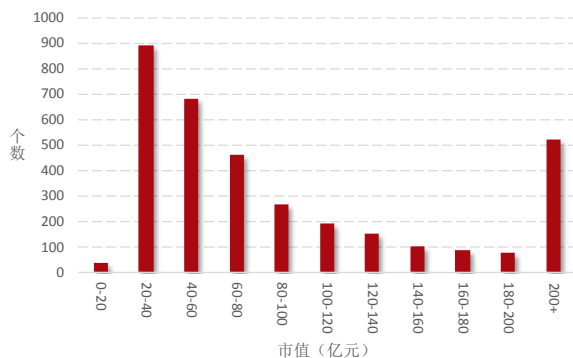
一个常见案例就是市值, 原始的市值因子呈现出明显的右偏、尖峰、后尾的分布特点, 其意义是 A 股小票众多, 但少数股票市值极大, 尤其是金融和两桶油。

以 2017/12/29 数据为例, 全市场近 3500 支股票, 上证 50 的 50 支成分股市值占到了全市场的 32.0%, 沪深 300 的 300 支成分股市值占到了全市场的 58.1%。

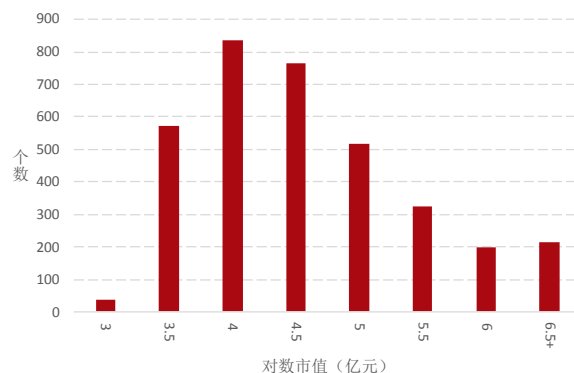
下图展示了原始的流通市值和对数处理后的流通市值分布, 后者更接近于正态分布。经典金融学模型假设股价服从几何布朗运动, 即收益率服从正态分布, 股价服从对数正态分布。因此一般市值因子都会取对数处理, 修正后尾分布。

图表9: 原始流通市值因子

图表10: 调整后的市值(对数流通市值)



资料来源: Wind 资讯, 方正证券研究所



资料来源: Wind 资讯, 方正证券研究所

4.4 数据标准化

由于各个因子的量纲不一致, 为方便后续因子加工和回归需要对数据作标准化处理。标准化处理的方式主要有三种:

- 1、对因子原始数据做 ZScore 标准化
- 2、对因子原始数据进行加权标准化
- 3、将因子原始数据转换为排序值再归一化

第一种方式是最常用的方法, 好处在于能够更多保留因子之间原始的分布关系, 但是进行回归的时候会受到极端值的影响; 第二种方法是 Barra 收益-风险模型中使用的方法, 通过市值加权, 能更好的体现出股票对于基准股票池的风险暴露; 第三种方式的好处在于排序值的稳健性, 容易看出因子和收益率之间的相关性的方向。

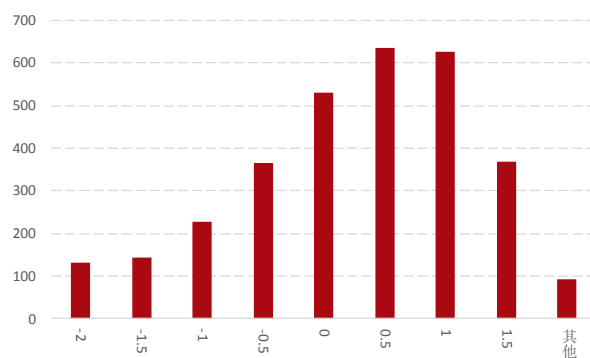
一般而言, 去极值和标准化两步会一起使用, 先采取中位数极值法, 后用 ZScore 做标准化处理。

图表11: 非流动性因子原始分布



资料来源: Wind 资讯, 方正证券研究所

图表12: 非流动性因子调整后分布



资料来源: Wind 资讯, 方正证券研究所

4.5 因子中性化处理

我们希望一个因子具有选股能力, 更希望一个因子具有额外的选股能力。如果其选股能力完全被已知因子所解释, 那么他给我们带来的增量信息就比较有限。如果在做完因子中性化, 新的因子仍具有选股能力, 这才是我们真正期待的结果。

常用的因子中性化有两种方法: 一是线性回归 (也叫作正交化), 二是分层测试。本节介绍回归法, 我们将分层放到 6.7 节再做介绍。

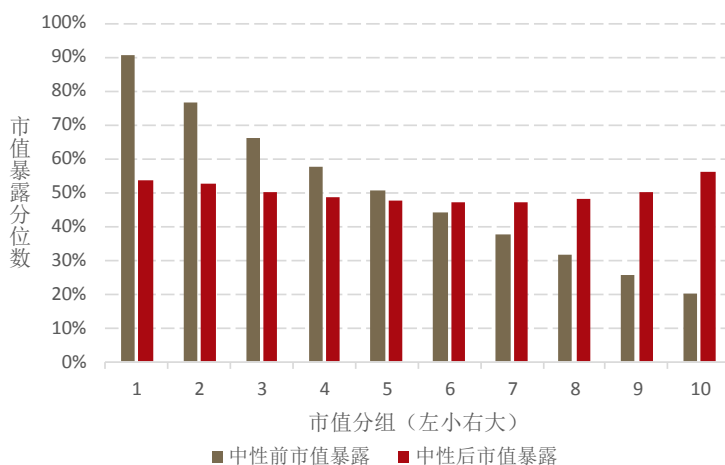
前者对数据的处理更为漂亮，可以很好地处理线性相关带来的影响，但对于非线性相关的关系可能效果稍差。后者更为稳健，但带来的问题是中性化处理相对粗糙，丢失一定的信息。

因子正交化的方法是横截面多元回归，应变量是待测因子 F ，自变量是行业 and 各类待正交化的因子，回归得到的残差就是中性后的待测因子。以剔除常见因子市值、动量、换手率、波动和行业为例，其中行业为 0-1 哑变量，其余为因子值，回归得到的残差 ε 就是纯净的 F 因子。

$$F = \beta_1 \text{LogMV} + \beta_2 \text{Mom} + \beta_3 \text{Turn} + \beta_4 \text{Vol} + \sum_{i=1}^n \beta_{5i} * \text{CiticInd}_i + \varepsilon$$

以非流动因子对市值中性为例，中性前非流动性因子与市值高度相关，相关性达到 -0.7679，而中性后相关性大幅下降至 -0.008。从各组分位数来看，中性前多头在小市值上有暴露，中性后大幅改善。中性后的非流动性因子出现了轻微的非线性情况，两端市值偏小而中间市值略大，但这种偏差基本可以忽略。计算因子暴露方法详见第 7 章。

图表 13: 中性化前后非流动因子的市值暴露



资料来源: Wind 资讯, 方正证券研究所

5 IC 评价体系

只用一个指标来衡量股票的价值，PE（市盈率）可能是首选；对于债券，YTM（到期收益率）可能是首选；对于期权，IV（隐含波动率）可能是首选；如果要衡量因子的选股能力，IC（信息系数）可能是最有效的指标。IC 评价体系为我们提供了一把公允的尺子，使我们用之丈量不同因子的预测能力。

