

证券研究报告



分析师:

徐寅

xuyinsh@xyzq.com.cn

S0190514070004

任瞳

rentong@xyzq.com.cn

S0190511080001

研究助理:

胡顺泰

hushuntai@xyzq.com.cn

### 报告关键点

我们基于机器学习中的 AdaBoost 算法, 深入挖掘传统因子的选股能力, 构建了非线性的多因子模型, 得到了兴业 Smart Alpha 动态选股模型, 并构建了相应的 SA 因子。该因子巧妙的利用了已有的传统因子, 极大地改善了它们效果, 为多因子选股模型的建立提供了一个有意义的新方向。

### 相关报告

《中证 1000 指数投资价值分析与投资策略》2015-06-08

《利用期权市场进行择时之二: 依据期权指标判断市场走势》2015-05-20

《“50 朋友圈” 交易策略之三: 分级基金与 50ETF 期权投资策略探索》2015-05-19

## 聪明的 Alpha, 机器觉醒!

2015 年 06 月 15 日

### 投资要点

- 在这篇报告里, 我们详细描述了新一代多因子选股模型——兴业 Smart Alpha (SA) 动态选股模型的基本构成。兴业 SA 动态选股模型以机器学习中的经典算法——AdaBoost 算法为核心基础, 通过动态选择结合当期有效因子的方式建立了具备一定自适应性且预测能力较强的 SA 因子。该因子 2006 年以来的 Rank IC 月平均值高达 9.16%, 多空组合年化夏普比率达到了 2.73。
- 进一步地, 我们深入分析了 SA 因子的日历效应, 通过使用相同月份的数据作为训练样本, 建立了考虑日历效应的 Smart Alpha 模型, 并构造了表现更为稳定的 SA\_12M\_5SM 因子, 其月 Rank IC 均值提高到 10.09%, 多空组合年化夏普比率更是高达 3.08!
- 最后, 基于 SA\_12M\_5SM 因子, 我们以中证 500 为基准构建了量化对冲选股策略。在考虑了交易成本的情形下, 该策略年化超额收益率达到了 16.31%, Sharpe 比率为 2.79, 而最大回撤仅为 4.93%。

## 目 录

1、引言.....	- 5 -
1.1、多因子选股体系.....	- 5 -
1.2、市场生存之道.....	- 6 -
2、揭开 Smart Alpha 的神秘面纱.....	- 7 -
2.1、机器学习简介.....	- 7 -
2.2、训练因子池.....	- 8 -
2.3、算法原理.....	- 9 -
2.4、算法对于传统因子的改善.....	- 10 -
2.5、模型关于参数的敏感性.....	- 12 -
2.6、玩转风格.....	- 14 -
3、SA 因子测试.....	- 15 -
3.1、因子 Rank IC.....	- 15 -
3.2、分位数组合测试.....	- 19 -
4、百尺竿头，更进一步.....	- 21 -
4.1、日历效应.....	- 21 -
4.2、SA_12M_5SM 因子 Rank IC.....	- 23 -
4.3、SA_12M_5SM 因子分位数组合测试.....	- 25 -
4.4、SA_12M_5SM 因子在不同规模股票池中的表现.....	- 27 -
5、基于 Smart Alpha 的选股策略.....	- 28 -
5.1、策略表现.....	- 28 -
5.2、策略对交易费率的敏感性.....	- 30 -
5.3、策略滞后成交表现.....	- 32 -
6、总结.....	- 33 -
附录 1：兴业因子池.....	- 34 -
附录 2：算法流程.....	- 36 -

图 1、多因子投资体系.....	- 5 -
图 2、有监督机器学习流程.....	- 8 -
图 3、Smart Alpha 模型流程图 .....	- 9 -
图 4、MACD 因子分位数组收益率和调整后因子分位数组收益率 .....	- 11 -
图 5、MACD 因子和调整后因子 Rank IC 和风险调整后 IC.....	- 11 -
图 6、MACD 因子和调整后因子 IC 的 12 个月平均 .....	- 12 -
图 7、不同弱分类器数量下 SA 因子的 Rank IC 和风险调整后 IC.....	- 13 -
图 8、不同训练样本长度下 SA 因子的 Rank IC 和风险调整后 IC.....	- 13 -
图 9、模型每期选择的前三个因子 .....	- 14 -
图 10、模型每期选择的风格.....	- 15 -
图 11、入选模型前三名因子的频数.....	- 15 -
图 12、SA 因子的 Rank IC (2006-1-25 至 2015-4-30) .....	- 16 -
图 13、SA 因子 IC 的衰减 (2006-1-25 至 2015-4-30) .....	- 16 -
图 14、SA 因子 的行业调整 Rank IC (2006-1-25 至 2015-4-30) .....	- 17 -
图 15、SA 因子的股票覆盖数和因子换手率 (2006-1-25 至 2015-4-30) .....	- 17 -
图 16、因子 Rank IC 平均值和风险调整后 IC 比较 (2006-1-25 至 2015-4-30) ..	- 18 -
图 17、SA 因子分位数组收益率和 Sharpe 比率 (2006-1-25 至 2015-4-30) -	19
图 18、SA 因子分位数组超额收益率和信息比率 (2006-1-25 至 2015-4-30) ..	20
图 19、SA 因子分位数组多空净值和多头净值 (2006-1-25 至 2015-4-30) -	20
图 20、SA 因子分位数组换手率 (2006-1-25 至 2015-4-30) .....	- 21 -
图 21、SA 因子 Rank IC 和多空组合收益率分月度统计 (2006-1-25 至 2015-4-30)	- 21 -
图 22、SA-5SM 因子 Rank IC 和多空组合收益率分月度统计 (2006-1-25 至 2015-4-30)	- 22 -
图 23、SA-12M-5SM 因子 Rank IC 和多空组合收益率分月度统计 (2006-1-25 至 2015-4-30) .....	- 22 -
图 24、SA-12M-5SM 因子的 Rank IC (2006-1-25 至 2015-4-30) .....	- 23 -
图 25、SA-12M-5SM 因子 IC 的衰减 (2006-1-25 至 2015-4-30) .....	- 23 -
图 26、SA-12M-5SM 因子 的行业调整 Rank IC (2006-1-25 至 2015-4-30) ..	- 24 -
图 27、SA-12M-5SM 因子的股票覆盖数和因子换手率 (2006-1-25 至 2015-4-30)	- 24 -
图 28、SA-12M-5SM 因子分位数组收益率和 Sharpe 比率 (2006-1-25 至 2015-4-30)	- 25 -
图 29、SA-12M-5SM 因子分位数组超额收益率和信息比率 (2006-1-25 至 2015-4-30) .....	- 26 -
图 30、SA-12M-5SM 因子分位数组多空净值和多头净值 (2006-1-25 至 2015-4-30)	- 26 -
图 31、SA-12M-5SM 因子分位数组换手率 (2006-1-25 至 2015-4-30) .....	- 27 -
图 32、SA-12M-5SM 因子在大盘股中的 Rank IC 和多空组合表现.....	- 27 -
图 33、SA-12M-5SM 因子在小盘股中的 Rank IC 和多空组合表现.....	- 28 -
图 34、SA-12M-5SM 因子策略净值 (2007-1-31 至 2015-4-30) .....	- 29 -
图 35、SA-12M-5SM 因子策略超额收益率和净值 (2007-1-31 至 2015-4-30) -	29
图 36、SA-12M-5SM 因子策略年度统计对比.....	- 30 -
图 37、SA-12M-5SM 因子策略换手率 (2007-1-31 至 2015-4-30) .....	- 30 -
图 38、策略多头在不同交易费率下的年化收益率和 Shape 比率 .....	- 31 -
图 39、策略在不同交易费率下的年化超额收益率和信息比率 .....	- 31 -

图 40、策略在不同交易费率下的多头净值和超额净值 .....	- 32 -
图 41、SA-12M-5SM 因子策略滞后成交超额收益率和净值( 2007-1-31 至 2015-4-30 ) .....	- 33 -
表 1、SA 因子 Rank IC 统计数据.....	- 16 -
表 2、SA 因子十分位数组表现统计.....	- 19 -
表 3、SA-12M-5SM 因子 Rank IC 统计数据.....	- 23 -
表 4、SA-12M-5SM 因子十分位数组表现统计.....	- 25 -
表 5、SA-12M-5SM 因子在不同股票池中 Rank IC 和多空组合表现.....	- 27 -
表 6、SA-12M-5SM 因子策略统计数据.....	- 28 -
表 7、SA-12M-5SM 因子策略分年度统计 .....	- 29 -
表 8、SA-12M-5SM 因子策略滞后成交统计数据.....	- 32 -
表 9、兴业因子池.....	- 34 -

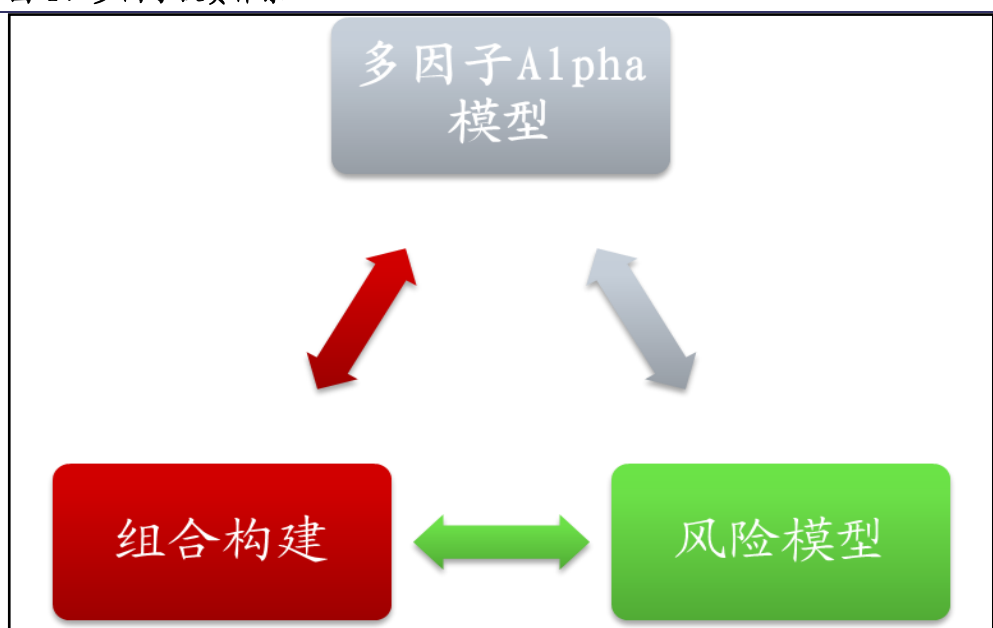
## 1、引言

### 1.1、多因子选股体系

众所周知，alpha 因子、风险模型与组合优化是现代多因子量化投资体系不可或缺的三个重要组成部分。三十多年来，多因子量化投资体系不断丰富发展。

“铁三角组合”中的风险模型和组合优化器先后都有了相对成熟的商业解决方案，并在投资实践中形成了一定的标准(如:Barra 的风险模型、Axioma 的优化器等)，其背后相对完善、精确的数量金融理论背景使得这两个部分的应用日趋成熟。而以 alpha 模型为代表的收益率预测模型，则一直是投资者所追寻的圣杯。新颖有效的 alpha 因子、稳定且具备较强适应能力的多因子模型是一个量化投资组合持续良好表现的根本保证。

图 1、多因子投资体系



资料来源：兴业证券研究所

虽然 alpha 模型十分重要，但他的构建却是一件仁者见仁智者见智的事情。模型中哪些 alpha 因子能够入选，每个 alpha 因子的权重应如何配置，这些问题到目前为止都还没有一套完善的解决方案。不仅如此，随着市场中主动量化投资主体的增加，特别是传统线性静态多因子模型的广泛运用，许多策略的超额收益近年来不断下降，而波动性则有明显上升的趋势。

是什么让传统多因子模型的表现逐渐平庸化？是 A 股市场已经十分有效了吗？我们认为，中国的股票市场远没有达到欧美股市的有效程度，但却呈现出了两个重要特征：第一，市场波动水平较高，风格切换频繁，任何一种策略的持续性都

比较差；第二，投资主体表现出高度竞争的博弈状态，市场及存在于其间的投资者从数据和自身的经验中不断的学习，因此昨天的市场“异象”（或无效点），很可能会成为今天的随机噪声，今天的随机噪声也可能会成为明天的市场“异象”！那么如何能保持投资策略超额收益的持续稳定呢？对于量化多因子 alpha 模型来说，主要有两条途径。一条就是不断发现新的、市场还未充分反映的 alpha 因子。这些因子一般会在发现后的一个较长时间内保持稳定的选股能力。我们在以前的猎金系列报告中就曾沿着这条路向前探索，也发掘了不少在中国市场上有效的选股因子。然而新因子的发掘更多的源自于灵感、直觉和想象力，是一件艺术性多于科学性的工作。另一条路则是更加充分地应用传统因子。在之前的研究中，我们曾深入分析过传统因子的最佳使用场合以及因子之间的相互关系，提出过情景模型的想法。而现在呈现在您面前的这篇报告则是我们在因子轮动方面的一次尝试：基于传统因子池，我们借助机器学习的方法，动态挑选并组合每一期的有效的因子，建立**兴业 Smart Alpha 动态选股模型**。希望通过风格轮动来深度发挥传统因子的效力，让它们重焕活力，为我所用。

## 1.2、市场生存之道

通常来说，量化投资依赖于从历史上学习到的某种模式，而假定这种模式可以在未来持续有效。但不幸的是，变化是市场永恒的主题。08、09 年全球金融危机的余烬熄灭后，经过了长达 5 年的漫漫熊市，而今天的 A 股市场如日中天。谁能预见下一个危机的降临？谁又敢断言 A 股市场已是强弩之末？在这个风云诡谲，弱肉强食的残酷市场里，我们该如何生存？

**“生存还是毁灭，这是一个值得思考的问题！默然忍受命运暴虐的毒箭，或是挺身反抗人世无涯的苦难，通过斗争把它们清扫，这两种行为，哪一种更高贵？”**

——《哈姆雷特》 莎士比亚

对于这个问题，我们毅然决然的选择勇敢面对。但环境是恶劣的，敌人是强大的，我们不能赤手空拳，盲目上阵。我们需要强大的武器，坚实的防具，有效的战术。只有具备了这些，我们才能在这个快速变化的市场里游刃有余，发展壮大。

作为一名量化投资者，什么是我们的武器？很显然，在历史熔炉中淬炼出来的 alpha 因子就是我们的刀枪剑炮。但单独的 alpha 因子有效性难以持久，波动性较高。为了结合这些形形色色的 alpha 因子以得到一个威力更大，性能更稳定的选股策略，各种多因子模型被不断研发出来。依据是否动态调整因子，可以将多因子模型划分为静态和动态模型。

静态模型相对简单，首先选出认为比较有效的 alpha 因子形成因子池，然后



为每个因子赋予一定的权重，加总得到最终的综合因子作为选股的根据。静态模型一个最大的优势就是简单易行，但它的缺点也同样明显。正如我们所说，市场是变化的，风格是轮动的，而静态模型过于僵化，不能根据市场的当前状况调整自己的风格，因而难以发挥各个因子的优势。其次，静态模型需要大量的历史数据来观测因子的有效性和收益模式，对于刚刚走过 25 个年头的 A 股市场，历史数据并不充分，很多检验也不可靠。另外，静态模型最核心的问题是因子权重的设置，许多权重的设定都是依赖于经验和主观的偏好，这无疑是一件难以窥探的独门武功。鉴于静态模型的种种不足，我们把目光投向更加灵活的动态模型。下面的文章中也比较了我们的模型和简单等权静态模型的选股能力。

