目录

[摘要 2](#_Toc432583979)

[Alpha来源概述 2](#_Toc432583980)

[多因子框架概述 2](#_Toc432583981)

[因子分析 4](#_Toc432583982)

# 摘要

根据现代金融理论，证券投资收益主要取决于市场的平均收益（即Beta收益）以及独立于市场的超额收益（即Alpha收益）。由于，Alpha的不同导致不同证券之间收益率的显著差异。因此，寻找出有效的Alpha因子将能帮助投资者获得超越市场的超额收益。本文主要着重于Alpha因子的挖掘，建立有效的Alpha因子挖掘流程以及衡量因子有效性标准，通过回溯测试直观的反映因子的效用性。

关键词：Alpha多因子，回溯测试

# Alpha来源概述

现代金融理论认为，证券投资收益主要来自于两部分：市场的平均收益（即Beta收益）以及独立于市场的超额收益（即Alpha收益）。根据资产定价模型（CAPM），证券投资的收益率可表示为：

其中，为市场上无风险收益率，表示为收益率残差。

由于，Beta收益基于市场整体表现与个股特征无关，因此如何挖掘具有显著Alpha特征的个股，从而获得超越市场平均表现成为了投资者追逐的目标。

对于Alpha的来源可谓是五花八门，有源于财务指标的，有源于技术指标的，也有源于舆情指标的。因此，本文主要侧重于Alpha因子的挖掘，以求建立起一个完善的因子数据库，最后根据该因子数据库构建适用于当下市场特征的多因子Alpha策略。

# 多因子框架概述

本文主要侧重于挖掘有效Alpha因子上，由于市场所包含的数据相当庞大，各种选股因子也非常之多，因此量化选股的一个大前提，就是要建立一个尽可能完备的数据库，基于这个数据库，从中挖掘历史以及当前市场中具有显著效用的因子。由于需要处理的数据量之大，时间关系，在回溯测试部分，仅对部分因子进行测试，而且回测时间样本区间设置较短。

1. **各类因子划分**

由于因子种类繁多，我将因子划分为如下几类：

1. 估值类因子：与股票估值相关的因子，包括常见的PE，PB，PCF等等；
2. 成长类因子：与股票的业绩成长等相关的因子，包括常见的净利润增长率，主营业务收入增长率等等；
3. 规模类因子：与股票规模相关的因子，包括常见的总资产，总市值，流通市值等等；
4. 质量类因子：与股票财务质量，资本结构相关的因子，包括常见的ROA，ROE，资产负债率等等；
5. 技术类因子：与技术指标相关的因子，如MACD，RSI等等；
6. **因子测试数据处理**

我对各类因子建立起一个统一的数据处理过程：

1. 所有数据均来自底层数据库；
2. 回测周期为：2014.01.01-2014.03.31
3. 基准指数统一使用沪深300；
4. 因子数据预处理，包括去极值化以及标准化：

* 去极值化：本文使用“中位数去极值法”进行去极值化处理，公式如下：

其中为第i个描述性变量的观察值，为所有观察值得中位数，将观察值与中位数的绝对偏离记作，即=，将记作的中位数，和分别表示中位数去极值化后描述性变量的上下限，n用于控制和的常数，本文取n=3。

* 标准化：由于各个描述性因子所衡量的单位不同，导致因子数值范围差异较大，因此在进行因子分析之前，必须对其进行标准化，本文采用最常见的N(0,1)正态标准化处理，公式如下：

标准化后向量=（原向量-均值）/标准差

1. **因子评价标准**

如何对因子的有效性进行衡量是本文研究的核心，我认为一个有效的Alpha因子应该能够带来长期稳定的Alpha收益，同时因子在各期的表现应该具备较好的稳定性，即因子的beta系数相对较低，以保证因子的与市场的相关性相对较低。另外，市场上存在动量/反转效应，当股价持续大跌的时候往往会出现反转，因此对于多因子选股中的超低配组合是否具备较高的胜率也是考察的标准之一。

综上所述，本文采用五等分位法替代因子打分法来考察各个因子的表现，以有效性指标来衡量各个因子的有效性。

**有效性指标**：通过跟踪超配与超低配组合的表现来考察因子的有效性，包括IC，IR，因子胜率，累计收益率。

（1）**因子IC（信息系数）**：即每个时点因子在各个股票的暴露值与各股票下期回报之间的相关系数，本文认为如果一个因子的IC值的绝对值高于2%，则认为该因子在优选个股Alpha收益上有较好的效果，IC为正表示该因子与股票未来的收益率呈正相关，应该超配因子暴露值高的股票，反之若IC为负，则低配因子暴露值低的股票。

（2）**因子IR（信息比）**：即因子在样本期间的平均年化收益与年化平均标准差的比值，IR的绝对值越高，表明该因子在优选个股Alpha收益上效果越好，另外，经统计发现，IR的绝对值高于0.7时，Alpha因子的选股效果通常比较显著，另外，若IR为正，则代表超配因子暴露值高的个股，反之若IR为负，则低配因子暴露值低的股票。