5.1 IC 的定义

顾名思义，IC 评价体系的核心在于月度 IC（信息系数），全称为 Information Coefficient，其计算方法为本期的因子值 $Factor_t$ 与下一期收益率 Ret_{t+1} 的截面 Pearson 相关系数，该指标旨在衡量因子 对于下一期收益的预测能力。因子预测能力越强，则 IC 的绝对值越大。

$$IC_t = \text{cor}(Factor_t, Ret_{t+1})$$

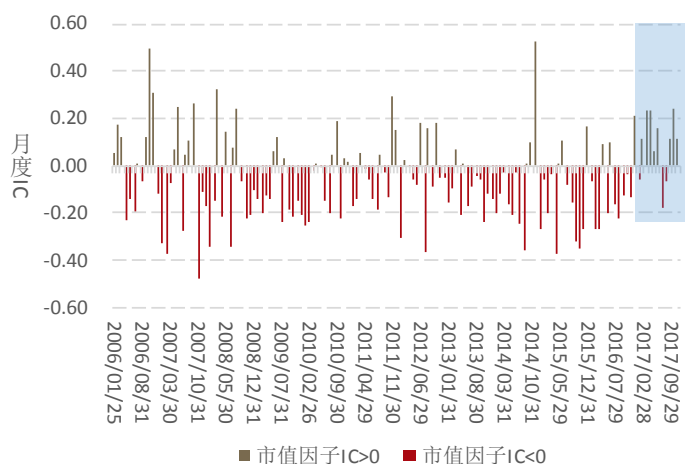
IC 可以和一元线性回归中的 R^2 联系在一起, 因为相关系数的平方等于带截距项的一元线性回归的 R^2 , 这两个变量的含义是因子 Factor 可以多大程度解释股票未来的涨跌幅。我们也可以用 t 值来衡量被检验因子的预测能力, 其效力与 IC 近似, 相关检验指标构造方法也相同, 本文不再做赘述。

$$Factor_t = \alpha + \beta * Ret_{t+1} + \varepsilon$$

当 IC 值为正, 表明因子值和未来收益正相关, 例如净利润增长率, 值越大股票未来收益越高。当 IC 值为负, 表明因子值与未来收益负相关, 例如 PE 因子, 值越小未来收益越大。由此可见 IC 的方向不重要, 重要的是 IC 的绝对值。

以总市值因子为例, IC 序列整体向下, IC 显著比率很高 (IC 绝对值 > 0.05), 表明市值因子对市场有非常强的影响能力。IC 为负表示小市值股票表现较好, 而 IC 为正表明大市值表现占优, 17 年以来因子的方向出现了明显的变化。

图表14: 总市值因子的 IC 序列



资料来源: Wind 资讯, 方正证券研究所

5.2 主动管理基本定律

IC 体系与实际投资效果有什么样的关系? 主动投资管理的基本定律将实际投资的效果与投资能力联系在了一起, 下述公式是 IC 因子评价体系的基石。

$$IR \approx IC * \sqrt{BR}$$

IR 是投资组合的收益风险比, IR 越高代表着实际投资中的低风险、高收益; IC 是因子的预测能力, IC 越高说明因子对于未来的预测能力越强; BR 是应用的广度, 如果我们在全市场选股, 应用的广度就是 3000 多, 而对于一次择时, 广度就是 1。在频率一致的情况下, 我择时能力须达到全市场选股的 $\sqrt{3000} = 54$ 倍, 才能有相同的 IR。

5.3 IC 衍生: RankIC

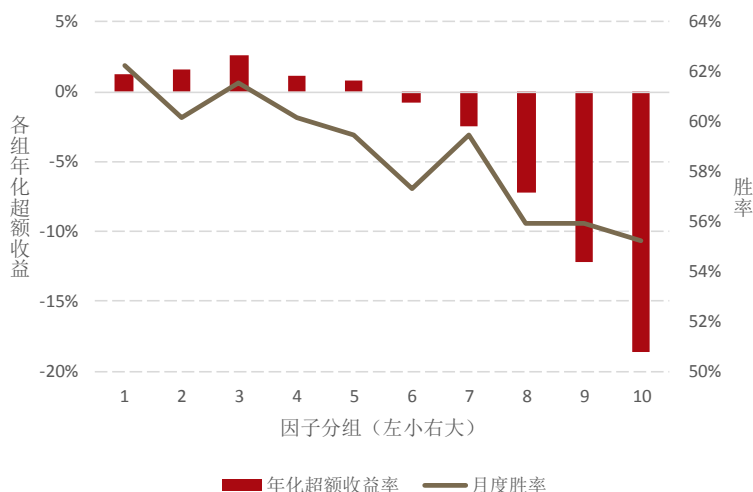
IC 的一个变形是 RankIC，即 Spearman 秩相关系数，其经济学意义与 IC 近似，用来衡量因子对于未来收益的预测能力。秩相关系数是将两列数的样本值按数据的大小顺序排列位次，以各要素样本值的位次代替实际数据而求得的一种统计量。

Pearson 相关系数要求数据是服从正态分布或者近似正态分布，如果不符合，一种可能就是采用 Spearman 秩相关系数来代替 Pearson 线性相关系数。

与 Pearson 相关系数相比，Spearman 秩相关系数可以排除非线性的因素，但也会带来一定的信息丢失。在实践中秩相关系数均值往往更大一些，显著比例也要更高一些。

比较极端的例子是一个月涨跌幅因子（短期反转因子），IC 为 -4.85%，RankIC 为 -6.42%。将因子值分十组后，以十组年化收益的中位数为基准绘制各组超额年化收益，因子空头上收益明显，体现出较强的非线性特点，因子 RankIC 会大幅超出 IC，我们在第 6.4 节因子的多空组中再次讨论这个问题。

图表15： 因子 IC 和 RankIC 统计



资料来源：Wind 资讯，方正证券研究所

5.4 IC 衍生：ICIR

ICIR 是 IC 的衍生指标，在 IC 评价体系中占据着重要地位，该指标用来衡量因子兼顾稳定性下的预测能力。ICIR 是 IC 指标的 IR，可以类比股票的 IR，用来衡量单位波动的收益情况。一般情况下，我们会比较因子的年化 ICIR。理想的选股因子 IC 均值大，预测能力强；另一方面 IC 波动率低，因子收益稳定波动较小，ICIR 指标将这两者做了有机的结合。

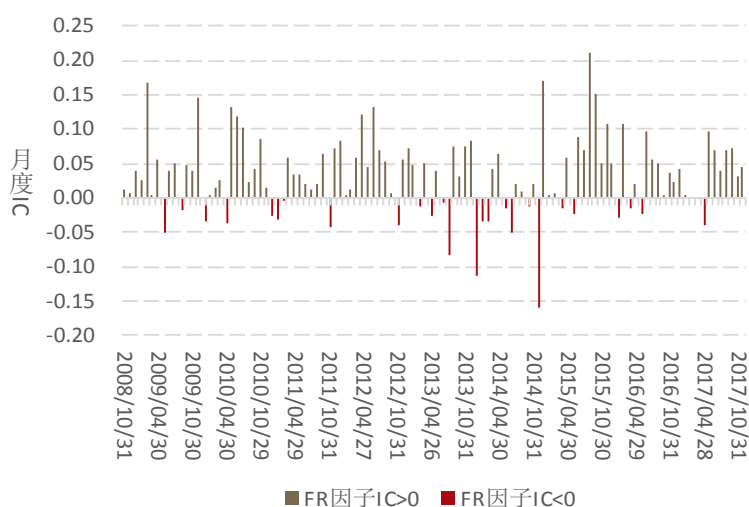
$$ICIR = \frac{\overline{IC}}{std(IC)} * \sqrt{12}$$