相比于静态模型，动态模型的种类和衍生更加丰富和微妙。动态模型根据市场条件的变化，动态的选择因子和调整权重，以期踏准市场节奏。动态模型的构建有很多方法，有一些也是依赖于投资者的主观判断和逻辑推理，这类模型不在我们的讨论之列。另一些模型则是依赖于数学、统计、计算机等量化工具，依据某种机制，构建出一个具有自适应和自动学习特征的模型。其中，机器学习的方法被广泛应用于此类模型的构建。我们的 Smart Alpha 模型正是基于一种称为 AdaBoost 算法的思想而构建的。

说到机器学习，很多人不禁有各种质疑。有人说机器学习是一个黑盒，很难从逻辑和经济金融原理上解释。又有人说机器学习依赖于很多参数，往往有过度优化的问题，样本外和样本内的表现也会大相径庭。我们承认，很多机器学习模型都有这些方面的问题。但是，我们在本文中选择的 AdaBoost 算法在这两方面却有其独到之处。首先，模型的机制较为清晰。我们是用过去一段时间的数据作为样本，选择选股区分度最好以及彼此相关性不高、互为补充的因子进行结合，同时兼顾收益和风险两个方面，并非完全不透明的黑箱。其次，相比于其他的机器学习算法，我们的模型参数较少，且参数的敏感性很低，这在下文中有详细的测试结果。第三，我们的模型对数据的噪声也不敏感，弹性比较大，可以将各种限制添加进去，比如我们下文中为了规避日历效应而对模型的改进。另外，我们模型的适用性较广，基本面投资者可以将其选股结果作为初始池来提高胜率，量化投资者既可以把它作为单独的 alpha 来源也可把它嵌入到传统的多因子模型中使用。下面我们将详细介绍兴业 Smart Alpha 动态选股模型。

## 2、揭开 Smart Alpha 的神秘面纱

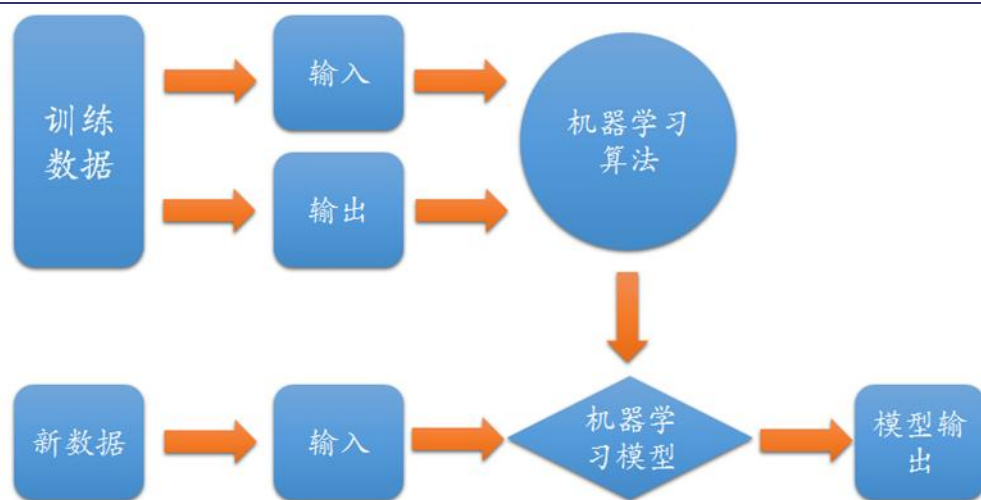
### 2.1、机器学习简介

机器学习是计算机科学的一个分支，通过算法的设计，使得计算机具有一定的人工智能特性，以完成更为复杂的任务，例如：模式识别、系统控制、推断和预测等等。机器学习算法基于已获得的经验数据，深度发掘数据的内在特征以及

不同观察变量间的隐秘关系，自动识别规律性的模式，据此作出合理的判断。

有监督的机器学习是机器学习的一个重要组成，其往往用于带有预测性的任务，例如：手写识别、语音识别、人脸检测、姿态估计等等。在本报告中，我们将有监督的机器学习算法用于预测股票的未来收益。有监督的学习基于一组已观测的历史样本，每个样本由一些输入变量和一个输出变量组成。一个有监督学习算法分析这些训练样本，得到输入和输出之间的一个推断函数。当有新的样本外输入发生，模型根据推断函数生成预测输出。有监督机器学习流程参见下图。

图 2、有监督机器学习流程



来源：兴业证券研究所

在本报告中，我们以因子数据作为模型的输入，股票的收益作为输出。根据历史的因子数据和股票收益作为训练样本，构建当前的因子值和下期股票收益之间的推断函数。股票市场是一个错综复杂的、由人类活动主导的环境，其内部很难挖掘出像自然环境中的物理定律一样恒定简单的规律。因而将擅于处理复杂关系、海量数据的机器学习方法用之于股票投资是一个值得考虑的研究和应用方向。近来，随着数据获得性的改良和计算机技术的革新，基于机器学习模型的选股策略在量化金融中变得越来越热门。中国的股票市场方兴未艾，最近更是欣欣向荣，将机器学习用于 A 股市场选股也成为一种可能。

## 2.2、训练因子池

我们长期关注和跟踪的因子有价值、成长、质量、动量、情绪、技术共 6 大类 65 个因子。我们以此为基础因子池，选择一个合适的子因子池作为下文中机器学习算法的数据来源。所有因子的定义请见附录。

考虑到因子覆盖度和有效性的问题，我们最终选择了附录兴业因子池表中标黄的 33 个因子作为我们机器学习的数据源。我们主要剔除了含有预期数据的因子，



因为其覆盖的股票数量较少，且时间较短，而这会影响机器学习算法的准确性。我们也剔除了像 LnFloatCap、TurnoverAvg-1M 等本身有效性很好或者市值偏斜较大的因子。

### 2.3、算法原理

我们将选股模型化归为一个二元分类问题。根据下期股票收益的大小，我们将股票池划分为两类：强势股和弱势股。强势股是下期股票收益居前的股票，而弱势股即是收益落后的股票。我们将要构建的分类器就是要区分某只股票属于强势股还是弱势股。算法流程参见下图。

图 3、Smart Alpha 模型流程图



来源：兴业证券研究所

我们依据被称为 AdaBoost 的算法构建分类器。AdaBoost 算法的基本思想是它自适应的构建一系列弱分类器，每一层弱分类器只关注上层弱分类器分类错误的样本，并对这些样本尽最大可能的区分，最终将这些弱分类器结合起来提升整体分类的效果。

分类器的输入是我们上一节选中的 33 个传统因子的当期因子值，而输出是每只股票的信心分数，它用于衡量该股票属于强势股的程度。一般来说，该分数大于 0 表示其更可能属于强势股，且值越大入选的可能性越高；而小于 0 则意味着其倾向于被划分为弱势股。

#### ● 训练样本集

首先我们要准备训练样本集作为构建分类器的基础。大多数的机器学习算法对于输入数据反应都比较敏感。如果训练数据噪声较多，一致性不好，那么算法极有可能陷入过度优化的泥沼。所以说数据准备对于机器学习是至关重要的一步，在某种意义上，其甚至超过了对学习算法的选择。

我们使用横截面上因子的序作为输入，而不是简单使用原始因子值。因为选股是比较股票的相对优劣，而并非关注具体数值的大小。另外，使用因子值的序也可以有效的避免极端值的影响。在每个时间点上，对于每个因子，我们计算每只股票按照这个因子排列的序，并除以总的股票数，使得所有因子的值都划归到(0,1]之间。

接下来，对于收益率数据我们也要做一些处理。我们将下期股票收益率从大到小排列，取前 30%的股票作为强势股，后 30%作为弱势股，中间的股票视作噪声弃之不用。强势股我们标记为+1，弱势股我们标记为-1。这样每个样本点可以记作 $(\mathbf{x}, y)$ ，其中 $\mathbf{x}=(x_1, K, x_k, K, x_K)$ 是股票在各个因子上的序化的暴露， $y=\pm 1$ 。

为了保证算法的稳定性，我们用过去 12 个月月底的面板数据来构建训练样本。

#### ● 弱分类器

每个弱分类器可以简单的被认为是一个因子。对于每个因子，我们按照其分位数将股票划分为若干组，每一组内我们分别计算强势股和弱势股的权重总和。直观上讲，区分能力最有效的因子应该是强势股和弱势股权重相差较大的因子。比较所有的因子，我们选择区分能力最强的作为该层的弱分类器。这样我们就用该因子每个组里强势股和弱势股权重孰大作为区分的标志。从而当新的因子值落入某个分位数组里时，如果该分位数组里强势股权重大，我们便将该股票化为强势股，反之则为弱势股。

开始时我们等权配置各个数据，在每次找到新的弱分类器以后，我们将调整数据的权重。每个被正确区分的数据权重将被调低，而区分错误的权重将加大。下一层的分类器将基于该调整的权重进行构建。这样每次我们都将重心转移到分类错误的股票上，最终将会慢慢改善整体的分类准确度。

#### ● 强分类器

最后，我们将所有弱分类器进行简单的加总得到最终的强分类器。分类器会赋给每一组因子值一个实数作为信心分数，用以衡量其属于强势股的程度，值越大越可能为强势股，越小越可能为弱势股。本质上，我们可以把这个分数作为一个复合因子的因子值。对于由最终强分类器生成的因子我们称为 SA(Smart Alpha)因子。详细的算法流程请参见附录。

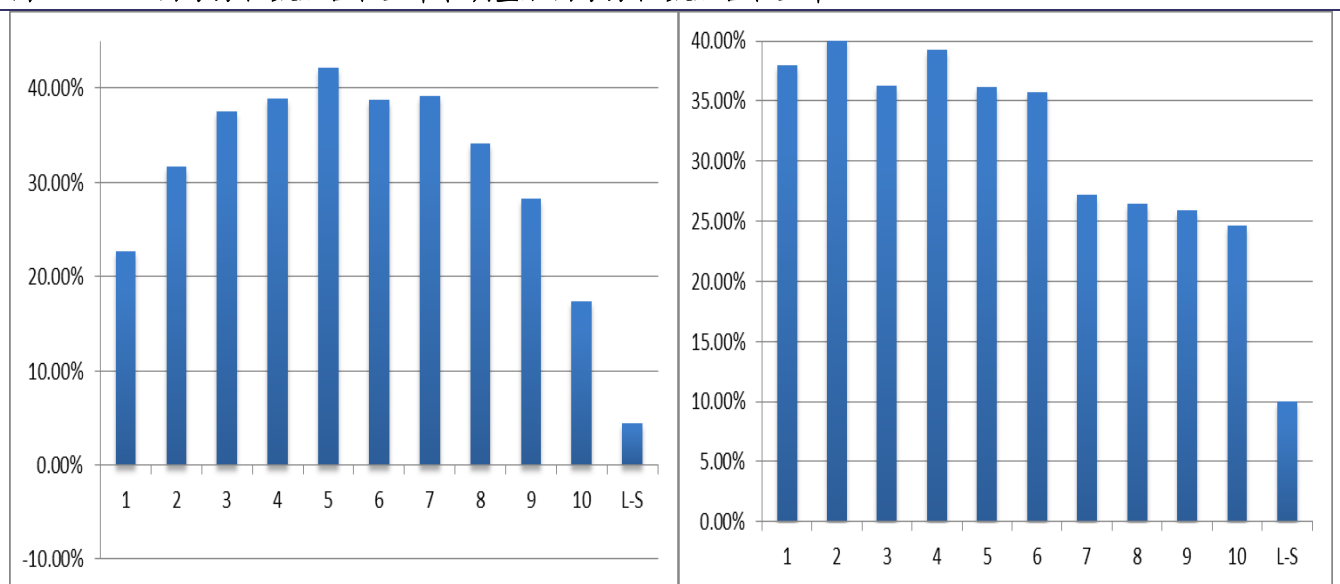
## 2.4、算法对于传统因子的改善

从上面的描述我们可以看出，算法中每一层的弱分类器本质上是把原本线性区分度不强的因子进行调整，把其非线性的区分能力进行了线性化。这里我们以兴业技术指标因子池中的 MACD 因子为例说明(技术指标因子的收益模式往往具有

非线性的特征)。

首先我们考察 MACD 因子的原始表现，然后仅以该因子作为训练样本，AdaBoost 算法中的弱分类器直接作为最终的强分类器使用，以其信心分数作为改良的 MACD 因子。下面是两者分位数组测试的结果。

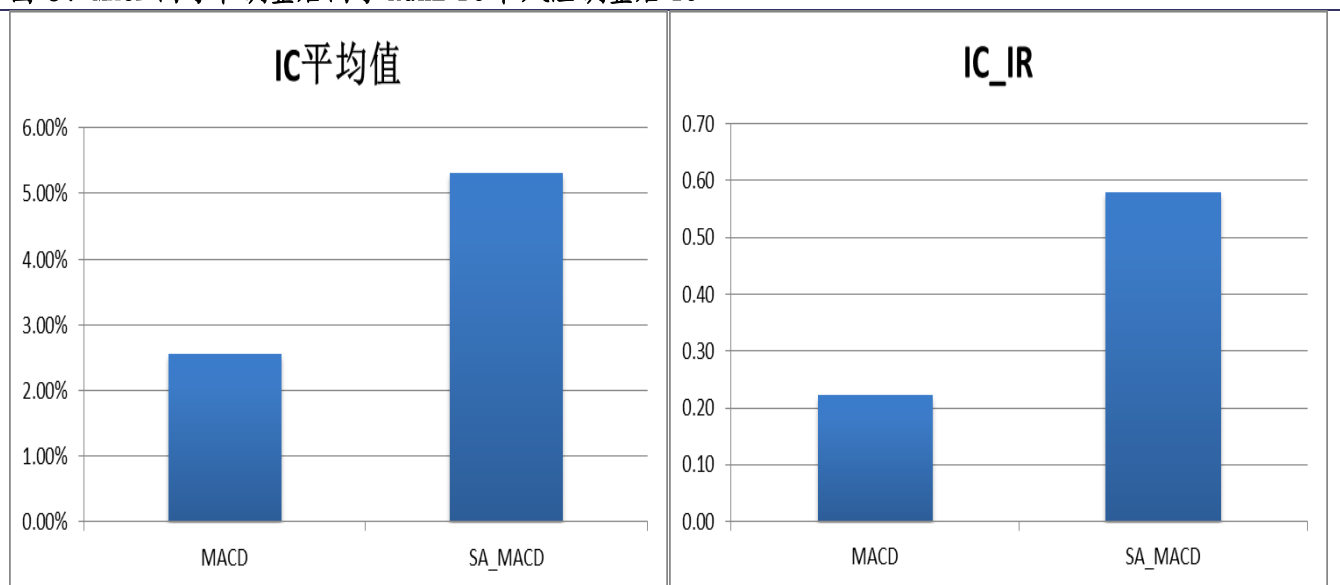
图 4、MACD 因子分位数组收益率和调整后因子分位数组收益率



资料来源：兴业证券研究所

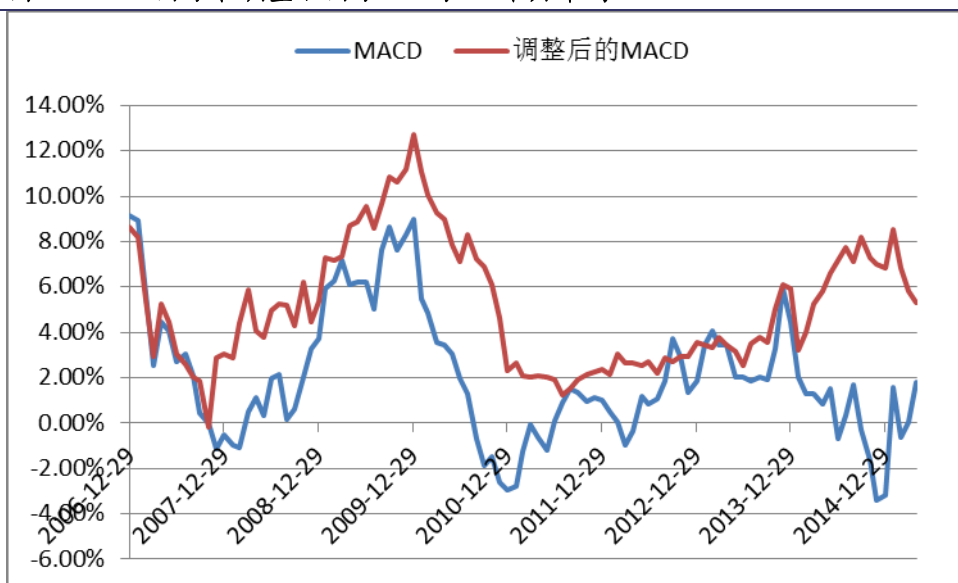
从上图可以看出，原始的 MACD 子具有明显的非线性形式，中间组合收益率高于第一组和第十组。而经过 AdaBoost 算法调整后，该因子非线性性得到了减弱，呈现了线性因子的特征。调整前后 MACD 因子的 IC 表现见下图。

