证券研究报告—深度报告

金融工程

数量化投资

数量化投资技术系列报告之五十三

2012 年 03 月 15 日

专题报告

基于核密度估计的选股因子分布差异测度和中性策略

相关研究报告:
《数量化投资系列之二十六：多因子 Alpha 选股－将行业轮动落实到 Top 组合》 ——2010-5-5
《数量化投资技术系列之四十五：基于 A 股市场选股因子边际效用和有效分散的动态区分度动量策略》 ——2011-08-31

证券分析师：董艺婷
电话：021-60933155
E-MAIL：dongyt@guosen.com.cn
证券投资咨询执业资格证书编码：S0980510120055
联系人：李荣兴
电话：021-60933165
E-MAIL：lirxing@guosen.com.cn

- **用核密度法估计因子概率密度分布**

因子分布是指股票组合中具有不同因子值的股票出现的概率密度分布。我们通过核密度估计法(Kernel Density Estimation)获得因子分布。核密度估计法作为常用的非参数估计方法，可以有效估计未知分布。核密度估计法可以选择不同的核函数，但是不同核函数对估计效果影响较小，而窗宽 h 对估计结果有重要影响。为了在获得因子分布之后能够分析组合分布与基准分布的差异，我们引入交叉熵及其变体来作为分布差异的测度。通过分析我们选择满足对称性并且可以减少差异点的非零约束的 L 测度来衡量差异分布。经过对累计差异分布函数的分析，我们可以发现组合中市值分布或者其他因子分布与基准分布之间的差异来源，并发现组合配置中存在的因子偏离状况，从而**控制组合在市值或者其他因子上暴露的风险**。

- **从因子分布出发构建因子中性策略**

为了减少组合在某个因子上暴露的风险，我们构建因子中性策略。此处因子中性是指组合在该因子上暴露的风险与指数基准股票池是一致的。在最理想的情况下，该因子不对组合贡献超额收益。目标函数是使得**组合分布与基准分布之间的差异最小化**。我们提出因子权重循环调整法，通过调整组合中不同股票的权重，使得组合最终的因子分布尽可能地贴近基准分布。根据情况不同可能使差异减少 30%到 99%以上不等。

某些情况下组合中会缺少某类股票，例如大市值股票，这将导致组合在市值因子上的分布与基准差异较大。我们可以通过因子分布的分析发现曲线差异来源于曲线尾部，即大市值部分。这种组合无法通过调整权重贴近基准分布，因此必须调整组合，添加大市值股票。尽管市值加权的方法也可以调整市值分布，但是由于市值加权必须已知市值因子在基准中的分布方法，不适用于其他因子，所以权重循环调整方法仍然具备较大的优势。

- **中性效果检验和策略评价**

由于回归方法在因子业绩归因中存在的各种问题，我们使用因子区分度来代替因子溢价进行因子收益归因，证明了减少分布差异后因子对业绩收益的贡献有效地减少，组合在因子上暴露的风险得到了控制。我们最后使用国信的策略模拟和评价体系对本报告使用的单因子策略的例子进行了评估和分析。

内容目录

图表目录

前言

多因子之惑

多因子模型作为现代量化投资体系的重要研究方向，其最初的目的是为了解释不同风险因素对股票或者债券的报酬的影响。在收益预测方面多因子模型也有重要的应用。对多因子模型的研究方法主要包括回归和非回归方法（如国信的因子区分度方法）等，不同的方法在多因子应用领域各有利弊。但不论使用何种方法，目前业内对多因子的研究都主要面临两类问题：因子的不完全挖掘和因子相关性问题。

因子的不完全挖掘对于回归模型有显著影响，目前有相当数量的研究报告集中于挖掘新的显著因子。但是与不完全挖掘相比，因子的相关性问题影响更为重要。这里的相关性问题包括两类：一类是因子的共线性问题，共线性直接影响到回归模型对于风险溢价的估计，并且常常使得多因子模型的风险归因出现较大的偏离。第二类是因子相关性的不缺定性问题。由于市场风格转换和因子轮动的存在，因子之间的关系表现极为复杂，因子的相关性随着时间推移有明显变化，以至于回归意义上的相关性研究常常无法得到在统计上显著的结果，甚至共线性本身也是不确定的问题。

在国信金融工程的专题报告《数量化投资技术系列之四十五：基于 A 股市场选股因子边际效用和有效分散的动态区分度动量策略》中我们指出，因子之间存在明显的互相干扰现象，随着因子数增加，多因素模型的收益稳定性逐渐下降，增加因子带来的边际效用逐渐减少。从中可以看出，因子的相关性是比不完全挖掘更本质的问题。对因子相关性的研究必然涉及到股票组合中因子的分布问题。因子在投资组合、股票池中的分布以及两个分布之间的差异包含着极大的信息。对多因子的联合分布的研究是极为复杂的工作，必须立足于对单因子的边缘分布的研究。本报告主要集中于研究单因子的分布以及因子在选股组合和股票池中的分布差异，作为因子相关性研究的第一步。在此基础上，我们构建了基于核密度分布估计的因子中性策略，作为因子风险控制的一个工具。

研究框架

为了研究股票组合中的因子分布，我们引入非参数估计的核密度方法对组合中的因子分布进行估计。投资组合与沪深 300 等股票池的分布对比可以直观地展示投资组合在因子上暴露的风险。为了定量地衡量两个不同分布之间的差异，我们引入了基于交叉熵（Relative Entropy）的 L 测度，在此过程中我们对于分布的尾部进行了探讨。L 测度对于我们控制组合在因子上的风险有很重要的意义。

在引入因子分布的基础上，我们提出构建单因子中性策略的方法。通过调整组合中不同股票的权重，我们可以使组合的因子分布不断接近市场基准的因子分布，使得组合在因子上暴露的风险与市场保持一致。最后我们对基于因子分布的中性策略进行了业绩归因意义上的检验，达到了较好的效果。

因子分布的估计与差异测度

当用多因素模型进行选股时，根据选择的因子不同，选出的股票组合在因子分布上与股票池的偏离也有所不同。我们常常希望了解选出的股票组合是否真的满足多因素模型的初衷。但是目前对组合的因子性质的分析主要来自于事后的收益分解，因子溢价等指标实际上也是对收益率进行回归的结果。为了能够更直观地了解在多因素模型下组合的性质，我们希望能够拟合组合股票在不同因子上的概率密度分布，同时也希望了解组合与基准股票池之间的分布差异。这里所说的因子分布，是指股票组合中具有不同因子值的股票出现的概率密度分布。我们将因子在股票池中的分布称为**基准分布**，因子在组合中的分布称为**组合分布**。

核密度估计法

对于某些指数，由于在选择成分股时对某类因子有一定约束，所以基准分布存在正态分布的可能。但是对于不同的因子，其基准分布千差万别，假设其满足正态分布或者其他预设的分布是不合理的。因此我们在此使用非参数估计方法，并不假设基准分布符合某种参数分布，在不确定基准总体分布的情况下通过非参数的方法进行估计和检验。

基本原理

在诸多非参数估计方法中，核密度估计法被广泛接受和使用。核密度估计法（Kernel Density Estimation）作为一种非参数估计方法，，由 Rosenblatt (1955)和 Emanuel Parzen(1962)提出，又名 Parzen 窗（Parzen window）。核密度估计认为某一点 x 处的密度函数估计值的大小与该点附近所包含的样本点的个数有关，若 x 附近样本点比较稠密，则该点附近概率密度的估计值应该较大。为此定义以 x 为中心，以 h/2 为半径的邻域，计算落入该邻域内的样本点，这里的 h 即为窗宽。最简单的 Parzen 窗密度估计将 x 领域内的所有点看成是同样重要的，这并不合理，实际上应该按照领域内各点距离 x 的远近来确定它们的贡献大小，模拟这个远近贡献的函数就是核函数(kernel function)。