值得注意的是在比较某个因子在全市场和局部例如沪深 300 内选股时，我们应该利用的指标是 IC 而非 ICIR，因为 IC 的波动率会受到预测广度的影响，广度越大则 ICIR 越大。

ICIR 我们在选股报告《抢跑者的脚步声：基于价量互动的选股因子》中提出的 FR 因子，就是一个 IC 值不高但 ICIR 较高的因子。IC 达到 3%，ICIR 达到 2.57。反观市值因子 IC 值达到 5.78%，ICIR 只有 1.10。

$$FR = \text{cor}(\text{turn}_{t-1}, \text{ret}_t)$$

图表16: FR 因子的 IC 序列



资料来源: Wind 资讯, 方正证券研究所

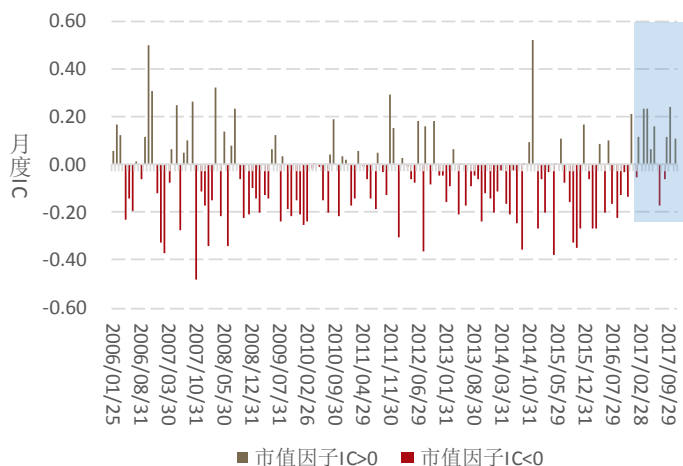
5.5 IC 衍生: IC 胜率

IC 胜率也是 IC 的一个衍生指标，算法 IC 为 IC 战胜基准的比率（比方 0），用来衡量因子预测能力的稳定性。类似的指标有 IC 显著比率，即 IC 大于（或小于）某显著水平，如 0.05 的比率。

一个 IC 绝对值的平均值大，IC 胜率越高则说明该指标的风格持续时间长，更偏向 alpha 因子，反之则更偏向 beta 因子。理想的 alpha 因子方向应该较少发生切换，拥有持续不断的同向高 IC。

仍以总市值因子为例，2006~2016 年期间，市值因子表现卓越，尤其是 13~16 年月度 IC 值几乎全部同向，IC 胜率极高。17 年来，市值类因子 IC 值衰减甚至反向，17 年 2 月~6 月、9 月~11 月 IC 方向均相反，逐渐从 alpha 因子向 beta 因子转变。

图表17: 总市值因子 IC 序列的变化



资料来源: Wind 资讯, 方正证券研究所

5.6 IC 的时间衰减

IC 衰减是指在时间维度和横截面股票维度上, 指标预测能力的流逝。IC 的时间衰减是被提及最多的一个概念, 用以衡量一个指标对未来的预测能力能持续多久。当一个指标时间衰减过快时, 可能会导致组合较高的换手, 交易成本会大幅侵蚀模型的盈利能力。通过计算当期因子值和滞后 N 期的收益率, 我们可以得到 IC 的时间序列。

$$IC_n = \text{cor}(Factor_t, Ret_{t+1+n})$$

可以简单的用 IC 半衰期来衡量 IC 时间衰减的快慢, IC 半衰期的定义为月度 IC 下降到一半所用的时间, 求解方法可以利用插值法, 如线性插值, 三次样条插值等, 以最简单的线性插值为例。

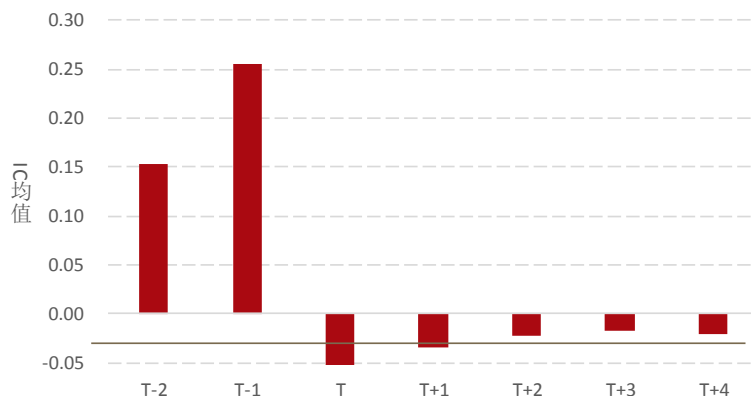
$$\rho(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x_1 - x_0)$$

$$\rho(0.5x_0) = 1.07$$

基本面因子 IC 衰减较慢, 半衰期在 6 个月以上, 我们将所有半衰期大于 6 个月的均记为 6。交易类指标如市值调整换手率因子半衰期短衰减快, 半衰期约为 1.07 个月。

图表18: 市值调整换手率因子 IC 时间衰减

市值调整换手率因子IC时间衰减

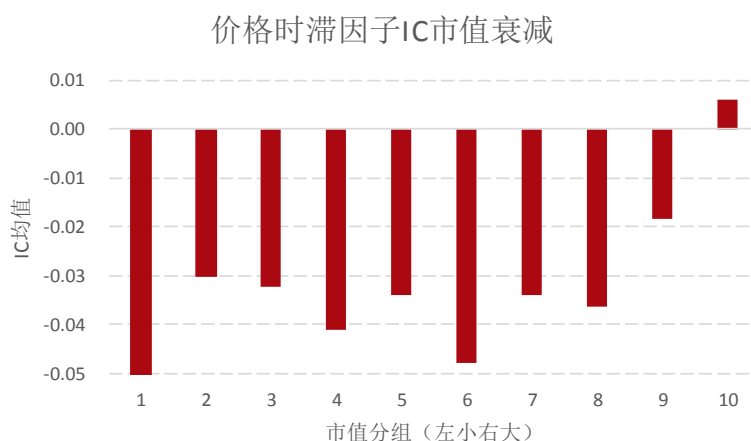


资料来源: Wind 资讯, 方正证券研究所

5.7 IC 的市值衰减

IC 的市值衰减在以往的研究被提及较少,但随着 17 年以来市值因子出现大幅回撤, IC 的市值衰减变得越来越重要。IC 的市值衰减不同于因子的市值暴露,前者用来衡量因子是否在大市值和小市值股票上都具有选股能力,后者用来衡量待检测因子和市值因子的相关性。通过将全市场股票按照市值大小分成 10 组,分别计算各组的 IC,观察比较 IC 的变化趋势来衡量 IC 市值衰减,该指标对应指标在各大指数内的选股效果。