（3）**因子胜率：**所谓因子胜率，就是统计在回测时间样本空间里，因子收益为正的次数占全样本长度的比值，比值越大，证明由因子挑选的超配组合收益战胜低配组合收益的成功率越高，该因子的有效性越显著！

（4）**累计收益率：**用于衡量因子Alpha在回测时间样本区间是否具有稳定且可持续收益。

（5）**最大回撤率：**统计在回测时间样本区间里，组合与最高点的亏损率，反应因子Alpha的抗风险能力，若回撤比率较大，则因子风险暴露值较大，反之，若回撤比率较小，则因子风险暴露值较小。通常最大回撤比率越小，表明因子Alpha越稳定且风险可控。

1. **挑选有效因子**

本文主要侧重于对有效选股因子的挖掘，而关于如何根据挖掘出的有效Alpha因子构建相应的Alpha策略，将在其他报告里进行介绍。

综合考虑各个Alpha因子在各项度量指标上的表现，本研究进一步从各类Alpha因子中挑选出具有较强代表性以及选股效果最佳的Alpha因子。

1. **备选因子**

由于因子的来源五花八门，为了更好捕捉到有效因子且适应不同投资时期。因此，对于各个因子进行归类，在尽可能考虑多维度的情况下，选取合适因子。同时，也要求所选因子具有经济逻辑以及投资逻辑。因此，本文根据经济逻辑以及投资逻辑将因子分为估值类因子，规模类因子，成长类因子，质量类因子以及技术类因子等五大类。并在日后根据相关逻辑进一步扩充因子维度。

# 因子分析

**回测样本标的：**全部A股，由于股票数据可能存在缺省问题，因此剔除ST，因子数据缺省股票，以及回测区间样本内上市不满一年的股票。

**回测区间时间样本：**2014.01.01-2014.03.31

**回测区间交易数据处理：**将回测区间样本进行前复权处理，以解决回测区间样本内可能存在的增转送股，股息分派导致股价差异巨大的问题。

**因子组合权重：**本文主旨在于建立起可用因子数据库，而非构建多因子Alpha策略，因此对于暂时使用等权重方式配比组合权重。

由于因子库数据处理量较大，本文以质量因子ROE为例进行数据处理。

**ROE因子分析结果：**

将ROE从大到小排序（去除ROE为负的证券），使用五等分位法分为五档收益率。其中，一档为超配组合，五档为低配组合。统计2014.01.01-2014.03.31区间内，一档超配组合表现跑输基准，五档低配组合表现相对优异。一档与五档胜率分别为45.61%，54.37%。各档Beta值均高于1，显示ROE因子在样本区间内与基准指数相关性较大。尽管各档最大回撤比率均低于20%，但是，各档组合IC，IR暴露值均不显著。其中，一档IR、IC为-0.0147，-0.0007，五档IR、IC为0.1373，0.0068，远低于限定暴露阀值。因此，在回测区间样本中，ROE因子表现并不优异，不呈现显著Alpha收益。

ROE因子在此区间内暴露值过小，可能原因有二。

1. 回测区间时间样本跨度较长，若改为以周为单位进行回测，某些时段暴露值应达到限定阀值，成为显著要素。
2. 在A股市场上，ROE并非投资者重要参考要素，投资者对上市公司的ROE好坏，并不敏感。

表一：2014.01.31-2014.03.31

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ROE | IR | IC | Sharp | Beta | 最大回撤率 | 累计收益率 | 胜率 |
| 一档 | -0.0147 | -0.0007 | -0.0649 | 1.0933 | 0.1233 | -0.0444 | 0.4561 |
| 二挡 | 0.0193 | 0.0010 | -0.0352 | 1.2208 | 0.1408 | -0.0257 | 0.4737 |
| 三挡 | 0.0764 | 0.0038 | 0.0046 | 1.2112 | 0.1195 | 0.0111 | 0.5437 |
| 四挡 | 0.0977 | 0.0048 | 0.0244 | 1.1404 | 0.1226 | 0.0304 | 0.5437 |
| 五档 | 0.1373 | 0.0068 | 0.0457 | 1.1966 | 0.0981 | 0.0505 | 0.5437 |

图1：区间每日波动

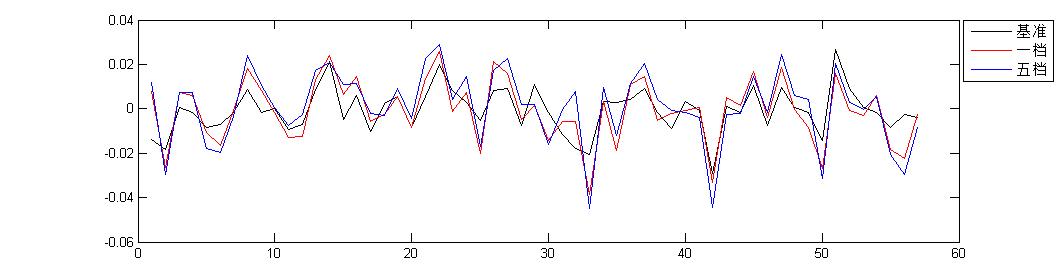


图2：区间累计收益率