表达式与窗宽

核密度估计常用的表达式如下：

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})$$

其中 K(.)称为核函数，h 称为窗宽，n 为样本个数。为了保证 $f_h(x)$作为密度估计函数的合理性，要求核函数满足：

$$K(x) \geq 0, \int_{-\infty}^{+\infty} K(x)dx = 1$$

即 K(x)是某个分布的密度函数。

常用的核函数包括均匀核函数、三角核函数、高斯核函数等。但是研究表明，核函数本身的选择对于核估计的结果影响不大，因此我们在本报告中统一使用高斯核函数进行估计。

核密度估计中的窗宽 h 会影响到估计曲线的光滑程度。h 选得越大，将会有越多的点对 x 处的密度估计产生影响，并且距离 x 较近和较远的点对应的核函数

值差距不大，此时估计曲线的图像会比较光滑，但是同时会丢失数据所包含的独立信息。如果 h 选得越小，则 $f_h(x)$是一条不光滑的折线，但它能反映各个数据所包含的信息。因此选择窗宽对于核估计的结果影响重大。一种常用的求最佳窗宽的方法是 MISE 法，令

$$MISE(f_h) = E\{\int [f_h(x) - f(x)]\,dx\}$$

其中 f(x)是真实分布。求 MISE 函数的最小值点，即可得到最佳窗宽的估计值。根据求导及极限计算可以得到 MISE 取最小值时，最佳窗宽为

$$h = \left\{\frac{\int [K(x)]^2 dx}{\sigma_k^4 \int [f''(x)]^2 dx}\right\}^{\frac{1}{5}} n^{-\frac{1}{5}}$$

其中$\sigma_k^2 = \int x^2 K(x)\,dx$。特别地，当总体服从$N(0,\sigma^2)$分布，核函数为高斯核函数时，最佳窗宽为

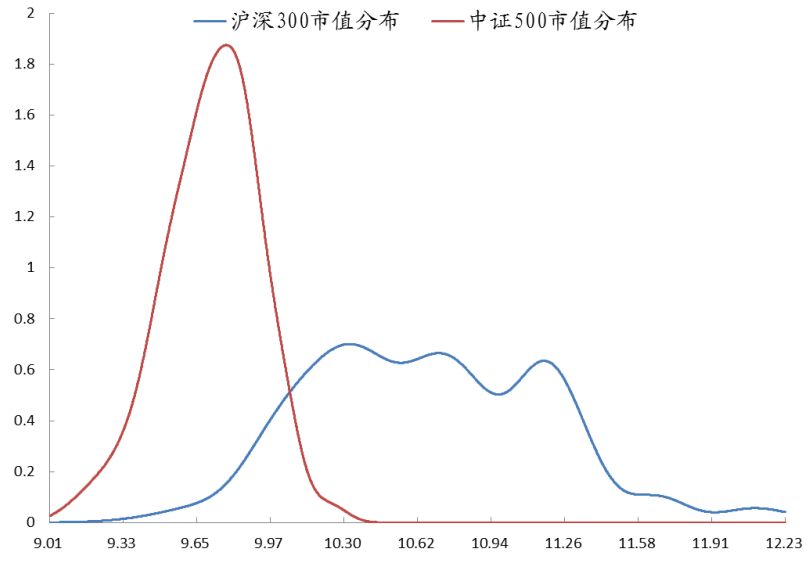
$$h = (\frac{4}{3})^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} = 1.06\sigma n^{-\frac{1}{5}}$$

因子分布的估计与检验

目前多种软件均提供核密度估计函数，我们使用 Matlab 中的 ksdensity()函数对核密度进行估计，该函数使用上述最佳窗宽作为窗宽默认值，高斯函数作为核函数，还可以对样本进行赋权估计。

如下图所示是以市值因子为例所画的沪深 300 和中证 500 股票池的概率密度分布图。由图可见中证 500 的股票市值集中在低市值范围，而沪深 300 主要是大市值股票（已加入权重，并对市值进行了对数化处理）。对于其他因子可以使用同样的方法进行估计。

图 1：沪深 300 和中证 500 市值因子概率密度分布图



资料来源：Wind 科技，国信证券经济研究所

我们使用蒙特卡罗法对核密度估计的结果进行非参数检验。首先从带权重的因子集和核密度估计得出的概率密度函数中分别随机抽取 N 个样本，通过多个检

验量检验这 N 对样本是否来自同一个总体分布，检验重复进行 10000 次，如果得到“两样本没有显著差异”的次数足够多，则我们可以认为原假设通过了检验。用此方法对上图中的沪深 300 和中证 500 分布进行检验得到的结果如下：

表 1：沪深 300 市值分布检验结果，N=300

原假设	检验方法	检验次数	通过	拒绝	正确率
两样本来自的总体分布无显著差异	K-S 检验	10000	9120	880	91%
	秩和检验	10000	9507	493	95%
	符号秩检验	10000	9527	473	95%

资料来源：Wind 科技，国信证券经济研究所

表 2：中证 500 市值分布检验结果，N=500

原假设	检验方法	检验次数	通过	拒绝	正确率
两样本来自的总体分布无显著差异	K-S 检验	10000	9209	791	92%
	秩和检验	10000	9501	499	95%
	符号秩检验	10000	9543	457	95%

资料来源：Wind 科技，国信证券经济研究所

由结果可知核密度估计的拟合结果是比较好的，对于较差的拟合结果，我们可以通过改变窗宽重新估计，得到新的估计曲线，直到通过检验为止。如果在应用中主要考虑定性的结果，则可以适当放宽检验要求。

分布差异测度

为了直观地了解投资组合的因子暴露，我们需要衡量组合分布和基准分布之间的差异。当我们需要比较多个组合时，不同组合分布和基准分布的差异的比较显得更加有意义。因此我们需要一个统一的指标测度来衡量概率密度分布之间的差异。

交叉熵及其变体

对于衡量两个概率密度分布之间的差异存在许多相关研究，在文献[1]中，作者总结了表征两个分布差异的多种方法。其中最为常用的是 Kullback-Leibler Distance 及其变体，也叫**交叉熵**(relative entropy)，其表达式为：

$$H(f_1, f_2) = -\int \frac{f_1(x)\, log\big(f_2(x)\big)}{f_1(x)} dx$$

其中，f1 和 f2 是我们要研究的两个不同的分布。上述表达式跟变量的连续型分布是相一致的，对于离散形式，表达式相应的转变为：

$$I(P_1, P_2) = -\sum_{x \in X} P_1(x)\, log \frac{P_2(x)}{P_1(x)}$$

但交叉熵测度存在的最大问题是不满足对称性，即 $H(f_1,f_2) \neq H(f_2,f_1)$，我们对于分布曲线的差异需要有一个统一的度量，因此这个问题对于我们的应用来说有比较大的影响。因此我们改用变形的 Jeffreys-Kullback-Leibler divergence:

$$\begin{aligned} J(f_1, f_2) &= H(f_1, f_2) + H(f_2, f_1) \\ &= -\int f_1(x)\, log \frac{f_2(x)}{f_1(x)} dx - \int f_2(x)\, log \frac{f_1(x)}{f_2(x)} dx \\ &= \int (f_1(x) - f_2(x)) log \frac{f_1(x)}{f_2(x)} dx \end{aligned}$$

在程序计算中我们只能使用离散形式：

$$J(P_1,P_2)=\sum_{x\in X}(P_1(x)-P_2(x))\,log\frac{P_1(x)}{P_2(x)}$$

很明显 J 测度满足对称性即 $J(f_1,f_2)=J(f_2,f_1)$。但 J 测度仍然存在问题，J 测度中的对数表达式要求 $P_1(x)$和 $P_2(x)$均不为 0。在核密度估计的实际应用中，我们通常是比较股票池和来自股票池的投资组合的分布差异，后者是前者的子集，因此 $P_1(x)$和 $P_2(x)$至少有一个不为 0，假设 $P_1(x)$为基准分布且不为 0，则只需要保证 J 测度的分母不为 0 即可，为此我们需要对分母进行处理。Jianhua Lin 在文献[2]中重新定义的表征分歧的指标：

$$K(P_1,P_2)=\sum_{x\in X}P_1(x)\,log\frac{P_1(x)}{\frac{1}{2}P_1(x)+\frac{1}{2}P_2(x)}$$

K 测度同样不满足对称性，因此进一步引入 L 测度：

$$L(P_1,P_2)=K(P_1,P_2)+K(P_2,P_1)$$

L 测度满足对称性的需求，同时可以解决分母为 0 的问题，因此我们将其选为衡量因子分布差异的绝对指标。

分布差异的相对指标

上述 L 测度是衡量差异的绝对指标，对于不同的基准曲线，差异反映出来的 L 测度没有可比性。而且由 L 测度反映出来的绝对差异仅仅是一个数值，没有直观的含义，不利于我们理解分布之间的差异大小。因此我们可以尝试用以下方法构建相对指标：