图表19: 价格时滞因子 IC 市值衰减



资料来源: Wind 资讯, 方正证券研究所

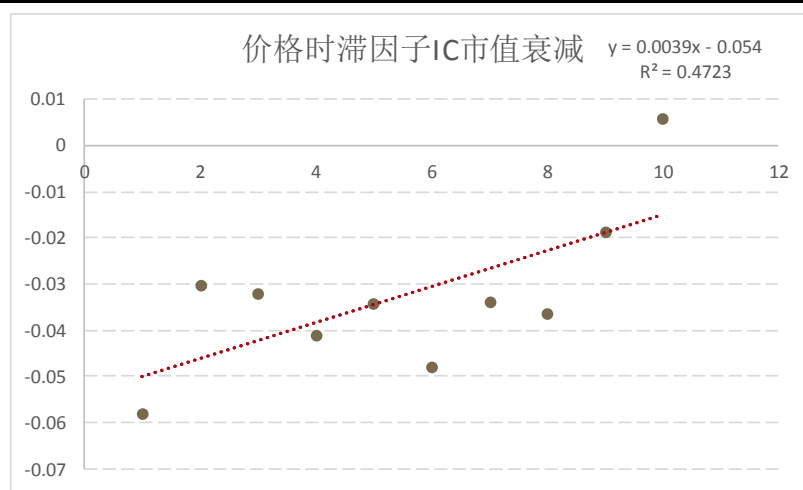
换手率、反转等价量因子体现出较强的市值衰减,即因子在小票上有很好的选股能力,但在大股票上效果欠佳。这解释了为什么有些因子在全局上具有很好的选股能力,但是在沪深 300 中选股效果一般。仍以价格时滞因子为例,因子在大市值上出现了明显的 IC 衰减现象,且单调性很强。如果以 IC 衰减速率的平均值来衡量这种现象,其含义是每经过一个因子分组 IC 下降的百分比,公式如下,价格时滞因子 IC 市值衰减速率高,达到 12%。

$$IC_{MV_Decay_V} = \left(\frac{IC_{MV_t}}{IC_{MV_{t-1}}} - 1 \right)$$

另一个方法是将市值分组 IC 回归，以回归的系数 k 乘以因子方向 I 代表 IC 的衰减，该方法更稳健，该指标衡量市值每增大一组 IC 平均下降的值。后续报告采取该方法计算因子 IC 市值衰减速度，价格时滞因子 IC 市值衰减量为-0.39。

$$IC_{MV_Decay_K} = K * I$$

图表20： 价格时滞因子 IC 市值衰减回归



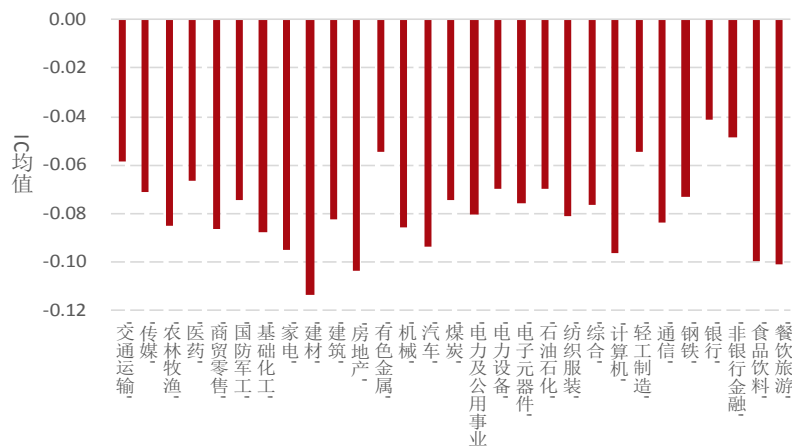
资料来源：Wind 资讯，方正证券研究所

5.8 IC 的行业衰减

IC 的行业衰减用来衡量因子在各个不同行业的选股能力，由于不同的行业有不同的投资逻辑，有的考察净利润，有的可能考察销量增速，有的更注重估值，观察指标的 IC 行业衰减有助于我们做更精细化的研究，进一步提高 IC 的预测能力。

以换手率因子为例，用 IC 均值来衡量因子在各个行业的选股能力。在银行、非银两个板块上，IC 衰减明显，换言之这两个板块上不适合用换手率进行选股。换手率因子一定程度可以衡量市场的交易热度，而金融类股票受交易情绪影响较少，股价操纵的程度不高，因此类似因子的选股能力偏弱。

图表21： 换手率因子 IC 行业衰减



资料来源：Wind 资讯，方正证券研究所

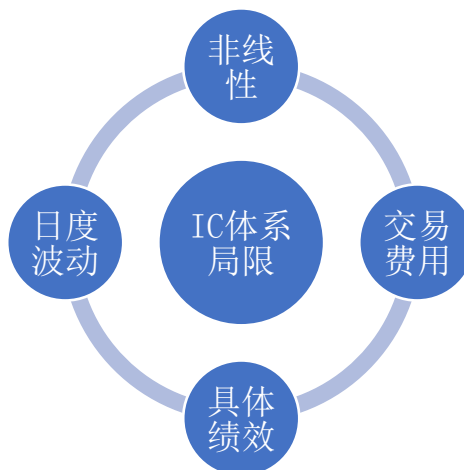
IC 行业衰减的结果可能稳定性稍差，因为行业分类相比市值分组更具主观，各个行业的股票数量有较大的差异。但是综合来看，仍能提供一定的指导意义。

6 因子评价之分组测试

上一节重点阐述了因子评价的 IC 体系，IC 体系简洁明了，一步相关性计算就能体现出因子的选股能力，在学术界备受追捧。不过在实际投资中，我们会将 IC 检验作为初步检验，而深入研究更多的还是会采取分组测试的方法来检验一个因子的选股能力，主要原因在于 IC 体系难以体现现实的复杂性。

例如反转类因子的多头收益钝化，指标收益非线性；例如交易费用对于收益的实际侵蚀；例如实际策略的日度波动特性；例如收益和回撤的具体大小等等。这些问题只有落实在组合中才能得到回答，这也是为什么有些高 IC 指标其实不能带来多少超额收益的原因。

图表22： IC 体系的局限性



资料来源：方正证券研究所

6.1 数据的分组

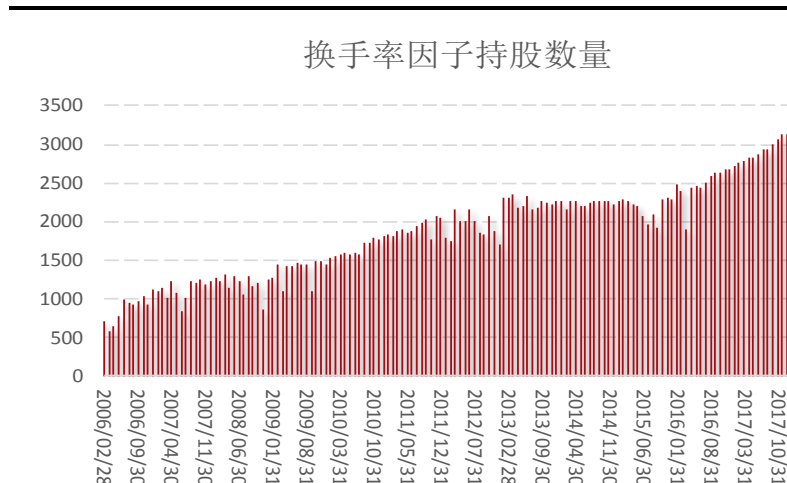
相比于 IC 在于数学上的优美，数据分组测试则更为实用。在每一个时间节点上，我们根据指标的大小等量地将所有股票分成 N 组，滚动计算每一组的累计收益率，真实模拟根据单因子进行投资的历史收益和回撤的情况。