图 5、MACD 因子和调整因子 Rank IC 和风险调整后 IC



资料来源：兴业证券研究所

图 6、MACD 因子和调整后的因子 IC 的 12 个月平均



资料来源：兴业证券研究所

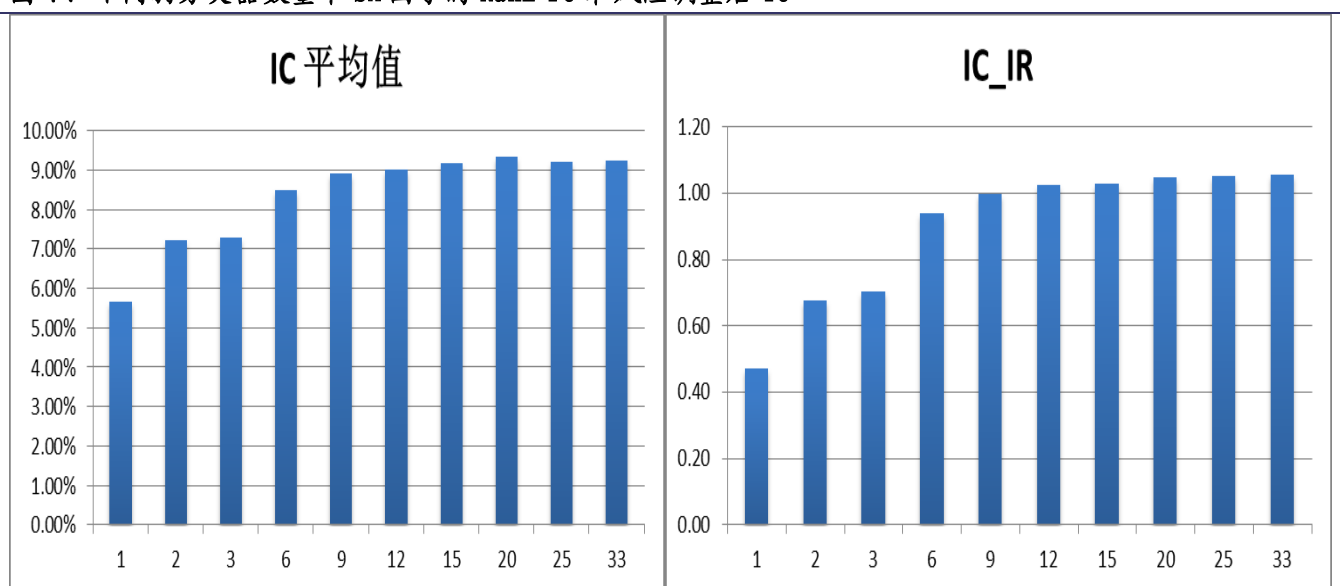
## 2.5、模型关于参数的敏感性

细心的读者可以发现，我们的 Smart Alpha 模型中有一些重要的控制参数。一般来说，一个稳健的模型不能对其中的参数过于敏感。下面我们对模型中的参数进行分析，首先我们观察用于构建最终强分类器的弱分类器数量，这是模型的一个关键参数。

直观上看，弱分类器的数量越多，模型应该更稳定、更有效。在训练样本内，这一点是肯定的。但一般来说，在训练样本内过度优化会导致样本外模型失效，这也是人们诟病机器学习和数据挖掘的一个主要问题。我们不禁要问，对我们的模型，是不是也存在这个问题呢？

下面两幅图展示了不同弱分类器数量下 SA 因子的 Rank IC 和风险调整后 IC 的表现。实证发现，模型的表现随着弱分类器的增加逐渐提高，并且在弱分类器数量较少的时候提高尤为明显，而后逐渐减弱，最终稳定在某个水平。这是因为最初选择的因子是表现最好的因子，而后的因子相关性也在上升，所以边际效应逐渐减弱。模型的表现关于弱分类器个数单调递增的特点说明我们的模型不容易产生过度优化的问题，甚至弱分类器个数与因子数量相等模型的表现依然出色。权衡了计算的成本和模型的表现后，我们选择 20 个弱分类器构建强分类器。

图 7、不同弱分类器数量下 SA 因子的 Rank IC 和风险调整后 IC

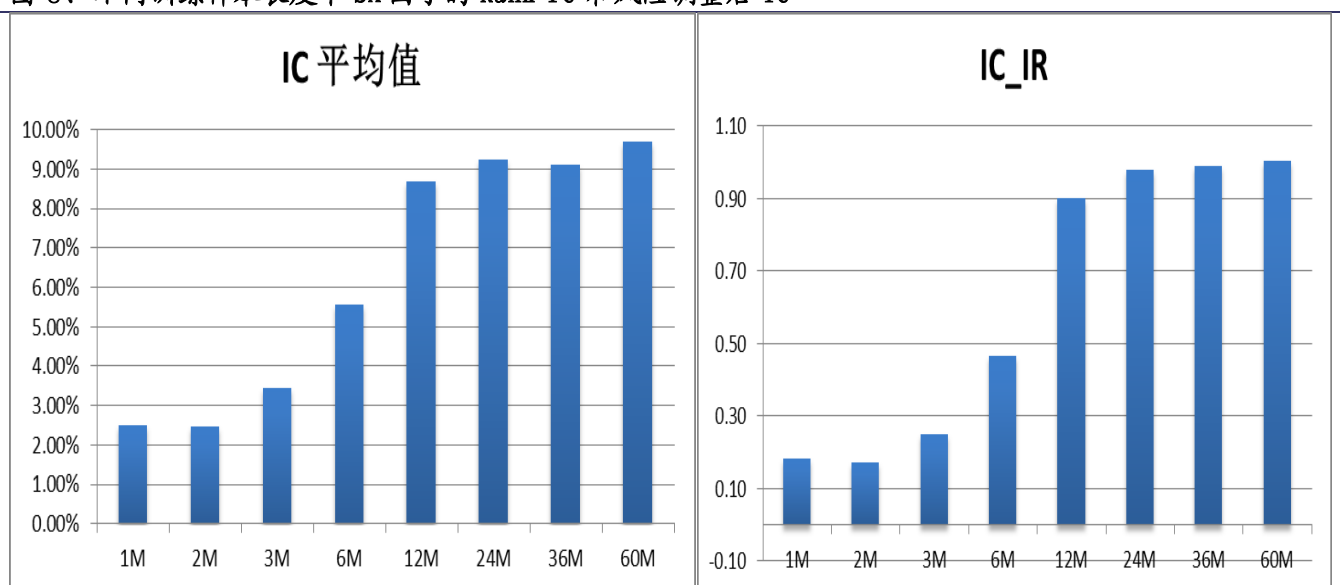


资料来源：兴业证券研究所

接着我们观察训练样本的长度对模型表现的影响。是不是训练样本越多，表现越好呢？一方面，较长的训练样本有更充足的数据，这会调高模型的准确度；但另一方面，过长的训练样本会引入大量过时的数据，这可能与当前的市场环境不符，反而会降低模型在样本外的适应能力。

下面两幅图展示了不同训练样本长度下 SA 因子的 Rank IC 和风险调整后 IC 的表现。我们发现，模型的表现随着训练样本长度的增加逐渐提高，但在 12 个月的训练长度以后边际改善逐渐减弱，最终稳定在某个水平。我们最终选择了 12 个月作为模型训练样本长度。

图 8、不同训练样本长度下 SA 因子的 Rank IC 和风险调整后 IC



资料来源：兴业证券研究所

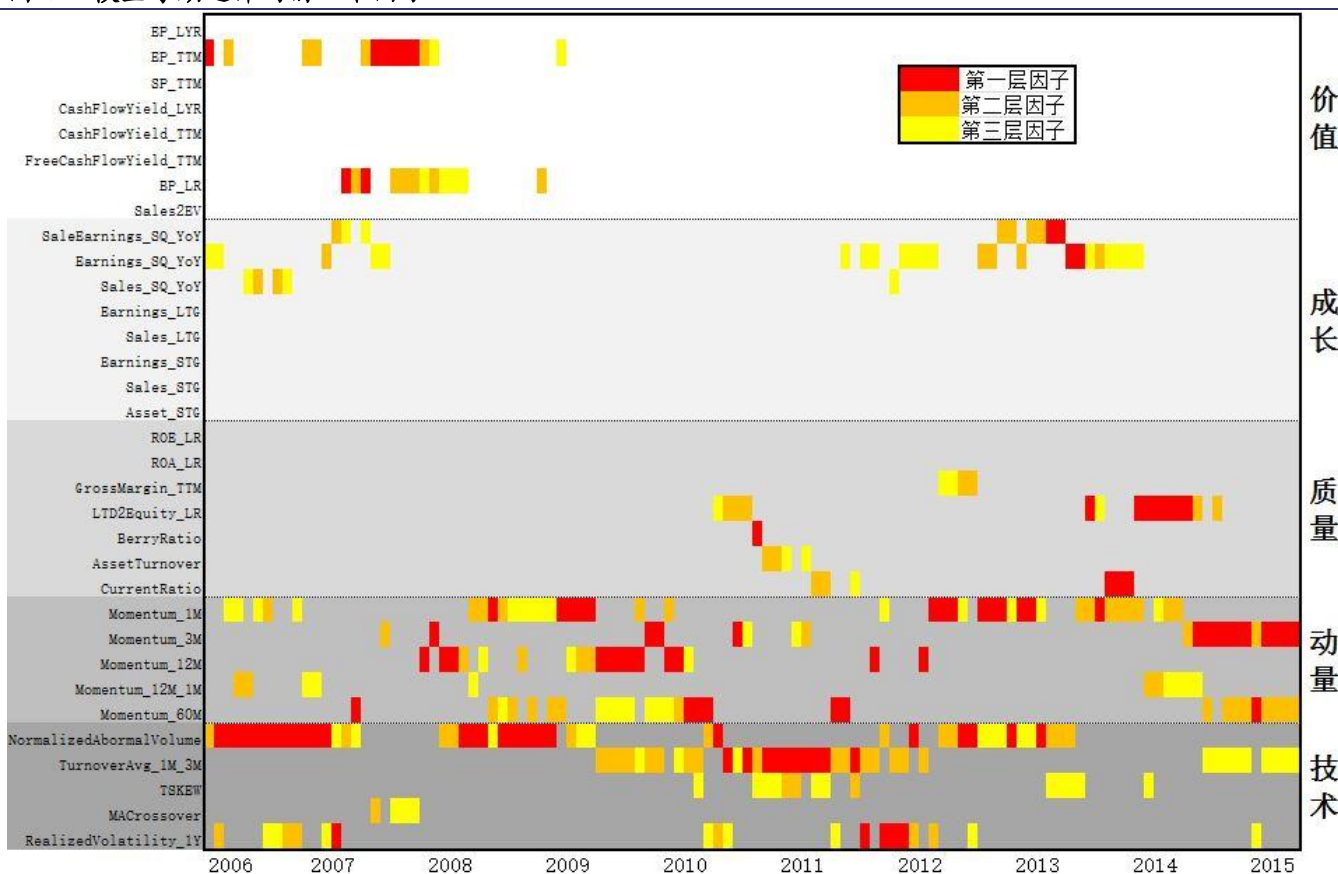


## 2.6、玩转风格

让我们深入观察一下模型对于因子的选择。下图展示了每一期模型所选择的前三个因子。可以发现入选因子多数情况下随着时间不断改变，且选中的前三个因子通常属于不同的风格类别，这也从结果的角度证明了我们的 Smart Alpha 模型倾向于选择彼此相关性较低的信息来构建 SA 因子。

从最终的选择结果来看，价值类因子在 2009 年以前常常出现在我们的模型中，成长类因子是 12 年到 14 年的“常客”，动量与技术类因子表现出了长期的有效性，而财务杠杆因子则从 13 年开始频繁进入我们的模型。以上的观察，应该说与我们对大类风格有效性的历史观察大体上保持了一致。我们认为，这一方面说明了 Smart Alpha 模型在因子选择上的有效性，同时也从一个侧面表明，只要谨慎地选择方法，基于机器学习算法的选股模型依然是可以理解的。