对于概率密度分布 f(x)，对于任意 x，令

$$ftop(x)=(1+10\%)f(x)$$

$$fbottom(x)=(1-10\%)f(x)$$

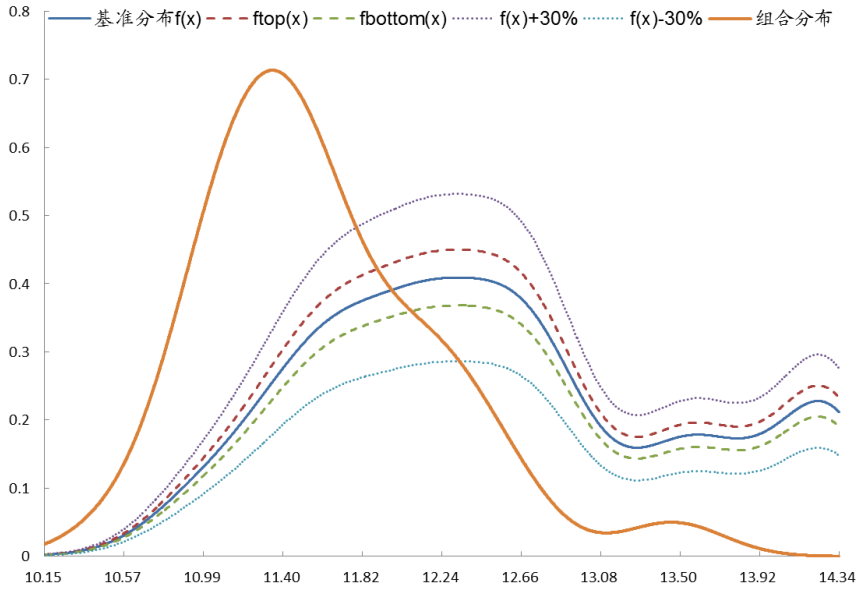
所有落入 ftop(x)和 fbottom(x)之间的曲线，其与 f(x)的差距都将小于 ftop(x)或者 fbottom(x)与 f(x)的差距，更确切地说，由于 fbottom(x)与 f(x)的差距更大，因此落在此范围内的曲线与基准的差异不会大于 fbottom(x)。此处的 ftop(x)和 fbottom(x)只是虚拟的曲线，实际上这两条曲线所代表的概率密度分布在归一化后即为 f(x)，这两条曲线只是一种形式表达式。

设 fbottom(x)与 f(x)的差距是 d_0，对于不同的基准曲线 $f_i(x)$，d_0有不同的值。但是 d_0所反映的含义是密度分布曲线上下 10%的范围，这对于不同的基准曲线是一样的。假设组合的概率密度曲线和基准曲线的 L 测度差异为 d，则 $D=d/d_0$为组合和基准曲线的相对差异。这样即使基准曲线不同，D 也可以作为一个对比的指标。

如下图所示是某时点沪深 300 的股票市值分布和以沪深 300 为股票池的某个选股组合的市值分布。对于核密度估计后的概率密度曲线，如果取的点较少，将会出现较大的偏离导致计算错误。因此我们从中取 5000 个等间隔点计算两条曲线的差异，取值区间的起点是股票池中因子值的最小值，终点是股票池中因子值的最大值，样本点的概率密度值用插值的方法得到。为了应用离散形式，我们对样本点的概率密度进行归一化，使所有样本点的概率密度之和为 1。

我们根据计算得到 $d_0=0.1143$，而组合分布的概率密度曲线与基准分布的差异是 $D=95.6d_0$。

图 2：因子分布的相对差异测度



资料来源：国信证券经济研究所

为了对 D 的大小有一个更直观的认识，在上图中我们画出了另外两条虚拟曲线 f(x)+30%和 f(x)-30%，这是基准分布 f(x)上下 30%的范围内的两条虚拟曲线。这两条曲线和基准已经有 30%的较大差异了，其中 f(x)-30%的曲线与 f(x)的差异 L 测度是 7.45d_0。可以看到，组合分布与基准分布的差异竟然是 f(x)上下 30%曲线的 13 倍！该组合在市值上的分布与基准有极大差异，我们将继续进行解释说明。

因子风险控制

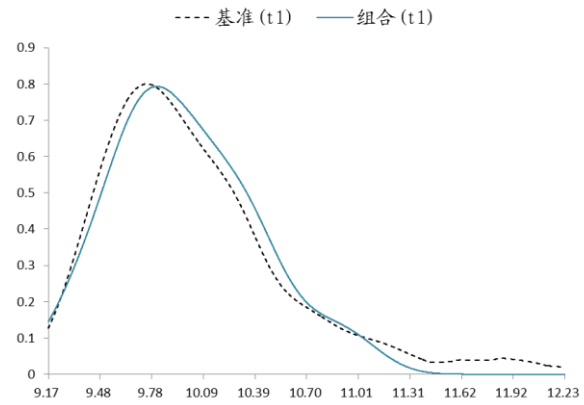
当从基准股票池中选出一个股票组合时，我们希望知道组合的性质与市场基准的差异有多大，差异出现的原因在哪里，进而才有控制差异的可能。传统的业绩归因理论只能得到简单的因子暴露的大小，并不能够进行更具体的分析。而且这种归因主要依赖于事后的回归，对于事前指导投资意义有限。

通过分布函数我们可以对因子的风险进行更细致的分析。

首先看以下两个分布的例子。左图是沪深 300 等权基准在 t1 时刻的市值分布以及策略在 t1 时刻选出的股票组合分布，右图是同一基准在 t2 时刻的市值分布以及策略在 t2 时刻选出的股票组合分布。图中基准的变化较为微小，两者几乎是一致的，而两个时刻选出的不同组合在市值分布上有一定差别。我们使用 L 测度对两者的差异进行计算。计算结果显示组合(t1)和基准(t1)的差异为 Diff1=9.96d_0，组合(t2)和基准(t2)的差异为 Diff2=7.79d_0。

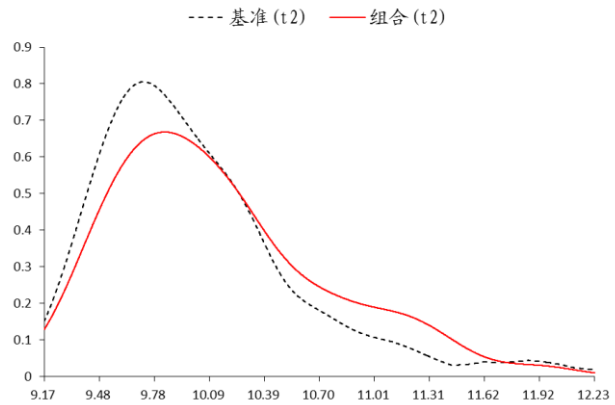
这个结果可能有些奇怪，因为组合(t1)的市值分布从直观上来看已经非常贴近基准分布，但计算结果显示 Diff1 要大于 Diff2，而且这个结果也大于基准上下 30%的虚拟曲线与基准的差异 7.45d_0，也就是说组合与基准存在较大差异。那组合的问题出在哪里？