常见分组方法为：假设被选股票数量为 N，将所有股票分为 10 组，则第 i 组的股票数量为 $[N/10*i] - [N/10*(i-1)]$ ，[] 为取整。该方法可以尽可能公允地将股票分到各个组别中。例如总股票为 301 支，则各组股票数分别为：30,30,30,30,31,30,30,30,30,30。

在个别案例中，若对极大值和极小值不敏感，也可采取将多余股票剔除，例如总股票为 301 支，则只选前 300 支进行选股，每组 30 支股票，共 10 组。

由于 A 股股票数量不断增加，在全市场选股时每个组合的持股票数量也会增多。如果在 300/500 指数内选股，则组合持股数量始终保持一致。

图表 23： 换手率因子全市场选股的持股数量



资料来源：Wind 资讯，方正证券研究所

6.2 组合的诞生

在分组构建组合时，我们需要特别注意的是交易条件限制，并不是所有股票每天都处于可交易的状态。停牌、一字涨跌停、成交量过低等客观原因会限制实际的买卖，另外我们主观上可能会倾向于交易非 ST 股票、上市一定时间的股票、剔除部分行业股票等等。

一般而言，我们会根据指标大小将股票分为 5 组或者 10 组，在月度调仓的频率下构建组合。假设初始资金是 1，那么等权组合的月度回报率 R 的算法如下所示。其中 T 为一个月的交易日个数，N 为组合内的股票数量， r_{it} 是股票每天的收益。

$$R = \sum_{i=1}^N (w_i * \sum_{t=1}^T (1 + r_{it})) \quad w_i = \frac{1}{N}$$

特别注意上述公式是站在每个调仓日的截点上计算，每次动
研究源于数据 19 研究创造价值

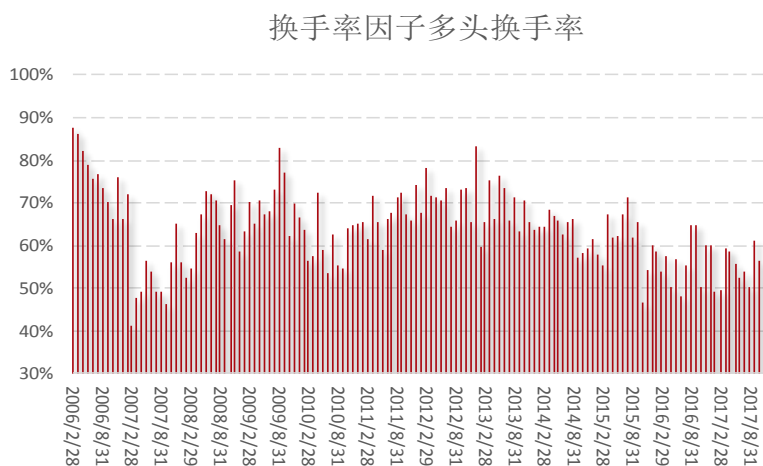
态平衡后会使得个股权重 w_i 重回 $\frac{1}{N}$ ，而事实上由于每日股票的波动，在非调仓日的截点上，个股权重并不相等。

每个调仓日，我们需要计算组合调仓的手续费 C ，手续费 C 可以简记为单边换手率 T 乘以单边费率 Fee 。单边换手率 T 的算法如下， P 为个股仓位，单边换手率等于当期仓位变动的绝对值总和的一半。

$$C = T * Fee$$

$$T = \frac{1}{2} * \sum_{i=1}^n \left| \frac{P_{2i}}{\sum_{i=1}^n P_{2i}} - \frac{P_{1i}}{\sum_{i=1}^n P_{1i}} \right|$$

图表24： 换手率因子的多头换手率



资料来源：Wind 资讯，方正证券研究所

6.3 指标评价体系

我们有一系列指标来衡量一个股票组合的表现是否优异，常见的指标有收益率、波动率、信息比率、最大回撤、换手率、胜率等等。

- 1、收益率用以衡量组合的盈利水平， R
- 2、波动率来衡量组合的风险，是收益率的标准差 $\sigma = \text{std}(r)$
- 3、信息比率来衡量组合的含义是单位主动风险所带来的超额收益，如果将基准设为 0，那么信息比率等于收益率除以波动

$$\text{率IR} = \frac{\bar{R}}{\text{std}(R)}$$

- 4、最大回撤用以衡量组合历史上发生过的最糟糕的情况，从历史高点下跌的最大幅度， $MD_t = \max(1 - \frac{p_t}{\max(p)})$

- 5、胜率用以衡量收益率为正的概率， W

图表25： 分组测试中的指标体系



资料来源：Wind 资讯，方正证券研究所

优秀的选股因子，各组收益率应出现明显分化，且具有较好的单调性。如市值调整换手率因子，IC 均值-7.31%、秩 IC 均值-10.25%、年化 ICIR -2.16、年化收益 40.26%、年化波动 15.56%、年化 IR 2.59、胜率 74.83%、最大回撤 16.41%、自相关性 63.11%、因子覆盖率 100.00%、IC 半衰期 1.07、IC 市值衰减速度-0.43%、多头年化收益 30.85%、多头收益占比 33.36%、因子方向-1。

图表26： 市值调整换手率因子的统计指标

IC均值'	秩IC均值'	年化ICIR'	年化收益'	年化波动'
-7.31%	-10.25%	-2.16	40.26%	15.56%
年化IR'	胜率'	最大回撤'	自相关性'	因子覆盖率'
2.59	74.83%	16.41%	63.11%	100.00%
IC半衰期'	IC市值衰减速度'	多头年化收益'	多头收益占比'	因子方向'
1.07	-0.43%	30.85%	33.36%	-1

资料来源：Wind 资讯，方正证券研究所

6.4 多头、空头与多空组

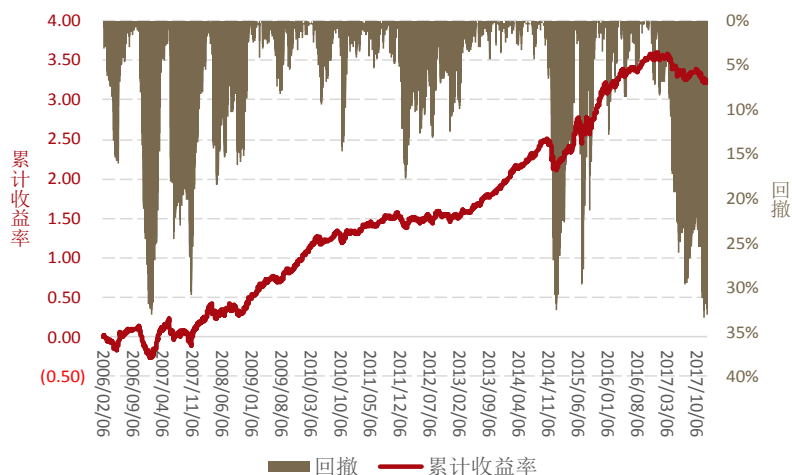
我们常用多空组的收益来衡量一个因子的表现，具体做法为做多第一组，同时做空最后一组，以多空组的表现来衡量因子的实际选股能力。理想状态下，多空两个组合在其他风险上暴露一致，仅在被检测因子上暴露不同，理想的 alpha 因子多空组净值（单利）应该是一条斜线稳定向上，多空组的收益和回撤可以较好地体现出因子的历史收益和回撤情况。