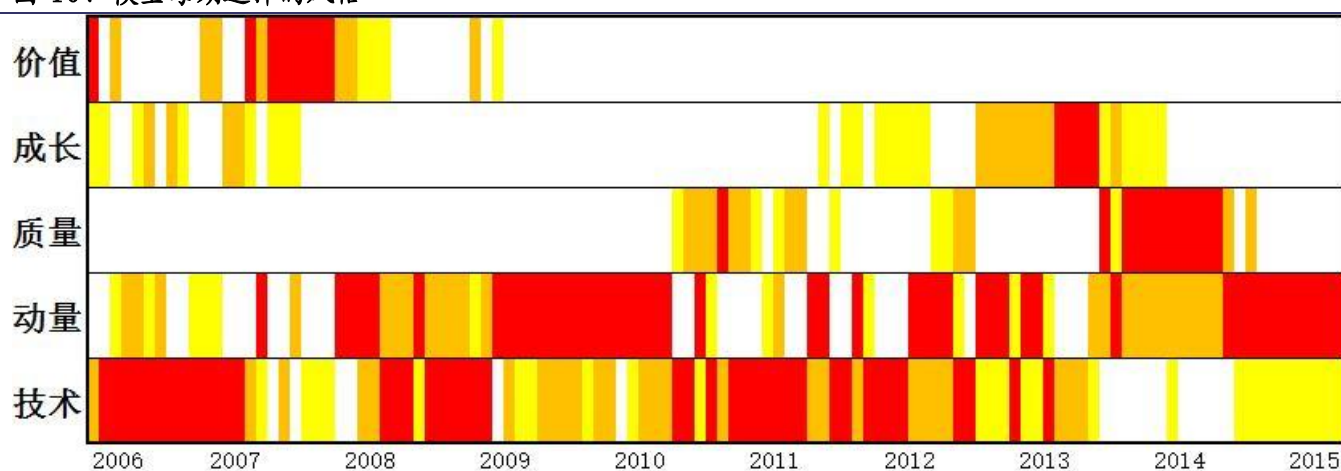
图 9、模型每期选择的前三个因子



资料来源：兴业证券研究所

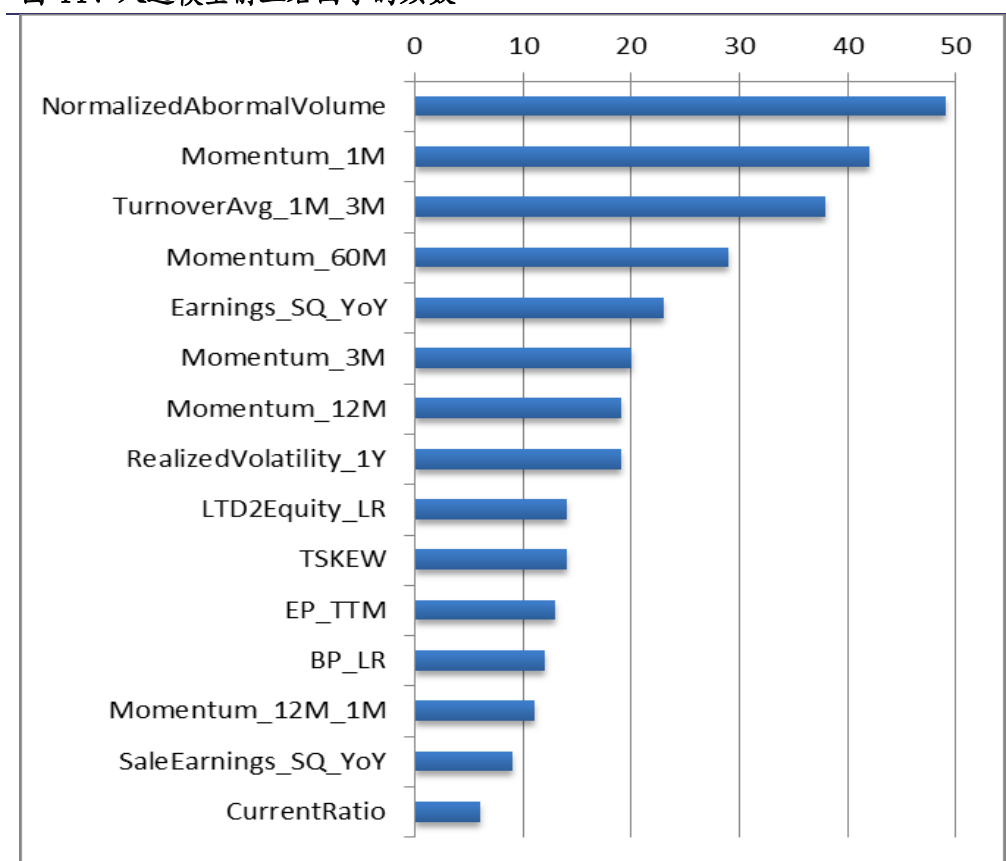


图 10、模型每期选择的风格



资料来源：兴业证券研究所

图 11、入选模型前三名因子的频数



资料来源：兴业证券研究所

### 3、SA 因子测试

#### 3.1、因子 Rank IC

对于由最终强分类器生成的 SA 因子，下面是其 Rank IC 的测试结果。其中，

SA 因子的 IC 平均值达到了 9.16%，标准差仅为 8.98%，风险调整的 IC 高达 1.02。

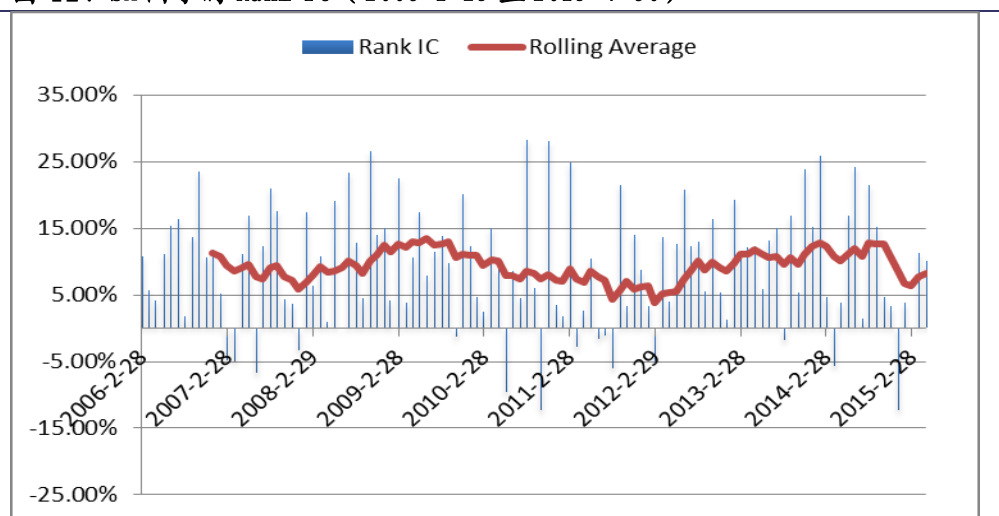
另外，考虑了行业调整后，SA 因子的 IC 表现更为突出了，IC 的平均值没有明显下降，但波动性却大幅减小，从而风险调整的 IC 也飙升到 1.34。

表 1、SA 因子 Rank IC 统计数据

因子	平均值	标准差	最小值	最大值	IC_IR	t 统计量
原始	9.16%	8.98%	-12.41%	28.24%	1.02	10.79
行业调整	8.96%	6.69%	-5.63%	24.06%	1.34	14.16

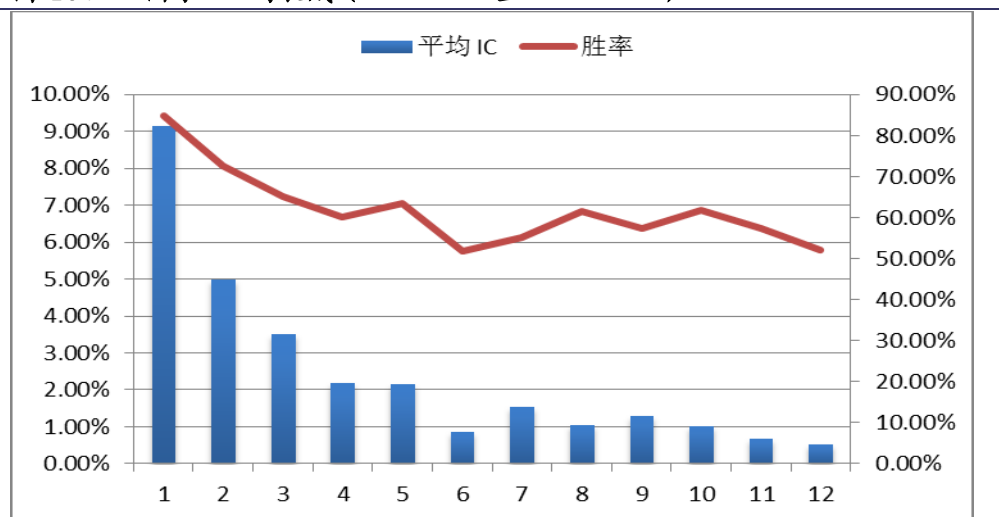
资料来源：兴业证券研究所

图 12、SA 因子的 Rank IC (2006-1-25 至 2015-4-30)



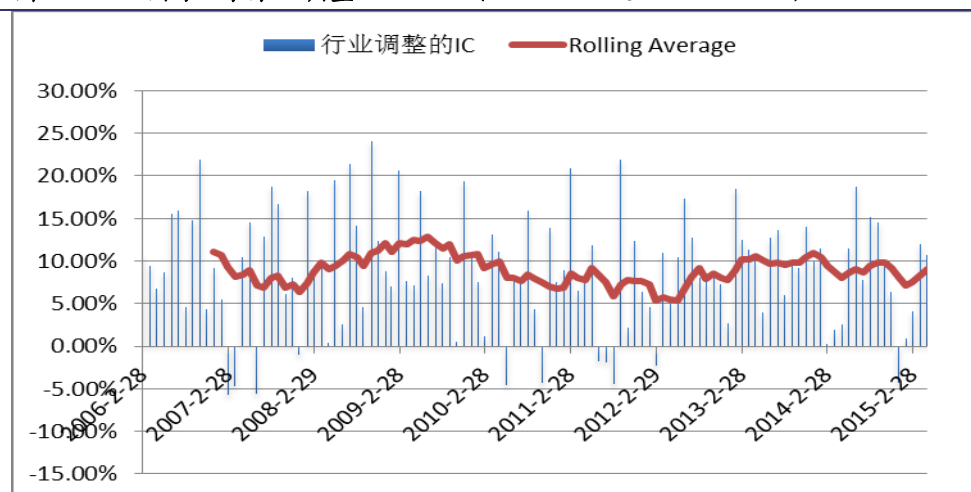
资料来源：兴业证券研究所

图 13、SA 因子 IC 的衰减 (2006-1-25 至 2015-4-30)



资料来源：兴业证券研究所

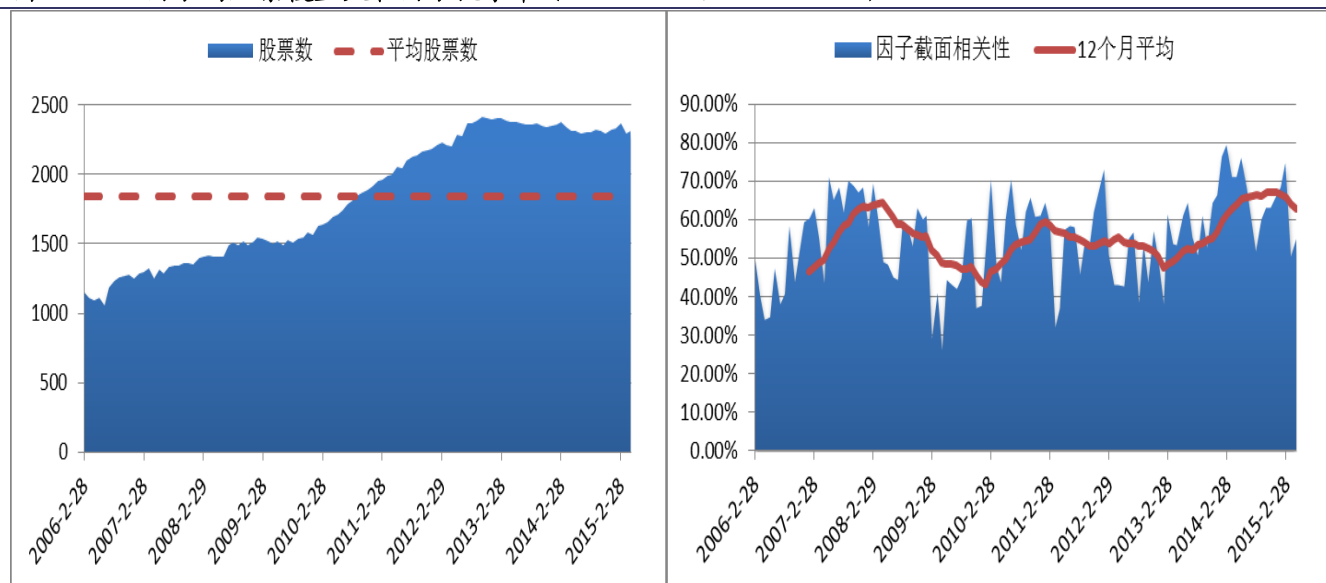
图 14、SA 因子的行业调整 Rank IC (2006-1-25 至 2015-4-30)



资料来源：兴业证券研究所

另外，我们再看一下 SA 因子的覆盖度和截面相关性。SA 因子的平均股票覆盖数约为 1838 只，对市场的覆盖性较好。SA 因子的截面相关性平均为 55.31%，相对较低，这是由于我们的算法每期都可能更换因子的选择而造成的。

图 15、SA 因子的股票覆盖数和因子换手率 (2006-1-25 至 2015-4-30)



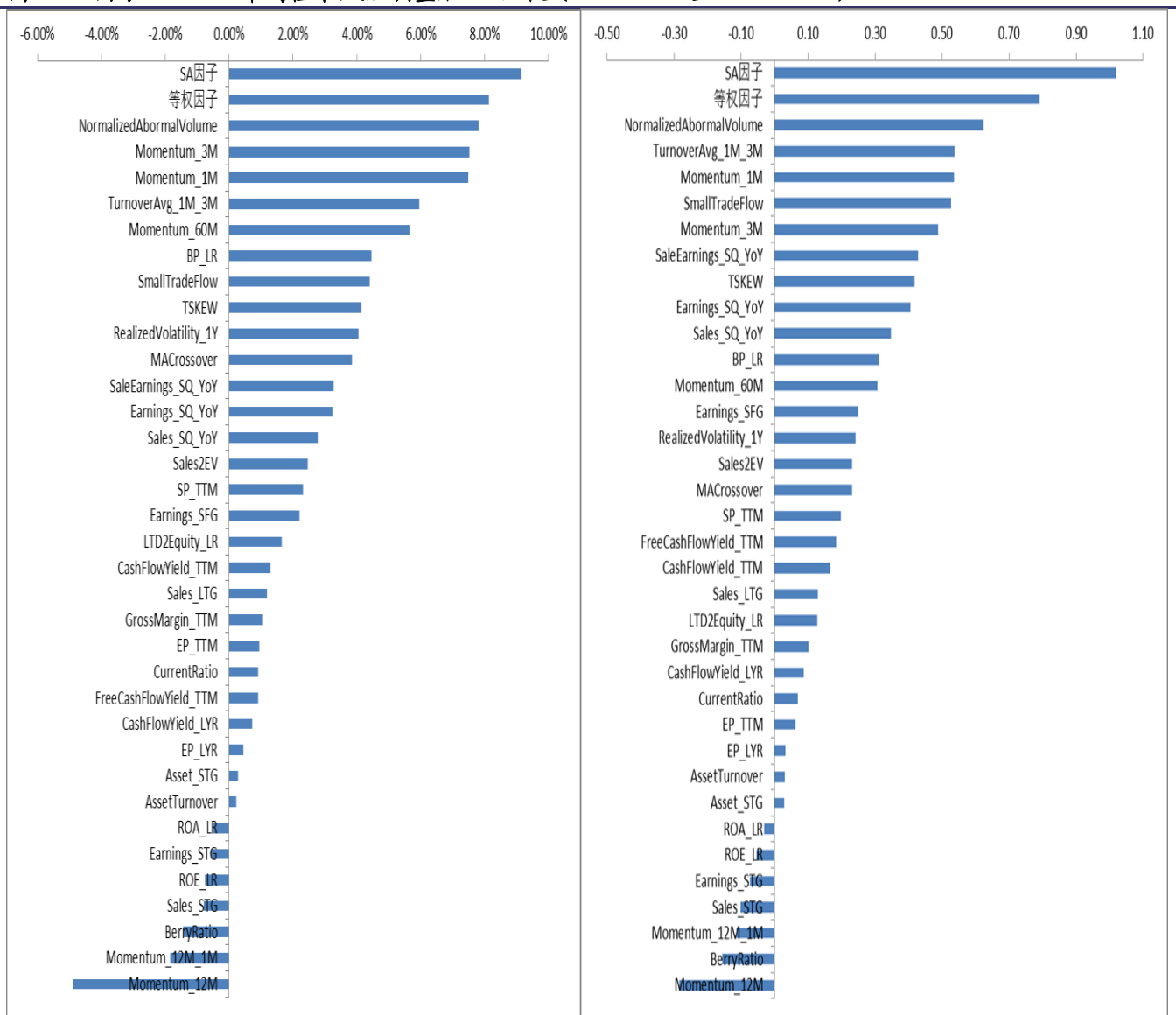
资料来源：兴业证券研究所

为了显示 SA 因子对传统因子的提升效果，我们对比了 SA 因子和用于训练样本的 33 个因子的比较结果。如下面两幅图所示。其中等权因子是将 33 个因子先按大类等权相加而后大类因子再等权相加生成的因子。不管从 Rank IC 还是风险调整的 IC 来看，SA 因子都当仁不让的位居第一，其次是等权因子。尤其从风险调整的角度看 SA 因子的表现远远超过传统因子，这说明，虽然某些单因子平均来