图 3：t1 时刻沪深 300 等权基准分布和 t1 组合分布



资料来源： WIND 资讯，国信证券经济研究所

图 4：t2 时刻沪深 300 等权基准分布和 t2 组合分布



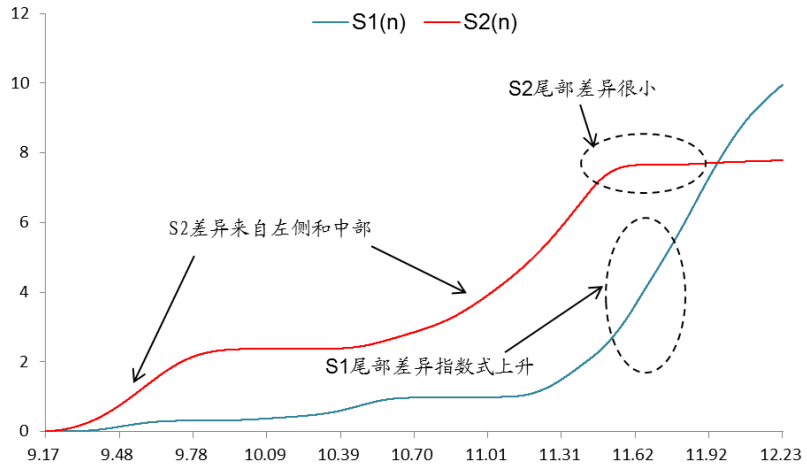
资料来源：WIND 资讯、国信证券经济研究所整理

为了进一步分析差异的来源，我们计算两者的差异累计函数：

$$S(n) = \frac{\sum_{i=1}^{n} \mathrm{diff}(x_i)}{d_0}$$

得到的曲线分布如下图所示：

图 5：组合与基准的差异累计函数



资料来源：国信证券经济研究所

从图中可以看到，组合(t2)的分布差异主要来自于分布曲线左侧和中部，这符合我们的直观感觉；组合(t1)的分布差异来自分布曲线尾部，而且几乎是指数式上升。观察分布曲线我们可以发现，在基准中存在一定数量的大市值股票，但是这些股票没有被选入组合中，因此组合的概率密度曲线在尾部非常接近于 0，与基准分布 $P_1(x)$的比值很大，最后导致两者差异明显大于 d_0！由此我们可以根据这个结果在组合(t1)中添加大市值股票，使之减少与基准的分布差异，减少在市值因子上的风险暴露。

我们再回到图 2 中的组合。这实际上是一个沪深 300 中选出的等权组合的市值分布。该组合由于是等权分布，放大了小市值股票的作用，因此整体分布向小市值方向偏移。该组合存在的另一个问题是缺少极大市值的股票，因此在曲线尾部出现了概率密度趋近于 0 的极小值，使得组合分布曲线与基准差异较大。解决的办法之一是提高大市值股票的权重，但是由于组合中不存在极大市值股

票，无法通过调整权重提高尾部曲线，因此必须加入极大市值股票。

在接下来的内容中，我们将提出一种调整股票权重的方法，在不需要改变股票组合的情况下使得组合的因子分布尽可能地接近于基准分布。

基于因子分布的中性策略

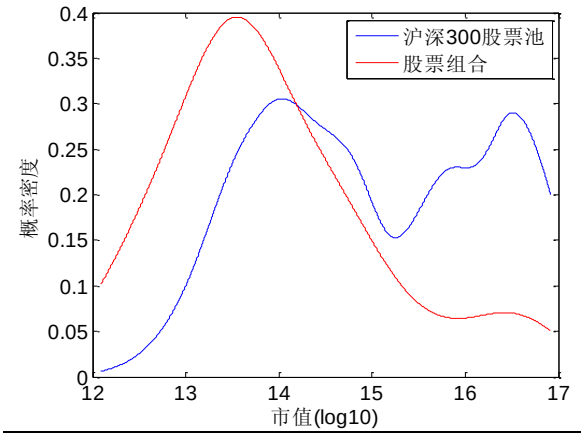
为了控制暴露在因子上的风险，我们提出一种因子中性策略。这里所说的因子中性是指组合在该因子上暴露的风险与指数基准股票池是一致的。在最理想的情况下，该因子不对组合贡献超额收益。结合前述内容对因子分布的研究，我们为因子中性策略设定一个目标函数，即使得**组合分布与基准分布之间的差异最小化**。为了实现这一目的，我们针对因子分布提出了一种权重循环调整的方法。

基于因子分布的权重循环调整方法

在给出权重调整方法之前，我们先给出一个简单的选股策略实例，在这个策略实例的基础上进行方法的演示。

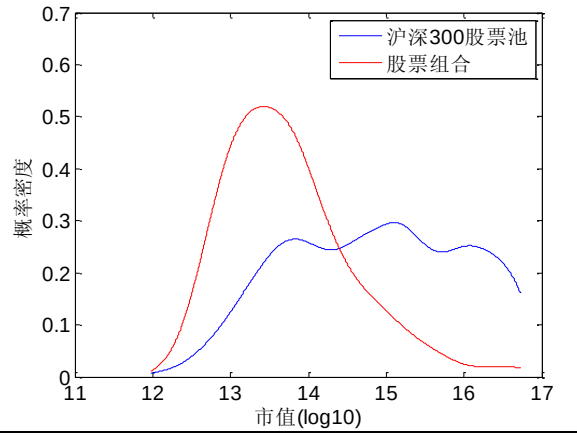
根据国信金工的区分度选股模型，我们使用估值因子市净率 P/B 在沪深 300 股票池内进行单因子选股。我们计算每天的 P/B 因子区分度，在市场偏好高 P/B 时选择高 P/B 的股票构成 Top 组合，在市场偏好低 P/B 的股票时选择低 P/B 的股票。该策略的具体实现可以参见国信金工报告《多因子 Alpha 选股－将行业轮动落实到 Top 组合》。该报告提出了一种多因子选股的方法，我们把它应用在单因子 P/B 上，在报告最后我们会对此策略进行模拟评价。现在我们的主要工作是，在通过 P/B 因子选出股票之后，我们希望该策略不要在市值因子上暴露过多的风险，因此我们希望达到市值中性的效果，使得组合的市值因子分布与沪深 300 的市值因子分布的差异尽可能小。该策略每次选出等权重的股票，我们都能得到当时的组合分布和基准分布。下面画出过去四年中四次选股组合的市值分布和基准分布的对比：

图 6：2008-03-20 市值加权基准分布和等权组合分布



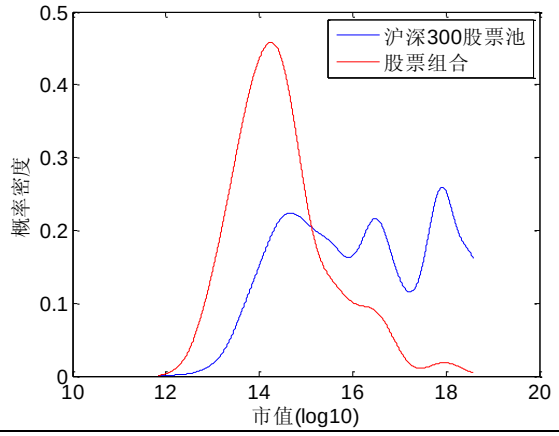
资料来源： WIND 资讯，国信证券经济研究所

图 7：2009-04-27 市值加权基准分布和等权组合分布



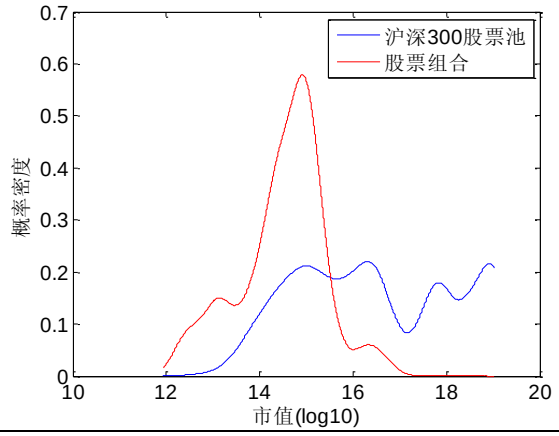
资料来源：WIND 资讯、国信证券经济研究所整理

图 8: 2010-10-18 市值加权基准分布和等权组合分布



资料来源： WIND 资讯，国信证券经济研究所

图 9: 2011-02-15 市值加权基准分布和等权组合分布



资料来源： WIND 资讯、国信证券经济研究所整理

这四个组合与基准分布的差异如下:

表 3: 策略组合分布与基准分布在调整前的差异

组合日期	组合分布与基准分布差异(单位:d0)	分布差异标准单位 d0
2008/3/20	64.93	0.11
2009/4/27	96.63	0.11
2010/10/18	113.93	0.11
2011/2/25	167.82	0.11

资料来源： Wind 科技，国信证券经济研究所

由上表可知这四个组合的市值分布与基准有巨大差异。等权组合的市值分布明显倾向于中小市值。我们现在使用权重循环调整方法进行权重调整，调整方法如下:

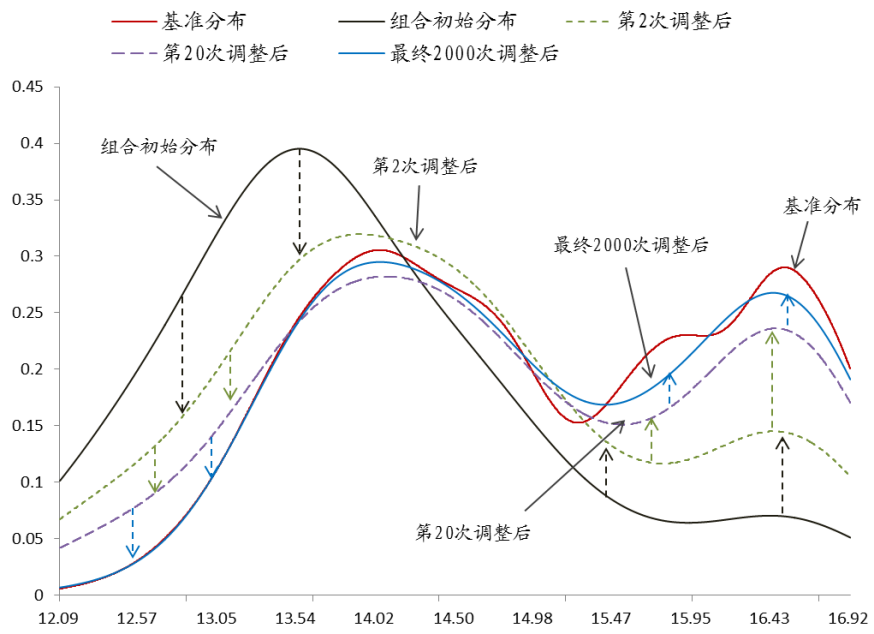
假设股票池有 n_0 支股票，现有 n_1 支股票构成策略选出的投资组合。首先用核函数方法估计因子在股票池中的概率密度分布 $fp_n(x)$，以及在组合中的分布 $f_n(x)$。对于第 i 只股票的因子值 x_i，对应着组合的概率密度 $f_n(x_i)$和股票池的概率密度 $fp_n(x_i)$。初始组合每支股票的权重都是 1，需要进行组合的权重调整，使其分布函数不断地接近股票池的分布函数。权重调整会循环进行 N 轮，每一轮的过程如下:

1） 从组合的第一支股票开始，对于第 i 支股票，如果 $f_n(x_i)$小于 $fp_n(x_i)$，则增加股票 i 的权重 w_i；如果 $f_n(x_i)$大于 $fp_n(x_i)$则减少股票 i 的权重。

2） 重新进行核密度估计得到新的概率密度分布 $f_{n+1}(x_i)$和 $fp_{n+1}(x_i)$，继续第 1 步，直到 n=N。

对于每次增加或者减少的权重，当 f(x)与 fp(x)相差越大时，调整的权重值应该越大，因此设置每次调整的权重值为 a*|f(x)-f(p)|，a 是一个实测的经验系数，目前取 a=0.03。循环次数 N 也是一个经验值，循环次数越多越好，但是需要与运行时间权衡，一般循环 100 到 2000 次。权重调整的情况如下图所示:

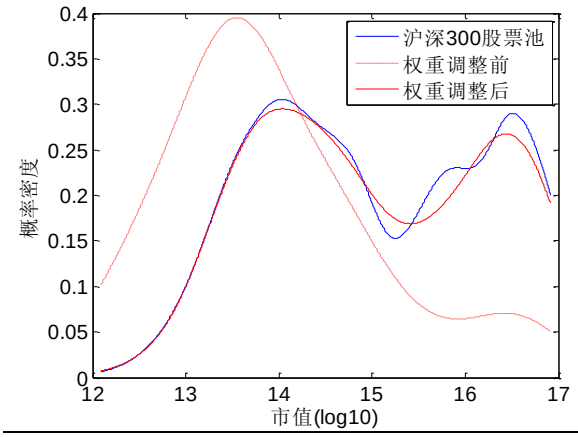
图 10：权重循环调整方法



资料来源：国信证券经济研究所

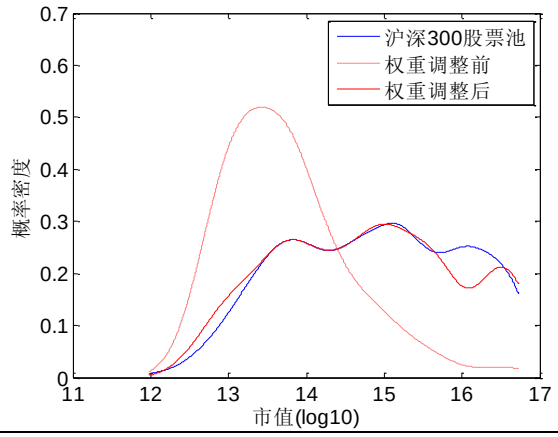
在进行权重循环调整之后，上面四个组合的组合分布变成如下图所示：

图 11：2008-03-20 组合权重调整前后分布



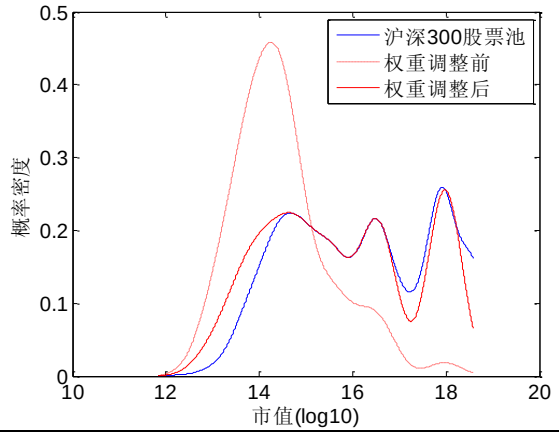
资料来源： WIND 资讯，国信证券经济研究所

图 12：2009-04-27 组合权重调整前后分布



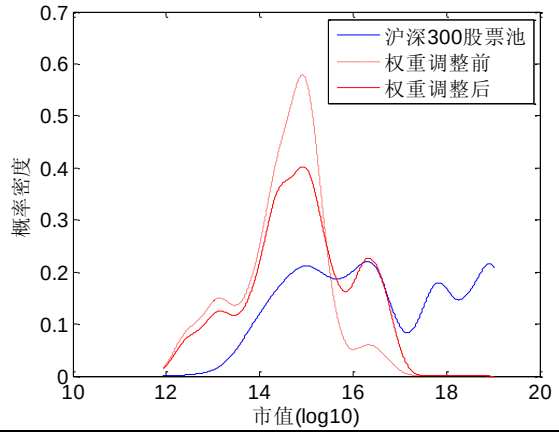
资料来源：WIND 资讯、国信证券经济研究所整理

图 13：2010-10-18 组合权重调整前后分布



资料来源： WIND 资讯，国信证券经济研究所

图 14：2011-02-15 组合权重调整前后分布



资料来源：WIND 资讯、国信证券经济研究所整理

从中可以看到，组合分布在经过权重调整之后不同程度的减少了和基准分布之间的差异，权重调整后的差异分布大小如下：

表 4：策略组合分布与基准分布在调整后的差异

组合日期	权重调整前(单位:d0)	权重调整后(单位:d0)	分布差异标准单位 d0
2008/3/20	64.93	0.31	0.11
2009/4/27	96.63	1.88	0.11
2010/10/18	113.93	8.14	0.11
2011/2/25	167.82	117.43	0.11

资料来源：Wind 科技，国信证券经济研究所

我们可以看到前三个组合调整权重的效果非常明显，全部将差距减少到原来的 1/10 甚至 1/100 以下，这是因为组合原有的股票较为均衡，经过权重调整之后可以较好地贴近基准分布。对于第四个组合，权重调整之后尽管差距减小了 30%左右，但是效果仍然很差，这是因为原有组合严重缺少大市值股票，原有组合分布曲线尾部几乎为 0，使得计算出的分布差距很大。对于这样的组合，权重调整不能解决组合和基准分布的差距，必须调整原有组合的构成，添加大市值股票。

此处我们以市值因子作为目标因子，由于沪深 300 指数本身是市值加权的，所以如果我们对组合做市值加权，也可以降低组合分布和基准分布的差异。但是市值加权是建立在我们已知基准的市值加权方式基础上的，对于其他因子，由于我们不知道基准的分布方式，不可能运用类似市值加权的方法进行权重调整，因此需要先估计因子分布，再进行分布拟合，也就是权重调整工作。适用于未知分布的不同因子，这是基于分布函数做权重循环调整的优势所在。