以投资者最感同身受的市值类因子为例，小市值股票从 2007 年到 2016 年，尤其是 2013 年到 2016 表现优异，仅在 2014 年 11 月出现较大的回撤，不过到了 2017 年方向出现反向，全年净值几乎都在下行。

我们多空累计收益率推荐以单利计算，因为我们相比较的是不同时期下的因子的净值走势，复利计算的净值曲线会加速加上，导致曲线失真。

图表27： 总市值因子多空组净值和回撤

研究源于数据 21 研究创造价值

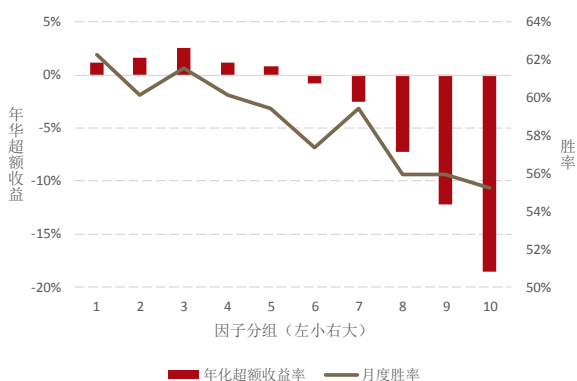


资料来源: Wind 资讯, 方正证券研究所

多空组的另一个隐含假设是因子收益在各组分布较为均匀, 那么只用多空组就可以较好的体现出因子的特征。但是实际情况是, 部分因子的单调性欠佳, 尤其是原始反转类因子多头收益钝化, 多空组的收益 94% 来自于空头的贡献 (行业中性后 82% 来自于空头的贡献)。另一方面, 10 分组下收益最高的组别出现在第 3 组, 单调性欠佳。由于 A 股实际交易中, 做空个股 (融券) 的成本较高限制较多, 该类因子的 alpha 较难获取。

反转类因子可以通过行业中性的方法加以改进, 其原理可能是行业上的动量干扰了个股上的动量效果, 月频上行业呈现出动量而个股呈现反转, 两者周期的错配可能导致个股动量的紊乱。

图表28: 反转因子各组收益



资料来源: Wind 资讯, 方正证券研究所

图表29: 行业中性反转因子各组收益

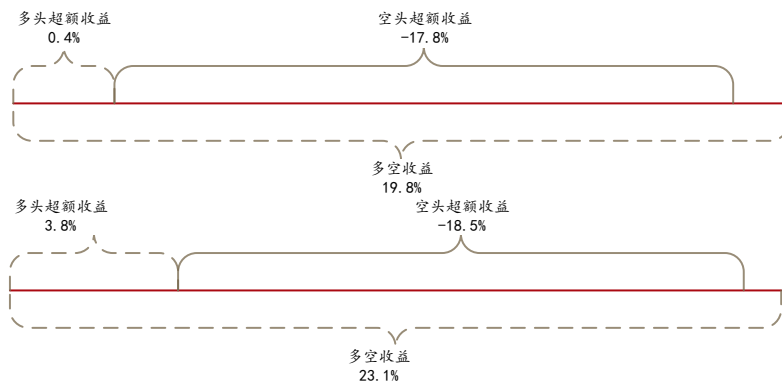


资料来源: Wind 资讯, 方正证券研究所

所以除了多空组的收益和分布外, 多头收益占比这个数据也值得关注, 可以参考下述算法, 该指标用来反映 10 分组下, 多头因子的收益占比。

$$\text{多头收益占比} = \frac{R_{\text{多头}} - \text{median}(R_i)}{R_{\text{多头}} - R_{\text{空头}}}$$

图表30: 多头收益占比

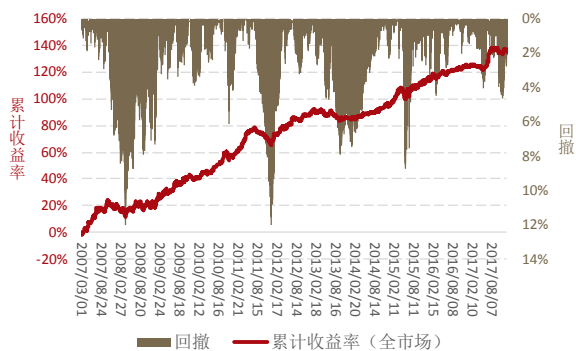


资料来源: Wind 资讯, 方正证券研究所

6.5 指数内测试

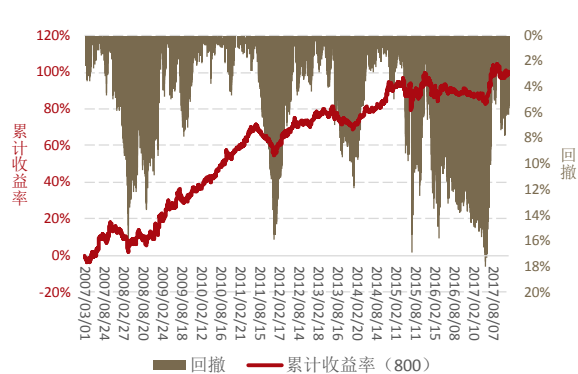
除了全市场选股, 我们常会做指数内选股, 常做的 300/500/1000 指数内选股的本质其实是考察因子的市值衰减。指数内选股又可以分为指数内等权和市值加权两种方法, 我们以等权为例展示单季度净利润同比增长率因子从 2007/02/01~2017/12/31 在全市场、800、500、300 上的选股表现。

图表31: 净利润同比增长率因子 (全市场)



资料来源: Wind 资讯, 方正证券研究所

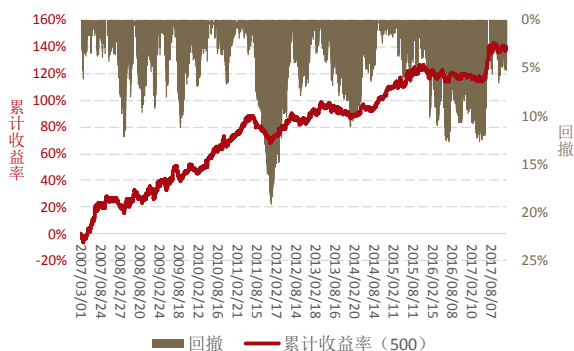
图表32: 净利润同比增长率因子 (中证 800)