看表现出较强的选股能力，但往往波动较大，而多因子模型的一个优势就是利用分散化效应有效地降低波动率。从风险调整的 IC 来看，SA 因子的稳定性很好，大大提高了传统因子的效果。

另外，虽然从 IC 上来看，线性模型因子的效果也同样比较出色，但线性模型常常因数据前视偏差问题而被人们所诟病。在构建线性模型时，我们已经知道了所有因子的历史表现，并按照它们历史回测下的排序方向进行了合成，而这在回测的历史时点上是无法确知的信息。我们基于机器学习的自适应动态选股模型并没有先验的获知关于因子的任何信息，而只是通过算法去挖掘因子的收益模式，有效规避了前视偏差，同时也增强了我们对模型在样本外表现的信心。

图 16、因子 Rank IC 平均值和风险调整后 IC 比较（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

### 3.2、分位数组测试

利用前面定义的 SA 因子，我们接下来将对因子的表现进行分位数测试评价。我们的回测范围是全体 A 股，剔除当天不交易以及因子值缺失的股票；回测的时间段是 2006 年 1 月至 2015 年 4 月，每月底调仓一次；我们在选股日当天构建十分位等权组合，以当天收盘价成交。下表给出了分位组合的表现统计量：

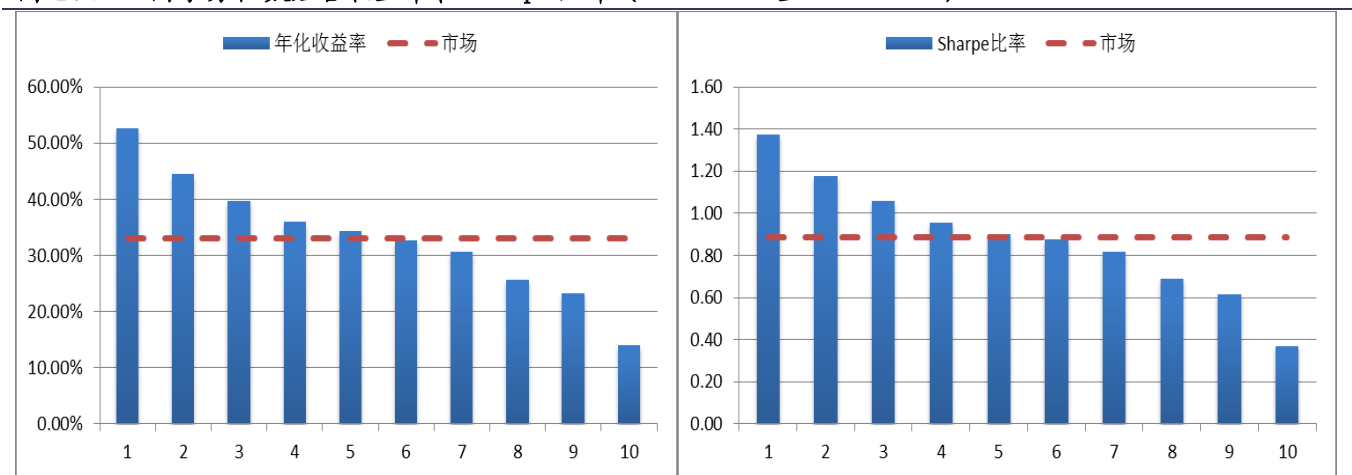
表 2、SA 因子十分位数组表现统计

组合	年化收益率	年化波动率	Sharpe 比率	超额年化收益率	跟踪误差	信息比率	胜率	双边换手率
1	52.64%	38.29%	1.37	15.09%	6.16%	2.45	75.00%	126.13%
2	44.46%	37.72%	1.18	8.71%	4.27%	2.04	71.43%	162.79%
3	39.71%	37.43%	1.06	5.00%	3.60%	1.39	70.54%	170.13%
4	35.94%	37.58%	0.96	2.21%	3.21%	0.69	61.61%	174.32%
5	34.37%	38.15%	0.90	1.21%	3.56%	0.34	52.68%	175.23%
6	32.60%	37.19%	0.88	-0.51%	3.99%	-0.13	47.32%	174.97%
7	30.71%	37.49%	0.82	-1.84%	3.66%	-0.50	40.18%	174.45%
8	25.64%	37.15%	0.69	-5.82%	3.90%	-1.49	30.36%	170.79%
9	23.28%	37.79%	0.62	-7.50%	5.67%	-1.32	31.25%	165.31%
10	13.91%	37.82%	0.37	-14.80%	7.43%	-1.99	22.32%	131.28%
多空	33.48%	12.25%	—	—	—	2.73	48.21%	—
市场	33.10%	37.36%	0.89					

资料来源：兴业证券研究所

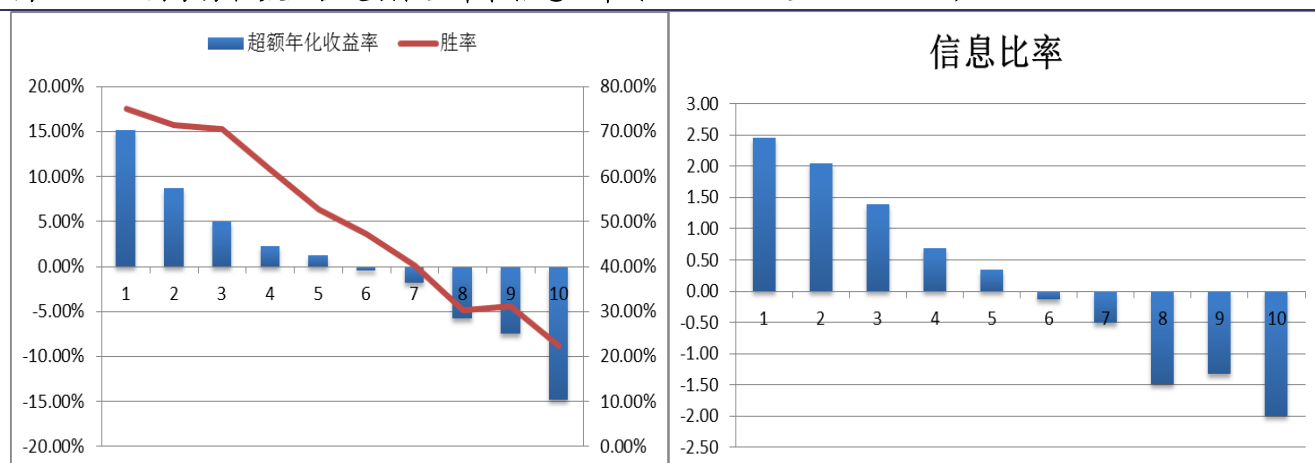
从上表中我们可以看到，SA 因子对超额收益率的预测具备良好的单调性，第一组和第十组的超额收益均十分显著，多空组合的年化信息比率达到了 2.73。

图 17、SA 因子分位数组收益率和 Sharpe 比率（2006-1-25 至 2015-4-30）



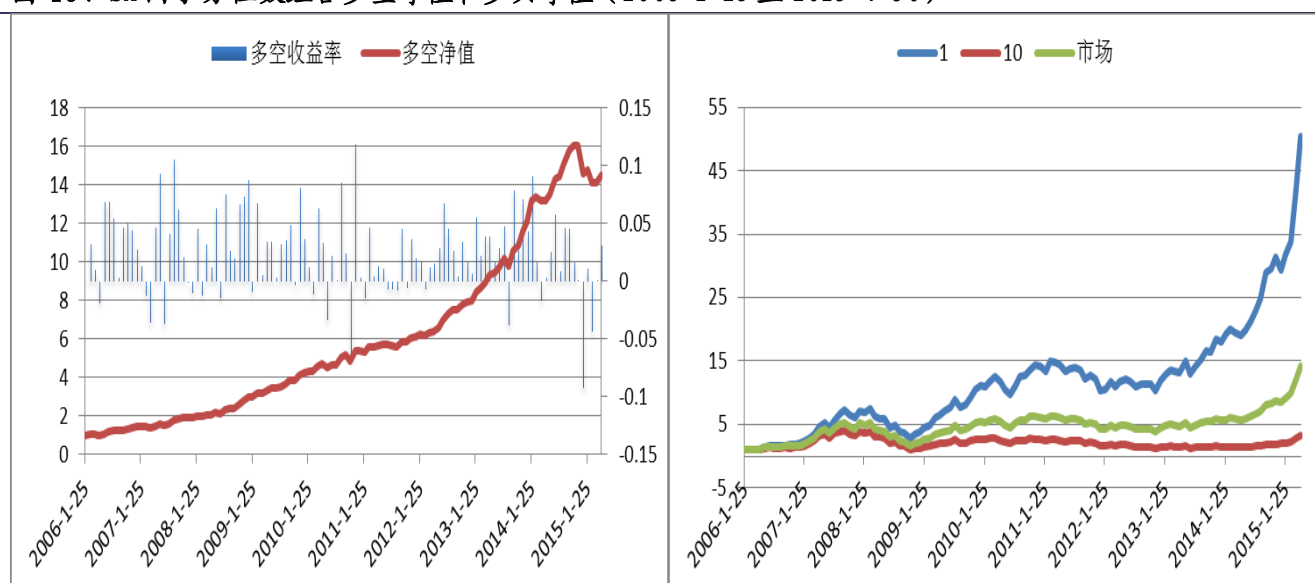
资料来源：兴业证券研究所

图 18、SA 因子分位数组合超额收益率和信息比率（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

图 19、SA 因子分位数组合多空净值和多头净值（2006-1-25 至 2015-4-30）

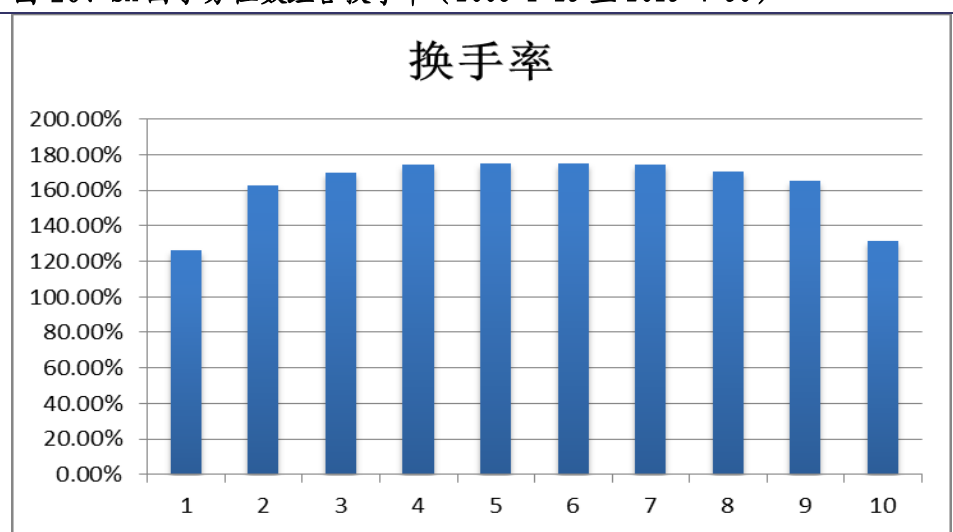


资料来源：兴业证券研究所

注意到，SA 因子的多空组合在 2014 年 11 月 28 日至 2015 年 2 月 27 日出现了明显的回撤，回撤率为 12.18%。我们认为这和当时的市场环境有很大关系。2014 年下半年市场风格突然切换，变化较快，而我们的模型是基于过去 12 个月的数据进行学习而得到的，因而滞后性是必然的。对于这种快速变化的风格切换，我们需要一个更为灵敏的模型来进行补充。在下一节中，我们将通过对日历效应的分析，进一步提升模型的稳定性。



图 20、SA 因子分位数组合换手率（2006-1-25 至 2015-4-30）



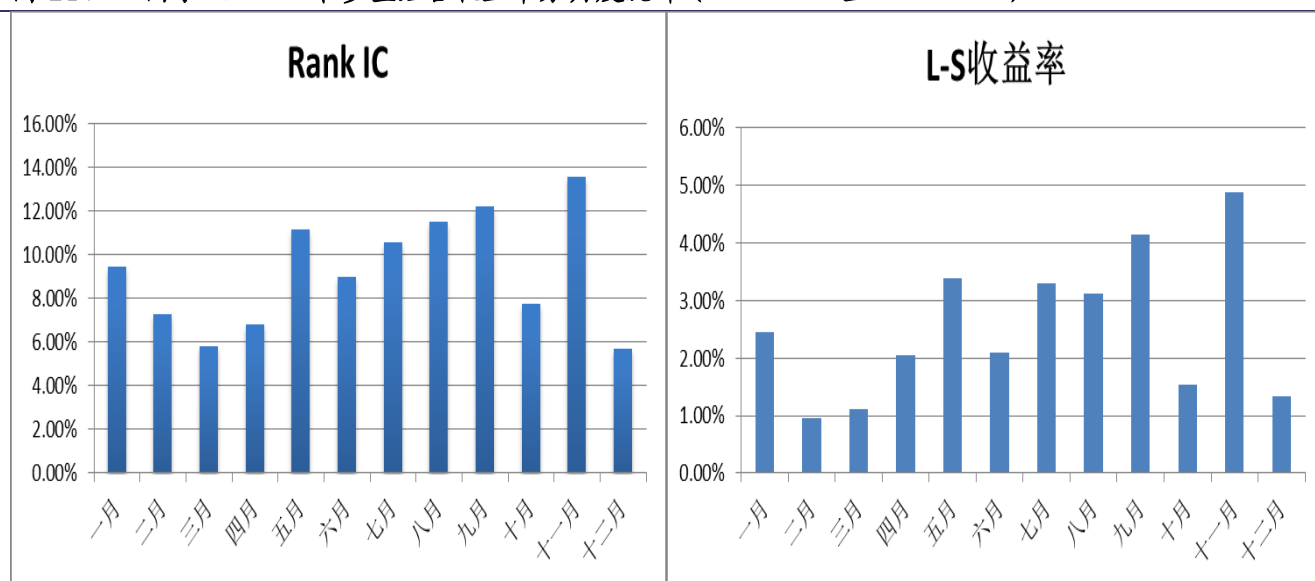
资料来源：兴业证券研究所

## 4、百尺竿头，更进一步

### 4.1、日历效应

我们上面构建的 SA 因子已经表现出较好的预测能力,但同时我们也看到了其波动性和回撤仍然不是特别令人满意。那该如何进一步提升模型的表现呢?尤其是如何让模型的表现更加稳定呢?如果我们更深入的考察模型,可以发现模型的收益呈现出较为明显的月度效应。下图展示了模型在不同月份的平均 IC 和多空组合收益率。模型在二月、三月、四月、十月以及十二月的表现较差,而在十一月、九月、八月、五月的表现比较突出。由此可见模型在不同的月份表现并不一致。