因子中性策略的效果检验

我们已经用权重循环调整的方法构建了基于分布函数的因子中性策略，现在需要进一步检验因子中性的效果。中性策略的检验是较为困难的事情。我们不能以业绩的提高作为一个衡量标准，而应该将业绩分解到单个因子。如果我们的策略对于某因子是完全中性的，那么业绩归因的结果应该会显示该因子没有贡献超额收益。因此我们可以用因子业绩归因的方法来衡量中性策略的效果。根据传统的归因理论，对于给定的收益区间 t:

$$r_P - r_B = \sum_{i=1}^{N}(w_i^P - w_i^B)r_i$$

$$= \sum_{i=1}^{N}(w_i^P - w_i^B)(\alpha_i + \beta_{i1}f_1 + \beta_{i2}f_2 + \cdots + \beta_{iK}f_K)$$

$$= \sum_{i=1}^{N}\widetilde{w_i}\,\alpha_i + \sum_{i=1}^{N}\widetilde{w_i}\,\beta_{i1}f_1 + \sum_{i=1}^{N}\widetilde{w_i}\,\beta_{i2}f_2 + \cdots + \sum_{i=1}^{N}\widetilde{w_i}\,\beta_{iK}f_K$$

其中r_P, r_B, r_i分别为收益区间内的组合收益率、基准收益率和第 i 只股票收益率，$w_i^P, w_i^B, \widetilde{w_i}$分别为第 i 只股票在组合中的权重、基准中的权重和两者之差，β_{ik}是第 i 只股票在第 k 个因子上的因子暴露，f_k 为第 k 个因子的因子溢价。上式将超额收益归因到各个不同因子中，例如第 k 个因子的收益贡献即为$\sum_{i=1}^{N}\widetilde{w_i}\,\beta_{ik}f_k$。

对于寻找f_k有多种方法，例如逐步回归、主成分回归以及非回归方法等。但是对于因子业绩归因而言，用回归方法找出特定因子的f_k是较为困难的事情。这是因为我们希望在锁定特定因子的情况下，能够在任意时间找到当时的因子溢价。但是回归方法常常不能得到显著的f_K。以逐步回归为例，当用大量因子进行逐步回归时，尽管我们可以筛选出当时显著的因子，但是并不能保证第 K 个因子是显著的，即很可能在归因时点某个特定因子并不显著，以至于无法对该因子归因。对于主成分回归而言，尽管主成分分析能够在主成分中保留大部分因子，有很大的可能性保留需要归因的因子，但是由于必须保留归因因子，主成分的回归常常不能得到显著的方程，拟合效果较差。

需要更进一步说明的是，对于我们的单因子中性策略而言，在不需要进行因子的横向比较的情况下，f_k的值并不重要。在同一时点，只要能减小$\sum_{i=1}^{N}\widetilde{w_i}\,\beta_{ik}$，中性策略就是有效的。因此只要能够找到持续有效而且一致的变量替代f_k作为公共因子即可。国信金工的多因子研究体系给出了因子区分度的定义，我们用因子区分度来替代f_k来进行业绩归因。

在之前的报告《数量化投资技术系列之二十六-多因子 Alpha 选股：将行业轮动落实到 Top 组合》中，我们提出了因子偏离度和因子贡献度的概念，并论述了两者之间的一致性。因此，我们选用因子贡献度作为因子对股票的区分度度量。

因子贡献度的含义如下：

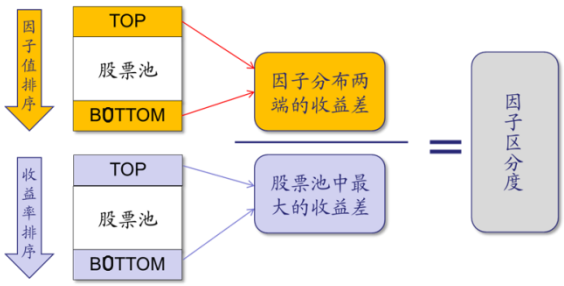
1.首先将股票池中的股票按因子进行排名，分别选出排名靠前的 20%和排名靠后的 20%股票构成两个组合；

2.我们将这两个组合各自的平均收益率相减，得到一个差值 D1；将股票池中所有股票中收益前 20%和后 20%的股票各自的平均收益率相减，得到第二个差值 D2；D1/D2 即为因子贡献度；

3.两个组合的收益率相差越大，则说明该时点此因子对股票的区分度越大；对于正的贡献度，因子值越大，平均收益率越小；对于负的贡献度，因子值越大，平均收益率越大。

因子区分度的概念如下图所示：

图 15：因子区分度（贡献度）的计算方法



资料来源：国信证券经济研究所

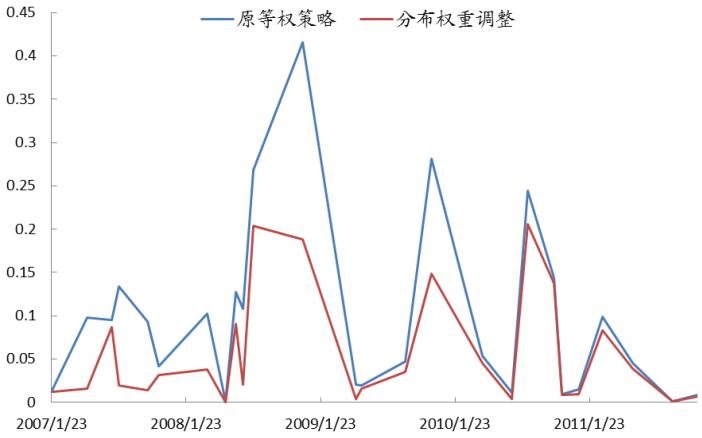
对于上文给出的 P/B 单因子策略，我们使用因子区分度对其市值因子的收益贡献进行归因，在归因中也对因子值进行了标准化。此外，我们将市值加权的结果也列出。在 2007 年 1 月至 2012 年 2 月的区间内，各个组合换仓时点的收益归因的绝对值如下：

表 5：市值因子对组合收益的贡献

日期	原等权策略	分布权重调整	市值加权
2007/1/23	0.0144	0.0119	0.0096
2007/4/30	0.0980	0.0160	0.0222
2007/7/4	0.0956	0.0866	0.0005
2007/7/25	0.1338	0.0193	0.0381
2007/10/11	0.0931	0.0140	0.0816
2007/11/9	0.0413	0.0319	0.0131
2008/3/20	0.1024	0.0382	0.0089
2008/5/9	0.0014	0.0007	0.0015
2008/6/5	0.1270	0.0901	0.0847
2008/6/26	0.1081	0.0204	0.1048
2008/7/24	0.2684	0.2040	0.1984
2008/12/3	0.4154	0.1883	0.2705
2009/4/27	0.0206	0.0037	0.0090
2009/5/13	0.0199	0.0164	0.0159
2009/9/8	0.0471	0.0349	0.0381
2009/11/19	0.2814	0.1488	0.4875
2010/4/7	0.0540	0.0453	0.0445
2010/6/25	0.0114	0.0041	0.0041
2010/8/6	0.2441	0.2057	0.1691
2010/10/18	0.1441	0.1376	0.1901
2010/11/8	0.0092	0.0083	0.0073
2010/12/24	0.0154	0.0096	0.0288
2011/2/25	0.0991	0.0834	0.0767
2011/5/19	0.0454	0.0389	0.0381
2011/9/5	0.0011	0.0010	0.0261
2011/11/10	0.0087	0.0069	0.0072

资料来源：Wind 科技，国信证券经济研究所

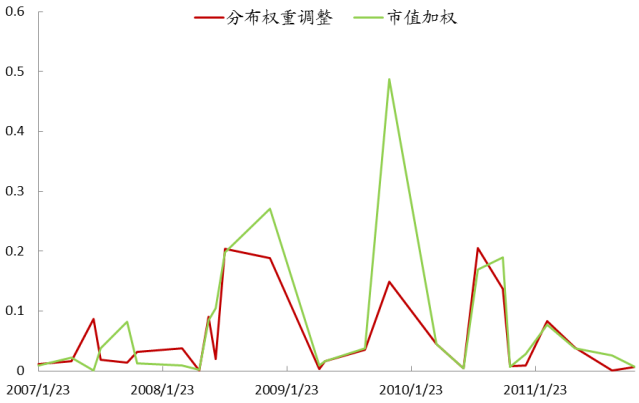
图 16: 原等权策略与分布函数下的权重循环调整后的市值因子贡献对比



资料来源：国信证券经济研究所

从中可以看到，每次调整组合后组合中市值因子的贡献明显小于权重调整之前。权重循环调整的中性策略明显达到了减少市值因子贡献的作用。为了对比市值加权方式的作用，我们将两者的对比作图如下，可以看到权重循环调整对市值加权也有一定的优势。