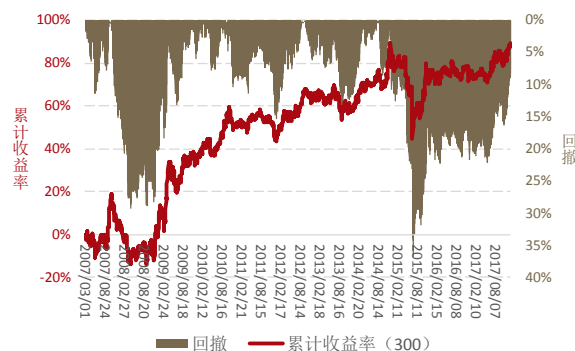
资料来源: Wind 资讯, 方正证券研究所

图表33: 净利润同比增长率因子 (中证 500)

图表34: 净利润同比增长率因子 (沪深 300)



资料来源: Wind 资讯, 方正证券研究所



资料来源: Wind 资讯, 方正证券研究所

6.6 基准的选取

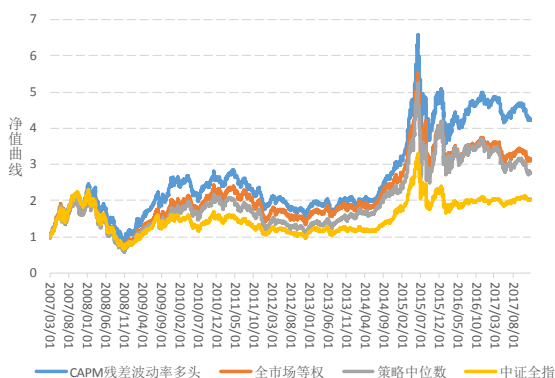
策略的比较基准也是一个值得讨论的话题, 一般我们会将策略的多头和基准指数进行对比, 比较两者在收益和波动上的差异。基准指数选的太弱, 可能会高估策略的收益; 基准指数选的太强则会倾向于低估。

我们展示一下 2007~2017 年 CAPM 残差波动率因子全市场选股多头与全市场等权、中证全指 (市值加权)、10 分组中位数的累计收益曲线与各组年化收益情况。

由于我国过去几年表现出强劲的小市值效应, 市值加权指数往往表现偏差, 普遍跑输等权指数。一般而言, 等权组合与等权基准比较, 加权组合与加权基准比较更为合理。由于小市值效应, 一个随机等权组合 (大猩猩选股组合) 在过去 10 年往往能跑赢多数加权指数, 但这并不能代表这个随机组合具有高超的选股能力。

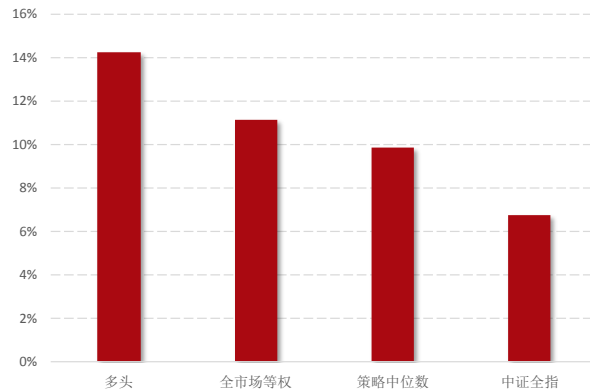
另外两个常被忽略的变量是分红和交易费用。交易费用会侵蚀我们的收益, 使得多头收益下滑, 降低超额收益。分红在组合中一般假设再分红投资, 而一般指数遇到分红会自然滑落, 导致超额收益虚高。严格的, 如果在多头上考虑分红再投资, 那在基准端我们应该用全收益指数进行比较。

图 35: CAPM 残差波动率因子累计收益



资料来源: Wind 资讯, 方正证券研究所

图 36: CAPM 残差波动率因子年化收益



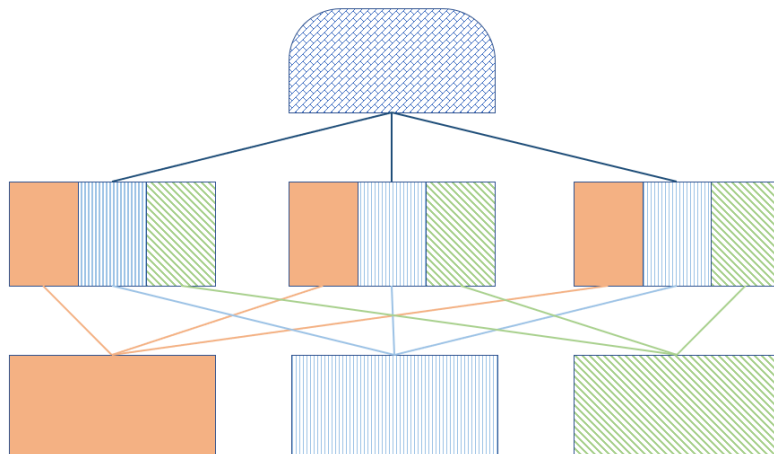
资料来源: Wind 资讯, 方正证券研究所

6.7 组合分层测试

常用的两种因子中性化方法，第一种在 4.5 节已经介绍，在因子的处理中先进行正交化再进行组合测试；第二种方法是做因子分层测试，具体方法为：

- 1、根据待中性因子 X 的大小将全市场股票分为 10 层
- 2、在每层中再根据待检测因子 F 将股票分为 10 组
- 3、每层中的第 1 组到第 10 组进行合并，得到新的分组

图表37： 因子分层示意图



资料来源：方正证券研究所

通过分组可以近似剔除 F 因子在 X 上的暴露，对比因子回归法，分组法的一个弊端在于无法对多个因子进行中性化，在行业中性化上也会遇到股票数量不对等的问题。模型构建中，我们采取回归法作为主要因子中性的手段，以分层法做辅助。

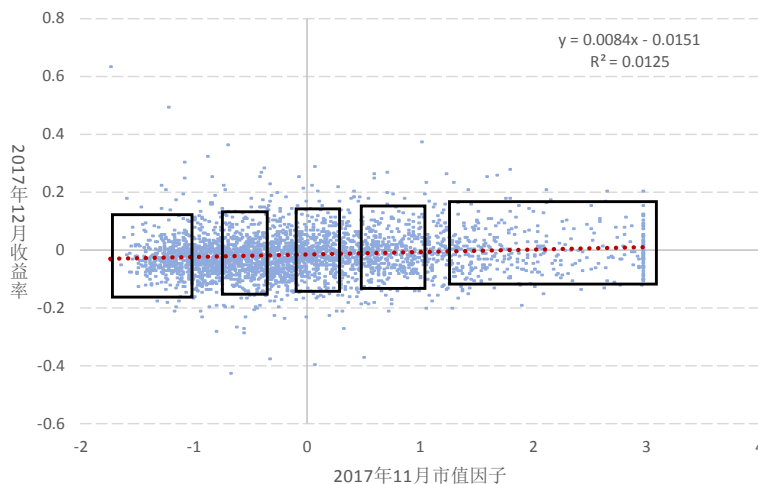
6.8 分组的本质

因子分组测试是弱化的线性回归，因此 IC 检验和分组测试的本质是一致的。两种方法都在回答，本期的因子值与下一期收益的联系。

以 2017/11/30 市值因子对下一个收益为例，下图横坐标是因子值，纵坐标是 12 月区间收益率。 IC^2 是回归中 R^2 ，用来衡量变量对于未来收益的解释程度。分组则是根据横轴因子值的大小，将股票等量分组，考察各组均值在纵轴即收益上的差异。