图 21、SA 因子 Rank IC 和多空组合收益率分月度统计（2006-1-25 至 2015-4-30）



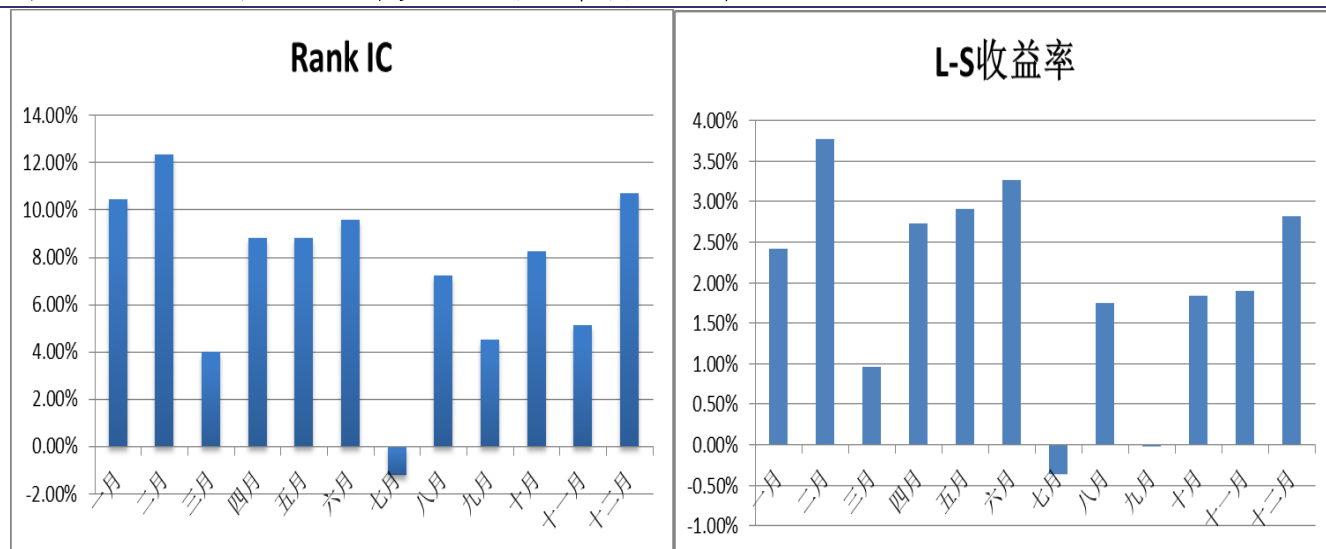
资料来源：兴业证券研究所

为了捕捉不同月份风格的特征，我们考虑使用同一月份的数据来训练样本。

具体的讲，在每个月末日，我们使用过去数据中与当前月份相同的数据来作为算法的训练集，其他设置不变。这里我们使用过去 5 年（不足 5 年时取尽可能多的数据）的相同月份数据来训练模型。我们称此模型输出的复合因子为 SA-5SM 因子。

下图是 SA-5SM 因子不同月份的平均 IC 和多空组合收益率，可以发现，与 SA 因子的表现是有很大区别的。其表现突出的月份为十二月、二月等，而表现落后的月份为七月、三月、九月等。SA-5SM 因子和 SA 因子的表现有一定的互补性，因而我们预期两个因子结合起来后表现的稳定性会有一定的提升。

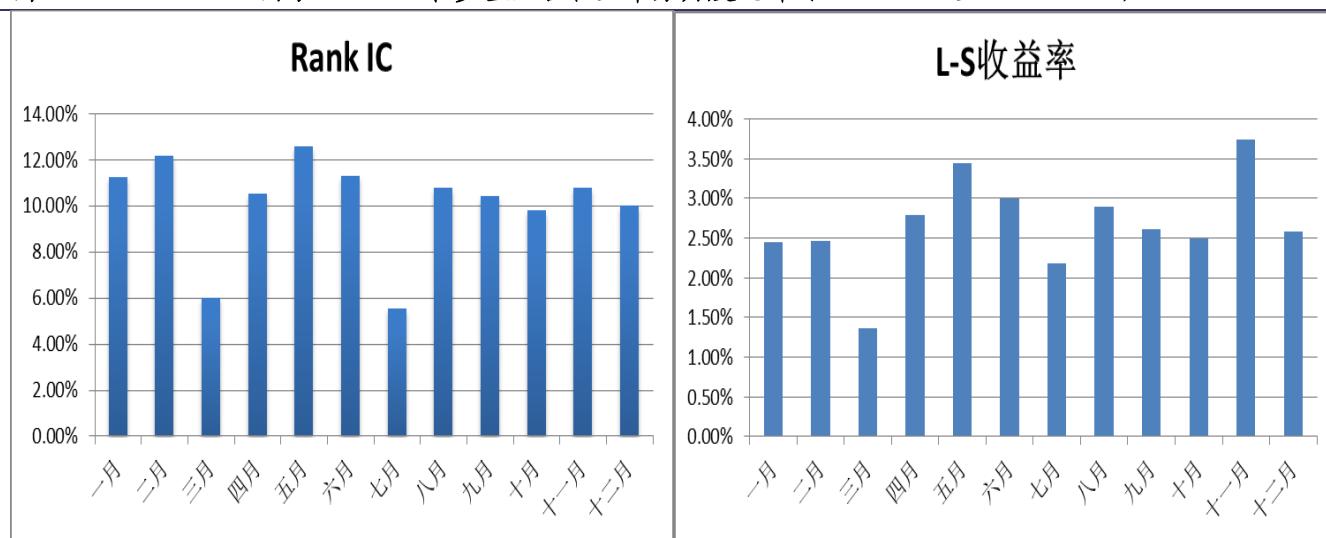
图 22、SA-5SM 因子 Rank IC 和多空组合收益率分月度统计（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

我们将 SA-5SM 因子和 SA 因子标准化后以等权的方式相加得到本文最终的复合因子：SA-12M-5SM 因子。下图是 SA-12M-5SM 因子的月度表现分布。可见，复合后的 SA-12M-5SM 因子在各个月份上的表现相较之前而言更具一致性。

图 23、SA-12M-5SM 因子 Rank IC 和多空组合收益率分月度统计（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

## 4.2、SA-12M-5SM 因子 Rank IC

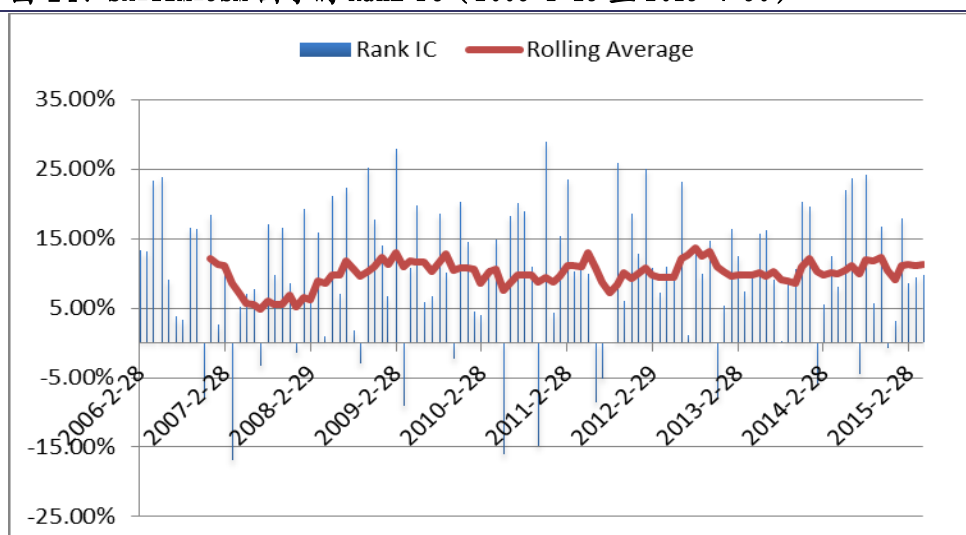
下表是 SA-12M-5SM 因子的 Rank IC 的测试结果。其中，SA-12M-5SM 因子的 IC 平均值已经超过了 10%，而标准差仅略为有所增长，其 IC 的信息比率更是高达 1.05。另外，考虑了行业调整后，SA-12M-5SM 因子表现也相应的有所提高，这和 SA 因子的情况是一致的。这个现象也提醒我们，行业中性可能会大幅调高模型的表现，这也是我们以后进一步改进模型的一个方向。

表 3、SA-12M-5SM 因子 Rank IC 统计数据

因子	平均值	标准差	最小值	最大值	IC_IR	t 统计量
原始	10.09%	9.65%	-17.00%	28.90%	1.05	11.07
行业调整	9.94%	7.90%	-15.77%	24.33%	1.26	13.31

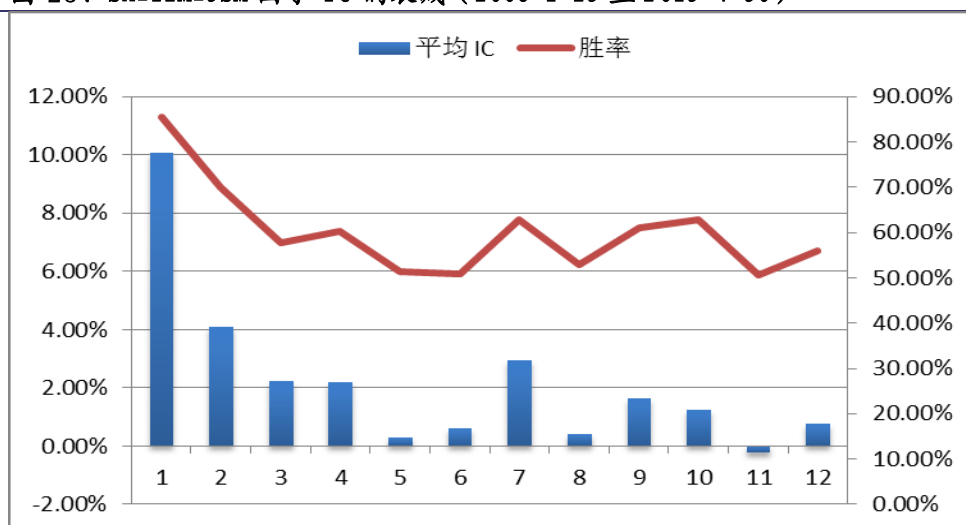
资料来源：兴业证券研究所

图 24、SA-12M-5SM 因子的 Rank IC (2006-1-25 至 2015-4-30)



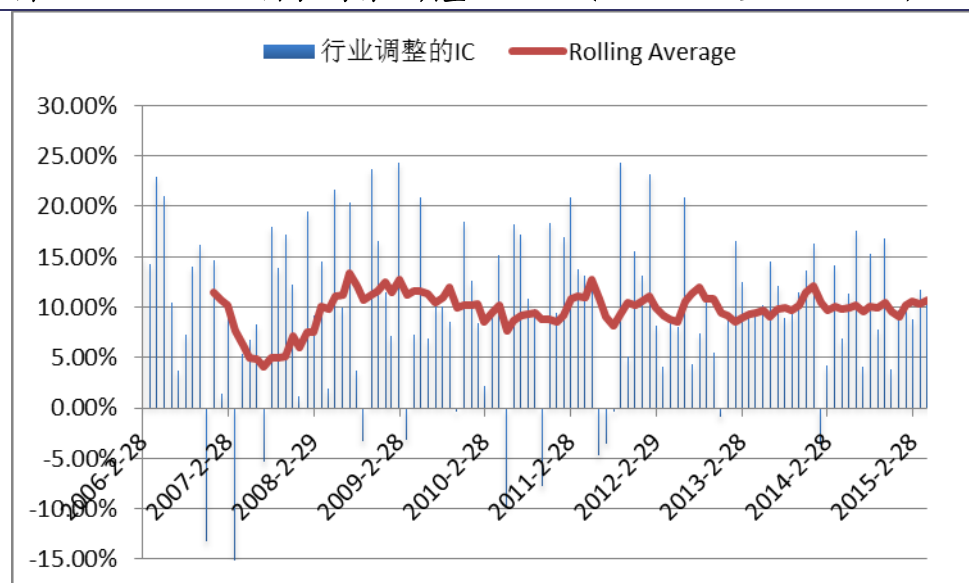
资料来源：兴业证券研究所

图 25、SA-12M-5SM 因子 IC 的衰减 (2006-1-25 至 2015-4-30)



资料来源：兴业证券研究所

图 26、SA-12M-5SM 因子的行业调整 Rank IC (2006-1-25 至 2015-4-30)

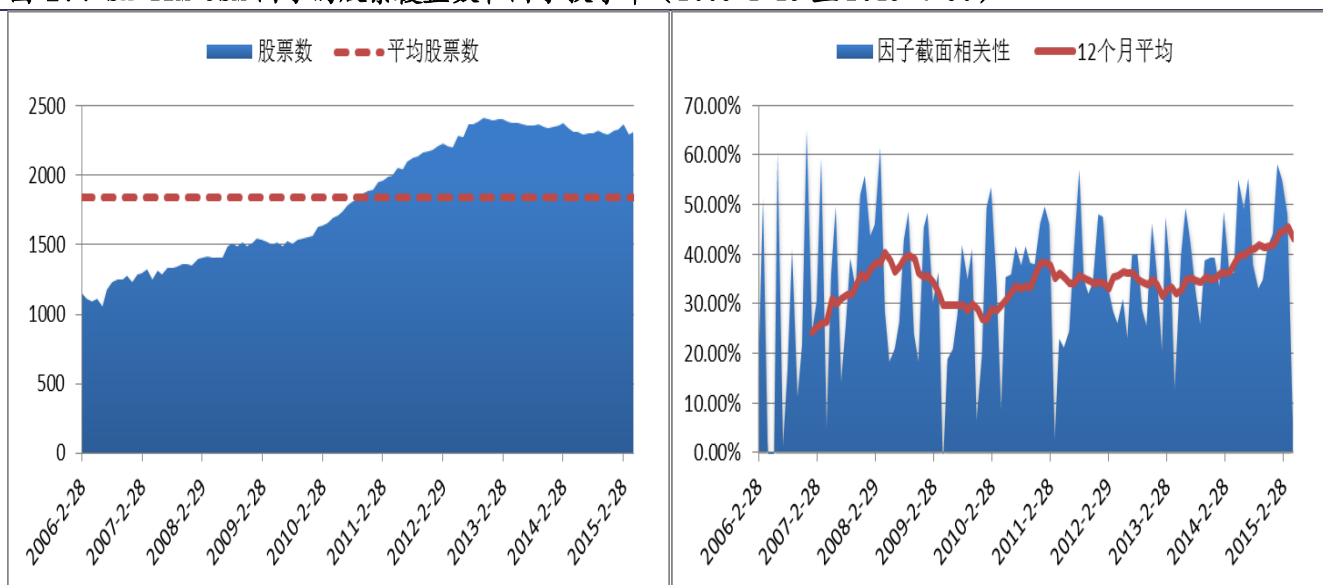


资料来源：兴业证券研究所

另外，从 IC 的时间变化的图像上来看，SA-12M-5SM 因子的主要回撤出现在 2007、2008 年，我们认为这可能和前期训练样本的数据较少有关。2010 年以后，训练样本的长度固定在 5 年，IC 的表现也稳定在了 10% 上下。