图 17: 市值加权策略与分布函数下的权重循环调整后的市值因子贡献对比



资料来源：Wind 科技，国信证券经济研究所

策略交易模拟与评价

上面我们对中性策略的中性效果进行了评价，下面对于权重调整后的策略进行一个模拟以完成完整的策略评价。我们假设用该策略构建一支指数基金，80%用于全复制，15%用策略选出的股票进行指数增强，5%持有现金。

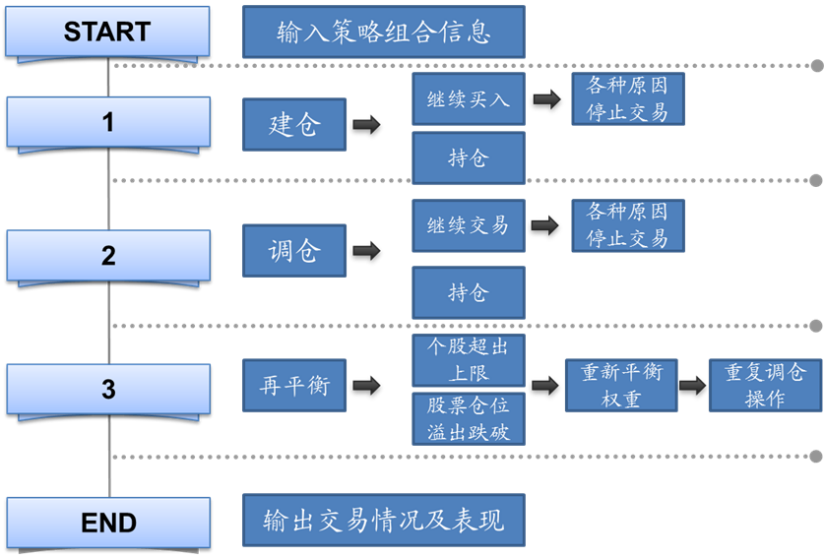
交易环境参数设置和流程示意图

1. 初始规模：10 亿
2. 交易费率：单边 1.5‰
3. 建仓、调仓交易期限：3 日
4. 再平衡触发天数：连续 3 日超出仓位上限或跌破仓位下限，触发再平衡
5. 最高成交比：20%
6. 股票仓位下限：90%，股票仓位上限：95%

7. 再平衡目标：超出上限，则降为最新仓位的 96%，低于下限，则升为最新仓位的 104%
8. 单一股票占基金总市值的上限：10%
9. 起始目标仓位：95%

流程如下：

图 18：模拟交易流程



资料来源：国信证券经济研究所

交易结果与业绩评价

交易结果如下：

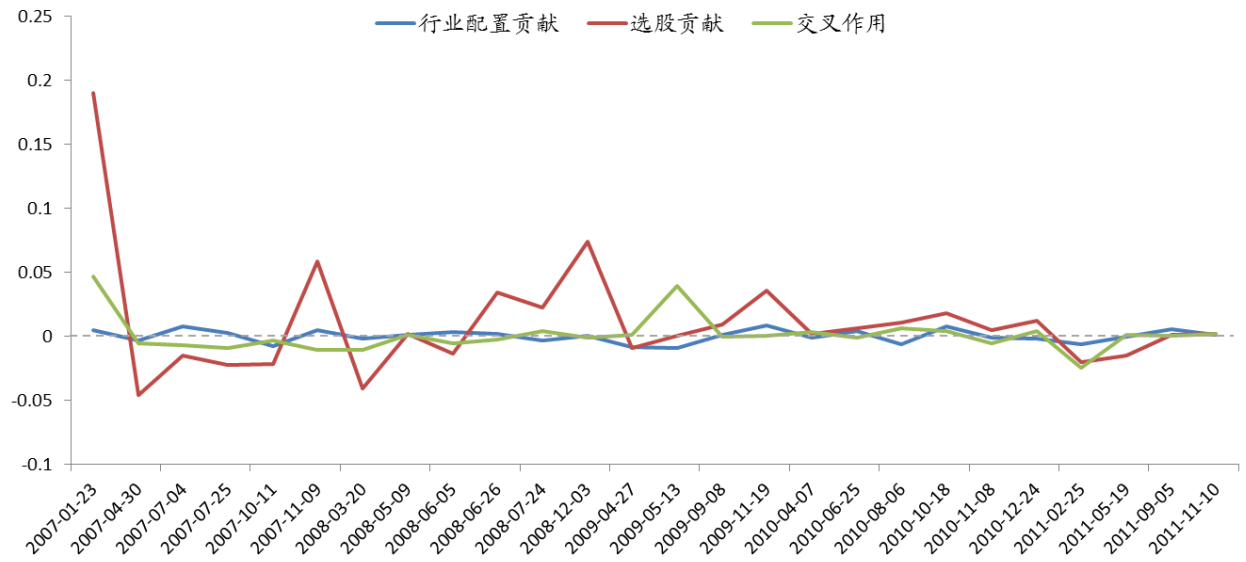
表 6：策略模拟业绩指标

指标	指标值
Beta	0.9589
Alpha	-0.00002
波动率	1.97%
年化波动率	31.25%
sharpe ratio	0.0119
年化 sharpe ratio	0.1891
基准 sharpe ratio	0.0130
基准年化 sharpe ratio	0.2064
Treynor 指数	0.0002
IR	-0.0135
年化 IR	-0.2135
TE	0.0023
年化 TE	0.0365
策略收益率	5.09%
基准收益率	7.26%
月度胜率	53.97%
最小换手率	38.59%
最大换手率	117.98%
平均换手率	76.51%

资料来源：天软科技，国信证券经济研究所

从业绩指标来看，这是一个表现一般的策略。我们进行业绩归因的结果如下：

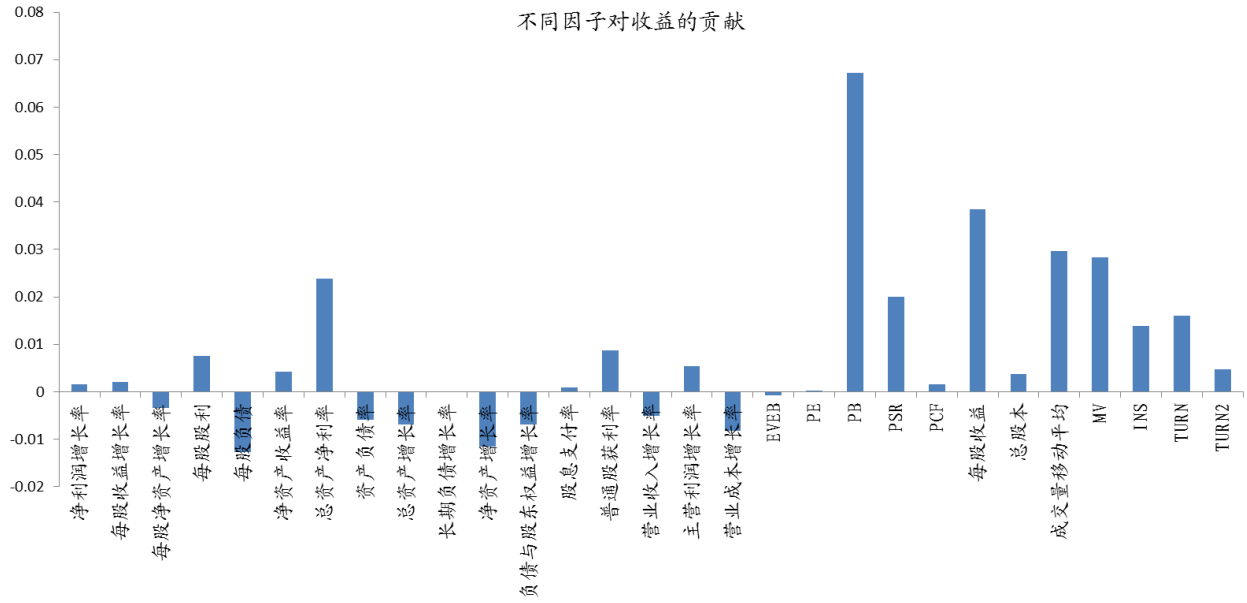
图 19：策略业绩 Brison 归因



资料来源：天软科技，国信证券经济研究所

可见选股的效果明显高于行业配置，我们进一步对因子进行收益分解，为了方便对比，我们给出了部分因子在整个区间内的收益贡献平均值，得到下图：

图 20：因子收益分解



资料来源：天软科技，国信证券经济研究所

我们从中可以看到，作为一个 P/B 单因子选股策略，P/B 因子贡献了最多的超额收益。图中显示市值因子 MV 也贡献了不少的超额收益。从整个策略区间来看，MV 的因子溢价（此处是因子区分度）要比 P/B 高很多，因此实际上策略在 P/B 因子上的暴露要比 MV 大很多。我们的策略在实现了获得 P/B 贡献的超额收益的目的。