一个优异的选股因子 IC 可以达到 0.03+，对应的 R^2 在 1% 左右。在统计上，我们一般都要求可决系数 R^2 达到 0.2 甚至 0.5，1% 的 R^2 一般认为统计上不显著。但在实际投资中，这个数量级的 R^2 已经可以为我们带来可观的 alpha，这或许是学术和实际投资的差异。

图表38： IC 检验与分组测试的关系



资料来源：Wind 资讯，方正证券研究所

7 因子相关性检验

一般而言，有两种常用的检验因子相关性的方法，第一种是计算 A 因子和 B 因子在横截面的相关性，再时间序列平均，第二种是计算 A 因子分组中 B 因子百分位数均值。前者简洁，后者可以为我们展示更多细节。通过计算因子两两相关性，我们得到了常见因子的相关性矩阵，可以发现换手率和市值负相关，和波动率正相关等特性。

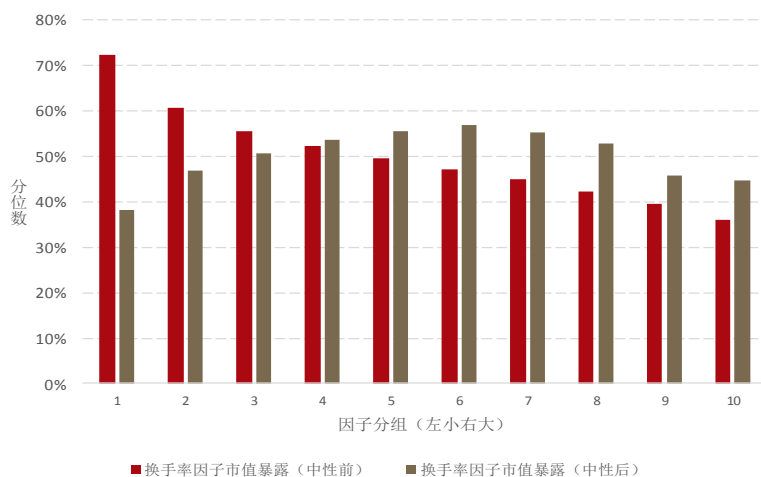
图表39： 因子相关性矩阵

	CAPM 残差波动率	资产负债率	总市值对数	20天反转	净利润增长率	ROA	PB	换手率
残差波动率		-0.07	-0.10	0.26	0.02	-0.02	0.30	0.44
资产负债率			0.07	0.00	-0.03	-0.37	0.03	-0.09
总市值对数				0.06	0.11	0.32	0.01	-0.12
20天动量					0.05	0.03	0.13	0.26
净利润						0.35	0.06	0.03
资产收益率							0.12	0.05
市净率								0.09
换手率								

资料来源：Wind 资讯，方正证券研究所

通过观察因子的相关性，可以对因子有更深入的了解。例如原始的换手率因子和市值相关性为负，通过第二种方法计算换手率 10 分组在市值上的暴露，可以发现低换手的股票更倾向于大市值，而高换手的股票更倾向于小市值。在中性化后，上述现象得到明显改善，换手率因子的收益也得到显著提升。

图表40： 换手率因子在市值上的暴露



资料来源: Wind 资讯, 方正证券研究所

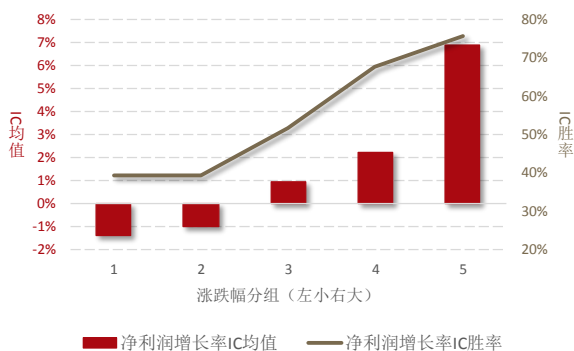
8 情景分析

在上述测试外, 我们还会对因子进行情景分析。我们发现有些因子在大盘上涨时表现更优, 有些因子在下跌中更好。通过对因子进行情景分析, 能帮助我们更好的构建后续的组合, 对因子进行择时。

根据下一期收益大小, 将 IC 值分成 5 组, 检验不同市场涨跌幅下的组合 IC 均值。CAPM 残差波动率因子在市场相对下跌时表现良好, 在市场上涨时表现较差。净利润增长率因子相反, 在市场上涨时 IC 均值更高, 预测能力更强。

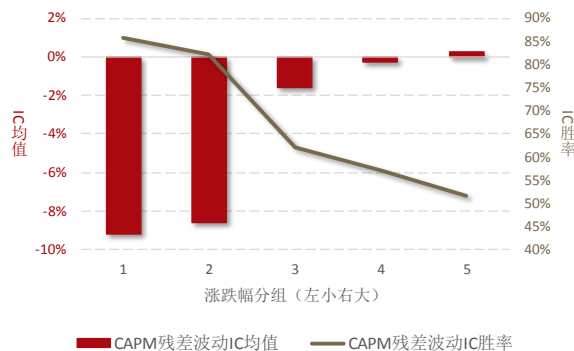
我们在后续的研究中会对该现象做进一步的研究。

图表41: 净利润增长率因子情景分析



资料来源: Wind 资讯, 方正证券研究所

图表42: CAPM 残差波动率因子情景分析



资料来源: Wind 资讯, 方正证券研究所

9 风险提示

本报告基于历史数据进行评价, 不构成任何投资建议。市场未来可能发生较大变化, 本报告因子评价模型仅供参考。

备注:

如没有特殊标注, 本报告回测期为 2006/01/01~2017/12/31, 月度调仓, 单边手续费为 6%。

感谢复旦大学实习生杨旭恒对本文的贡献!

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

方正证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。本报告仅供本公司客户使用。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司、本公司员工或者关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

本公司利用信息隔离制度控制内部一个或多个领域、部门或关联机构之间的信息流动。因此，投资者应注意，在法律许可的情况下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的情况下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

市场有风险，投资需谨慎。投资者不应将本报告为作出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“方正证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。

公司投资评级的说明：

强烈推荐：分析师预测未来半年公司股价有20%以上的涨幅；

推荐：分析师预测未来半年公司股价有10%以上的涨幅；

中性：分析师预测未来半年公司股价在-10%和10%之间波动；

减持：分析师预测未来半年公司股价有10%以上的跌幅。

行业投资评级的说明：

推荐：分析师预测未来半年行业表现强于沪深300指数；

中性：分析师预测未来半年行业表现与沪深300指数持平；

减持：分析师预测未来半年行业表现弱于沪深300指数。

	北京	上海	深圳	长沙
地址：	北京市西城区阜外大街甲34号方正证券大厦8楼(100037)	上海市浦东新区浦东南路360号新上海国际大厦36楼(200120)	深圳市福田区深南大道4013号兴业银行大厦201(418000)	长沙市芙蓉中路二段200号华侨国际大厦24楼(410015)
网址：	http://www.foundersc.com	http://www.foundersc.com	http://www.foundersc.com	http://www.foundersc.com
E-mail：	yjzx@foundersc.com	yjzx@foundersc.com	yjzx@foundersc.com	yjzx@foundersc.com