我们再看一下 SA-12M-5SM 因子的覆盖度和截面相关性。SA 因子的平均股票覆盖数约为 1836 只。SA-12M-5SM 因子的截面相关性平均为 34.31%，同 SA 因子相比有所降低，这是跟我们加入了 SA-5SM 因子有关，两个因子相关性较低，风格的变换更为频繁。

图 27、SA-12M-5SM 因子的股票覆盖数和因子换手率 (2006-1-25 至 2015-4-30)



资料来源：兴业证券研究所

### 4.3、SA-12M-5SM 因子分位数组合测试

我们接下来将对 SA-12M-5SM 因子的表现进行分位数测试评价。我们的回测范围依然是剔除当天不交易以及因子值缺失的全体 A 股；回测的时间段是 2006 年 1 月至 2015 年 4 月，每月底调仓一次；我们在选股日当天构建十分位等权组合，以当天收盘价成交。

下表给出了分位组合的表现统计量：

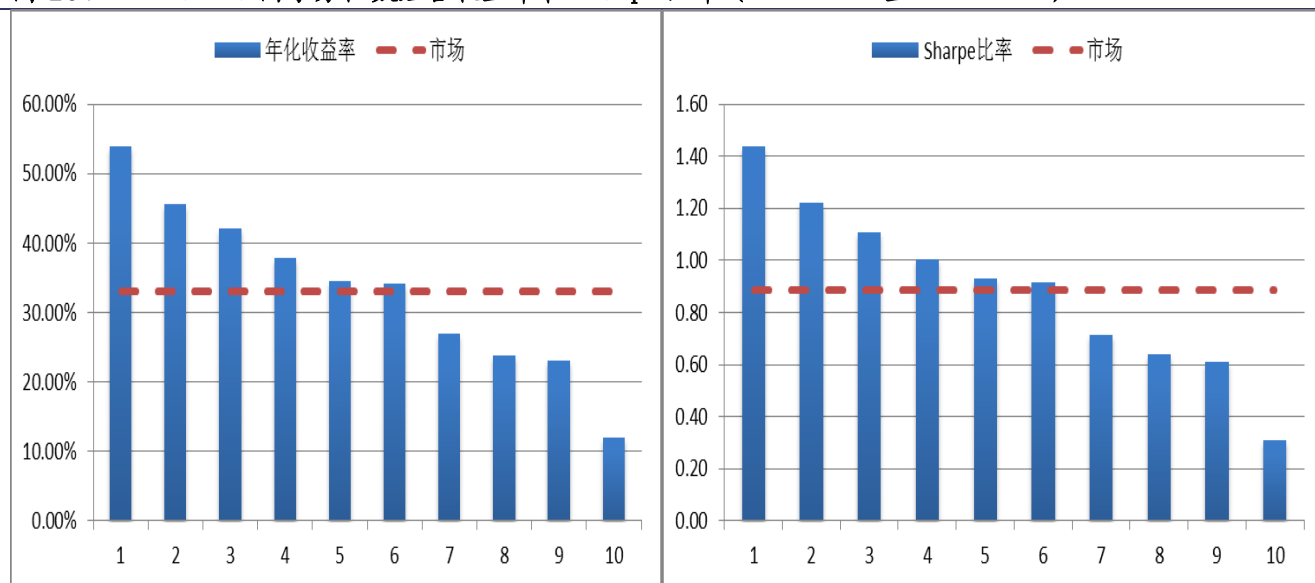
表 4、SA-12M-5SM 因子十分位数组合表现统计

组合	年化收益率	年化波动率	Sharpe 比率	超额年化收益率	跟踪误差	信息比率	胜率	换手率
1	53.99%	37.55%	1.44	15.84%	6.20%	2.56	78.57%	150.70%
2	45.65%	37.34%	1.22	9.42%	5.03%	1.87	65.18%	170.74%
3	42.14%	38.02%	1.11	7.03%	4.50%	1.56	68.75%	175.57%
4	37.88%	37.65%	1.01	3.72%	3.10%	1.20	65.18%	177.93%
5	34.54%	37.19%	0.93	0.95%	4.14%	0.23	50.00%	177.81%
6	34.21%	37.29%	0.92	0.77%	3.59%	0.22	52.68%	177.73%
7	26.93%	37.69%	0.71	-4.66%	3.96%	-1.18	35.71%	177.98%
8	23.71%	37.17%	0.64	-7.31%	4.36%	-1.67	25.00%	176.18%
9	23.07%	37.88%	0.61	-7.66%	5.33%	-1.44	33.04%	172.33%
10	12.04%	39.03%	0.31	-15.80%	6.97%	-2.27	23.21%	152.94%
多空	36.01%	11.71%				3.08	50.00%	
市场	33.10%	37.37%	0.89					

资料来源：兴业证券研究所

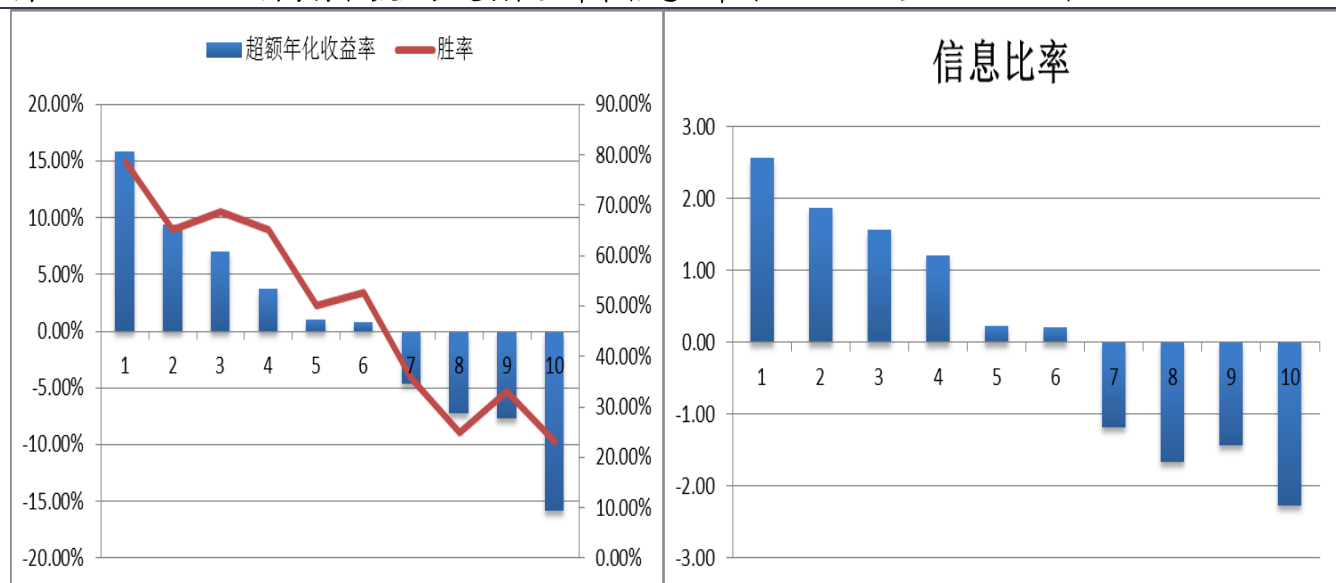
从上表中我们可以看到，相比于 SA 因子，SA-12M-5SM 因子对超额收益率的预测性更好，波动性也更低。第一组和第十组的超额收益均十分显著，多空组合的年化信息比率更是超过了 3。

图 28、SA-12M-5SM 因子分位数组合收益率和 Sharpe 比率（2006-1-25 至 2015-4-30）



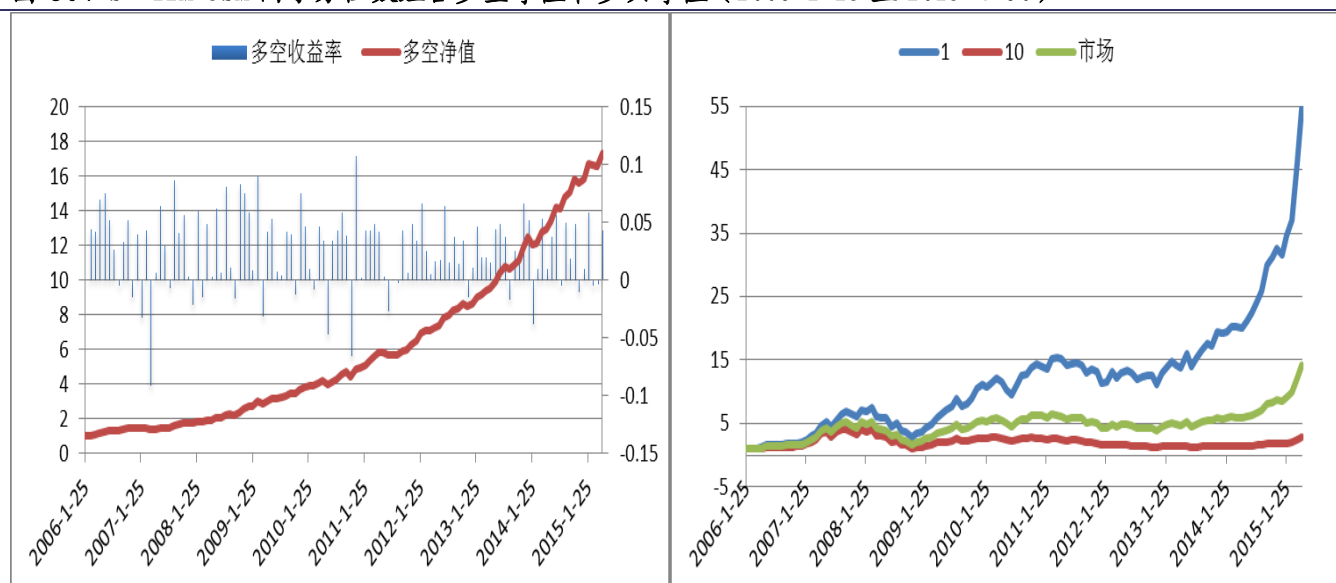
资料来源：兴业证券研究所

图 29、SA-12M-5SM 因子分位数组合超额收益率和信息比率（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

图 30、SA-12M-5SM 因子分位数组合多空净值和多头净值（2006-1-25 至 2015-4-30）

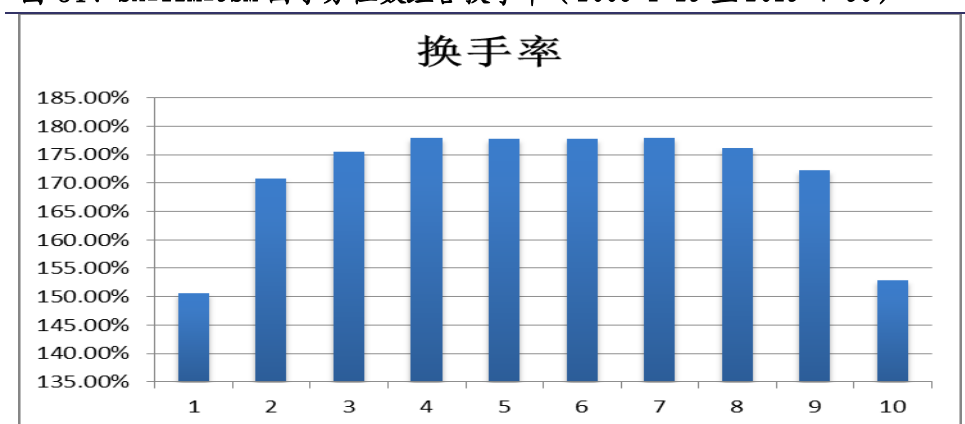


资料来源：兴业证券研究所

注意到，相比于 SA 因子的多空组合，在 2014 年底和 2015 年初的回撤在 SA-12M-5SM 因子上已经不再明显。从整体上看，SA-12M-5SM 因子的最大回撤率仅为 9.12%，而且发生的时间已经不再是 2014 年底了，是在 2007 年 2 月 28 日至 2007 年 3 月 30 日发生的。



图 31、SA-12M-5SM 因子分位数组合换手率（2006-1-25 至 2015-4-30）



资料来源：兴业证券研究所

#### 4.4、SA-12M-5SM 因子在不同规模股票池中的表现

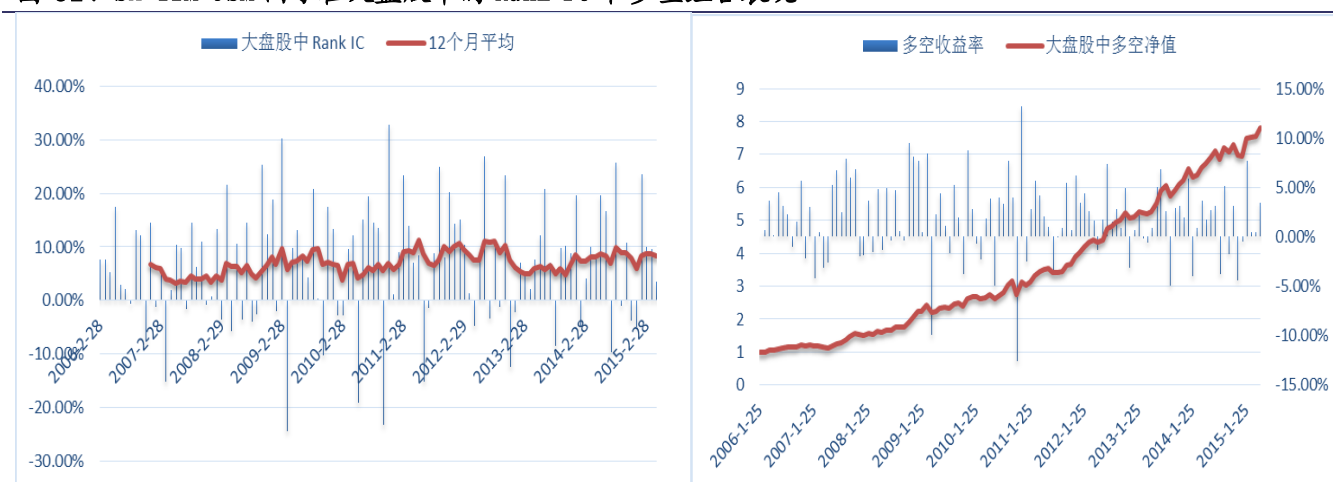
当我们缩小选股范围时，大多数因子的表现将会下降，甚至不再显著。这里我们考察 SA-12M-5SM 因子在不同规模股票池中选股的有效性。每一时点，我们将流通市值超过其中位数的股票称为大盘股，低于中位数的股票称为小盘股。下面的图表展示了在大盘股和小盘股中 SA-12M-5SM 因子的选股能力。可以发现，在大盘股池中，SA-12M-5SM 因子的表现虽然有一定的下降，但依然保持着显著的选股能力。在小盘股中因子的表现更为突出。这个实证检验也说明市值是我们要注意的风险点，在以后的改进中也是可以考虑的方面。

表 5、SA-12M-5SM 因子在不同股票池中 Rank IC 和多空组合表现

Rank IC	大盘股	小盘股	多空组合	大盘股	小盘股
IC 均值	7.00%	12.92%	年化收益率	24.89%	48.88%
IC 标准差	11.01%	8.43%	年化波动率	13.69%	10.80%
IC-IR	0.64	1.53	Sharpe 比率	1.82	4.52