在策略模拟中我们为了贴近实际操作，对交易成本、建仓、再平衡时间等做了

严格的设定，因此不可避免地会增大成本，对策略的收益造成了影响。

结语

我们在本报告中对因子的分布进行了一定的研究。使用核密度估计的非参数估计方法可以很好地得到因子分布的概率密度曲线，这种做法存在两方面的重要意义：

1. 对组合性质的分析评价。我们通过实例解释了如何使用分布函数来找出组合在某个因子上偏离基准的原因。这种方法可以让我们实现对因子风险的有效控制，而且这种方法不需要在事后对收益进行评估归因，在组合产生的时候就可以进行。

2. 因子中性和偏离策略的实现。通过权重调整的方法，我们可以构造出接近因子中性的组合。事实上这种方法并不仅仅限于中性策略，我们甚至可以构造因子偏离策略，使得组合在因子上的暴露通过权重的改变向我们所希望的方向偏离。换言之，我们实际上实现了控制因子分布的方法。

在实例展示过程中我们发现，对于某些股票组合，由于其在因子分布上存在极端差异（例如缺少大市值股票），因此这类组合不能通过权重调整的方法达到中性效果。为了能够更好的实现因子中性策略，我们可以在选股阶段对股票池进行分层选股，这样可以保证因子分布不会出现极端差异，在这种情况下进行权重调整，可以得到更好的中性效果。

因子的分布对于研究因子相关性有重要的作用。目前为止我们主要研究单因子的分布，多因子的边缘分布以及联合分布需要涉及到更复杂的方法。本报告只是研究因子分布和因子相关性的一系列报告的开始，我们会继续对多因子分布以及时间序列上的因子分布进行研究。后续工作有很大的难度，也有重要的应用价值，让我们拭目以待。

参考文献

[1]Entropy, divergence and distance measures with econometric applications, Aman Ullah, 1996

[2] Divergence measures based on the Shannon Entropy, Jianhua Lin, 1991

国信证券投资评级

类别	级别	定义
股票投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数±10%之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10%之间
	中性	预计 6 个月内，行业指数表现介于市场指数±5%之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

风险提示

证券投资咨询业务的说明

证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所团队成员

宏观

周炳林	0755-82130638
林松立	010-66026312
崔　嵘	021-60933159

固定收益

李怀定	021-60933152
侯慧娣	021-60875161
张　旭	010-66026340
赵　婧	021-60875168

策略

黄学军	021-60933142
林丽梅	021-60933157

技术分析

闫　莉	010-88005316

交通运输

郑　武	0755-82130422
陈建生	0755-82133766
岳　鑫	0755-82130432
周　俊	0755-82136085
糜怀清	021-60933167

银行

邱志承	021-60875167
黄　飙	0755-82133476

房地产

区瑞明	0755-82130678
黄道立	0755-82133397
方　焱	0755-82130648

商业贸易

孙菲菲	0755-82130722
常　伟	0755-82131528

汽车及零配件

左　涛	021-60933164

钢铁及新材料

郑　东	010-66025270
秦　波	010-66026317

机械

陈　玲	0755-82130646
杨　森	0755-82133343
后立尧	010-88005327

基础化工及石化

刘旭明	010-66025272
张栋梁	0755-82130532
吴琳琳	0755-82130833-1867
罗　洋	0755-82150633
朱振坤	010-66025229

医药

贺平鸽	0755-82133396
丁　丹	0755-82139908
杜佐远	0755-82130473
胡博新	0755-82133263
刘　勍	0755-82133400

电力设备与新能源

杨敬梅	021-60933160
张　弢	010-88005311

传媒

陈财茂	010-88005322
刘　明	010-88005319

有色金属

彭　波	0755-82133909
龙　飞	0755-82133920

电力及公共事业

谢达成	021-60933161

非银行金融

邵子钦	0755-82130468
田　良	0755-82130470
童成墩	0755-82130513

轻工

李世新	0755-82130565
邵　达	0755-82130706

家电

王念春	0755-82130407

建筑工程及建材

邱　波	0755-82133390
刘　萍	0755-82130678
马　彦	010-88005304

计算机及电子元器件

段迎晟	0755-82130761
高耀华	0755-88005321
欧阳仕华	0755-82151833

纺织服装

方军平	021-60933158

食品饮料

黄　茂	0755-82138922

新兴产业

陈　健	010-88005308

旅游

曾　光	0755-82150809
钟　潇	0755-82132098

数量化投资产品

焦　健	0755-82133928
周　琦	0755-82133568
邓　岳	0755-82150533

量化投资策略

董艺婷	021-60933155
郑　云	021-60875163
毛　甜	021-60933154
李荣兴	021-60933165
郑亚斌	021-60933150

量化交易策略与技术

戴　军	0755-82133129
黄志文	0755-82133928
秦国文	0755-82133528
张璐楠	0755-82130833-1379

基金评价与研究

杨　涛	0755-82133339
康　亢	010-66026337
李　腾	010-88005310
刘　洋	0755-82150566
潘小果	0755-82130843
蔡乐祥	0755-82130833-1368
钱　晶	0755-82130833-1367

数据与系统支持

赵斯尘	021-60875174
徐左乾	0755-82133090
袁　剑	0755-82139918

国信证券机构销售团队

华北区（机构销售一部）		华东区（机构销售二部）		华南区（机构销售三部）	
王立法	010-82252236 13910524551 wanglf@guosen.com.cn	盛建平	021-60875169 15821778133 shengjp@guosen.com.cn	魏　宁	0755-82133492-1277 13823515980 weining@guosen.com.cn
王晓健	010-82252615 13701099132 wangxj@guosen.com.cn	马小丹	021-60875172 13801832154 maxd@guosen.com.cn	邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn
焦　戬	010-82254209 13601094018 jiaojian@guosen.com.cn	郑　毅	021-60875171 13795229060 zhengyi@guosen.com.cn	段莉娟	0755-82130509 18675575010 duanlj@guosen.com.cn
李文英	010-88005334 13910793700 liwying@guosen.com.cn	黄胜蓝	021-60875166 13761873797 huangsl@guosen.com.cn	郑　灿	0755-82133043 13421837630 zhengcan@guosen.com.cn
原　祎	010-88005332 15910551936 yuanyi@guosen.com.cn	孔华强	021-60875170 13681669123 konghq@guosen.com.cn	王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn
赵海英	010-66025249 13810917275 zhaohy@guosen.com.cn	叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn	甘　墨	0755-82133456 15013851021 ganmo@guosen.com
甄　艺	010-66020272 18611847166 zhenyi@guosen.com.cn	崔鸿杰	021-60933166 13817738250 cuihj@guosen.com.cn	徐　冉	0755-82130655 13632580795 18022@guosen.com.cn
杨　柳	 18601241651 yangliu@guosen.com.cn	李　佩	021-60875173 13651693363 lipei@guosen.com.cn	颜小燕	0755-82133147 13590436977 yanxy@guosen.com.cn
		刘　塑	021-60875177 13817906789 liusu@guosen.com.cn	林　莉	0755-82133197 13824397011 linli2@guosen.com.cn
		汤静文	021-60875164 13636399097 tangjwen@guosen.com.cn	赵晓曦	82134356-1228 15999667170 zhaoxxi@guosen.com.cn
		梁轶聪	021-60873149 18601679992 liangyc@guosen.com.cn		