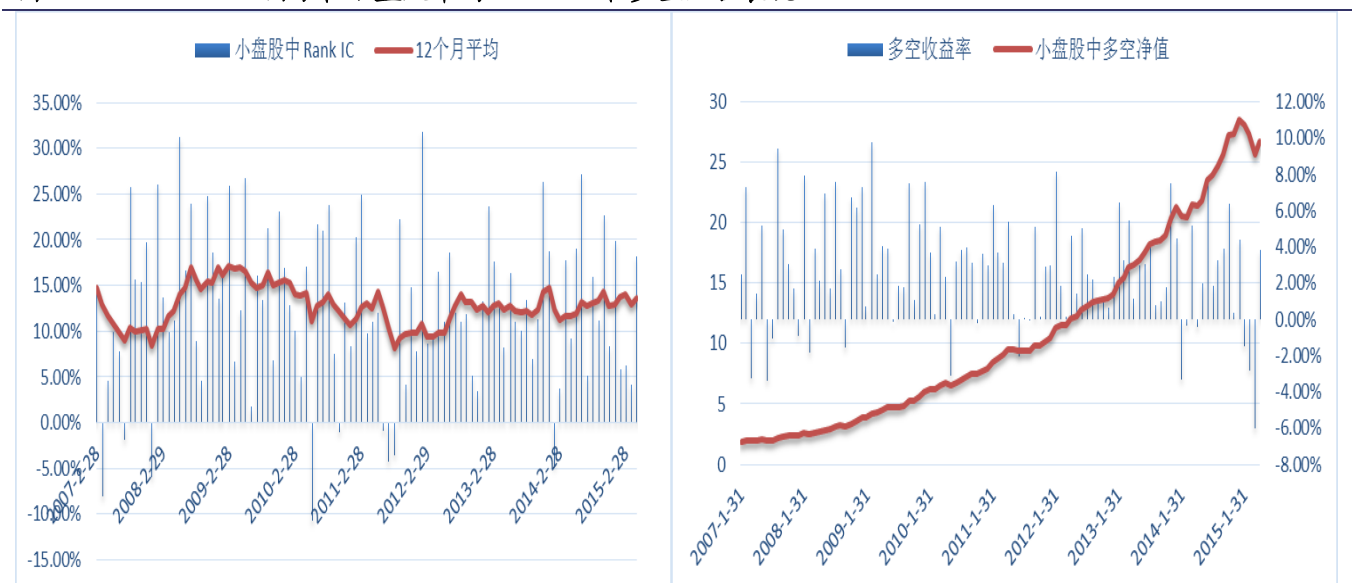
资料来源：兴业证券研究所

图 32、SA-12M-5SM 因子在大盘股中的 Rank IC 和多空组合表现



资料来源：兴业证券研究所

图 33、SA-12M-5SM 因子在小盘股中的 Rank IC 和多空组合表现



资料来源：兴业证券研究所

## 5、基于 Smart Alpha 的选股策略

### 5.1、策略表现

在本节中我们利用 SA-12M-5SM 因子构建我们的 alpha 对冲选股策略。具体的策略构建方式如下：

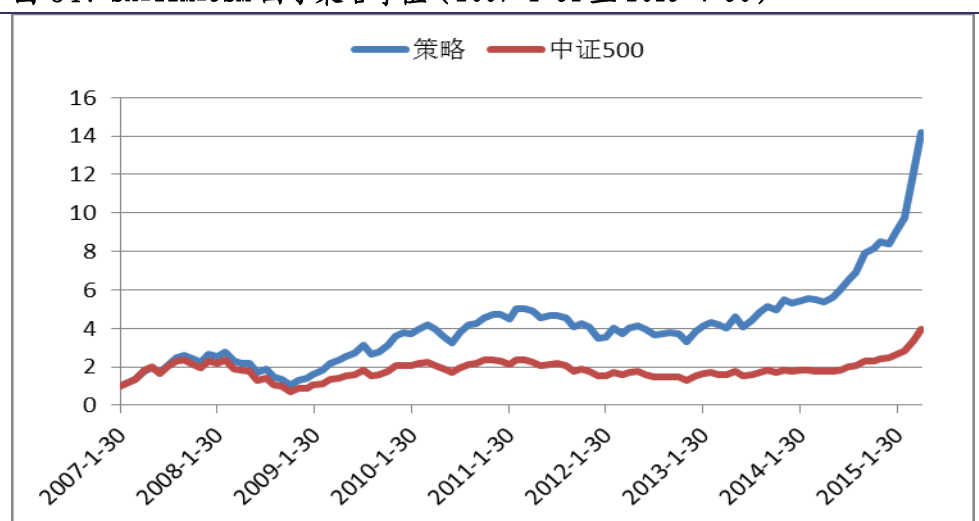
在每个月的月底，先剔除不能正常交易、涨跌停或者被特别处理的股票。为了规避行业的风险，在剩下的股票池中我们在每个中信一级行业内按照 SA-12M-5SM 因子对股票进行降序排列，选择前 10% 的股票形成投资组合。以中证 500 为基准指数，每个行业以其在中证 500 中的权重加权，行业内的股票等权。我们的回测区间是从 2007 年 1 月 31 日至 2015 年 4 月 30 日（中证指数公司在 2007 年 1 月开始公布中证 500 的成分股）。手续费率是单边 0.3%。以下是策略的回测结果。

表 6、SA-12M-5SM 因子策略统计数据

统计指标	策略	中证 500	多空
年化收益率	37.91%	18.13%	16.31%
年化波动率	36.15%	36.93%	5.84%
Sharpe 比率	1.05	0.49	2.79
最大回撤率	61.83%	69.27%	4.93%
最大回撤开始日期	20080229	20070928	20070228
最大回撤结束日期	20081031	20081031	20070430
胜率	63.64%	62.63%	82.83%

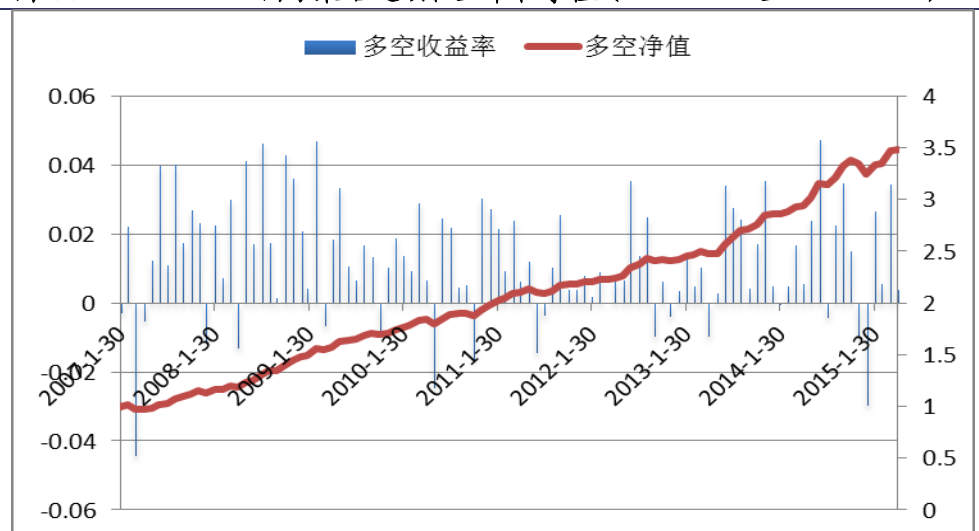
资料来源：兴业证券研究所

图 34、SA-12M-5SM 因子策略净值 (2007-1-31 至 2015-4-30)



资料来源：兴业证券研究所

图 35、SA-12M-5SM 因子策略超额收益率和净值 (2007-1-31 至 2015-4-30)



资料来源：兴业证券研究所

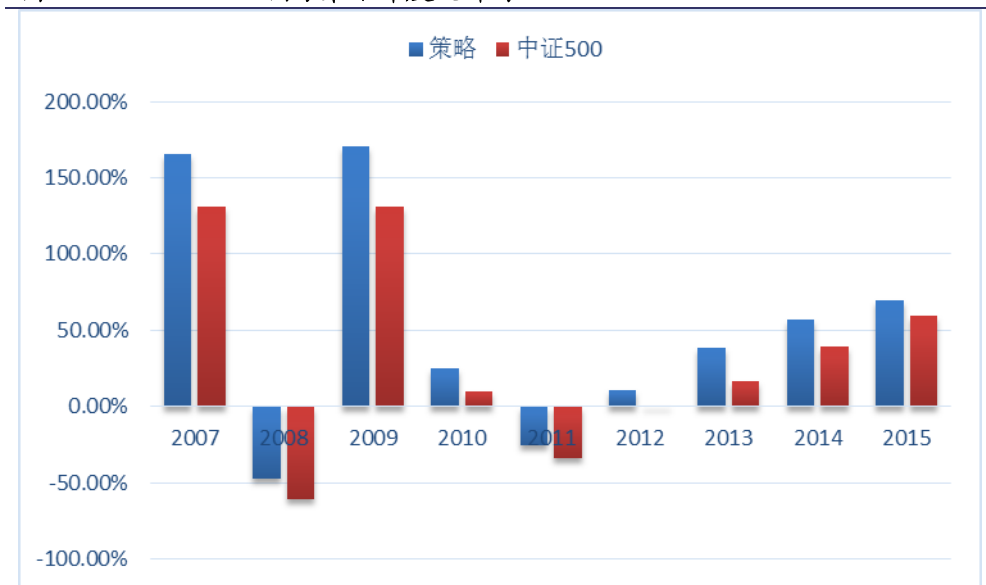
从上述的统计结果看，我们的策略的年化收益为 37.91%，超过基准近 20 个点。策略的最大回撤发生在 2007 年三、四两个月，而在 2014 年底的回撤仅为 3.93%。

表 7、SA-12M-5SM 因子策略分年度统计

年份	策略	中证 500	多空
2007	165.74%	130.88%	13.78%
2008	-47.51%	-60.80%	30.49%
2009	170.81%	131.27%	17.68%
2010	25.12%	10.07%	13.78%
2011	-25.77%	-33.83%	11.23%
2012	10.38%	0.28%	9.86%
2013	38.42%	16.89%	18.35%
2014	56.76%	39.01%	13.23%
2015	69.22%	59.12%	7.20%

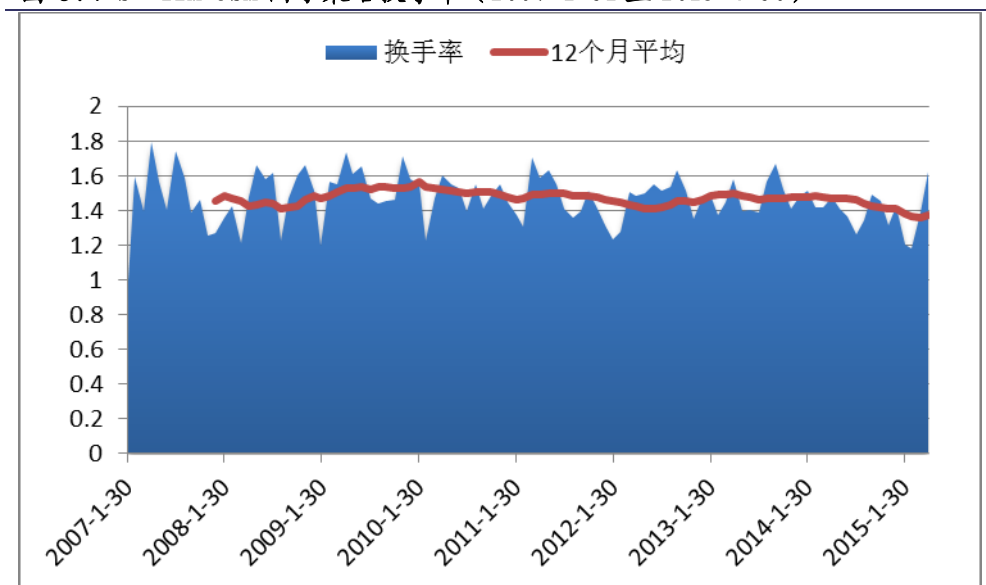
资料来源：兴业证券研究所

图 36、SA-12M-5SM 因子策略年度统计对比



资料来源：兴业证券研究所

图 37、SA-12M-5SM 因子策略换手率（2007-1-31 至 2015-4-30）



资料来源：兴业证券研究所

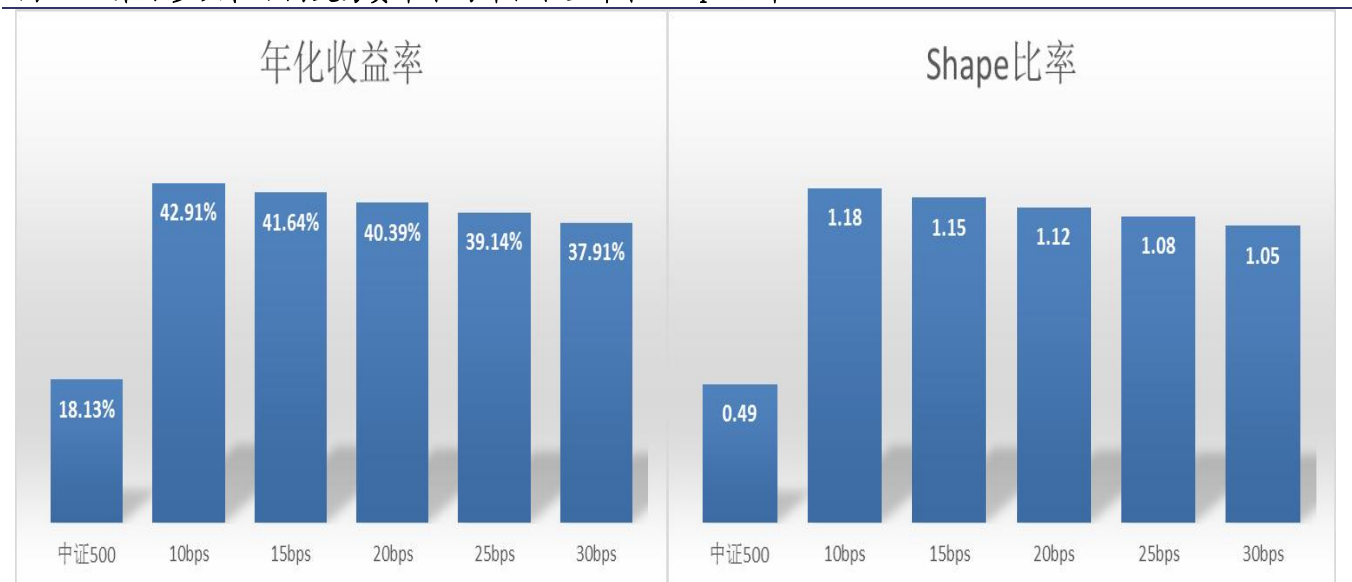
通过分年度统计可以发现，在所有年份，我们的策略都战胜了基准。另外策略的换手率比较高，因此在下一节中我们将着重分析交易成本对我们策略的影响。

## 5.2、策略对交易费率的敏感性

由于我们的策略换手率较高，交易费将对策略的表现产生较大的影响。因而本小节我们主要考察不同交易费率下策略的表现。以下的图表展示的是单边交易费率分别为 0.1%、0.15%、0.2%、0.25%、0.3% 是策略的收益和风险调整后的表现

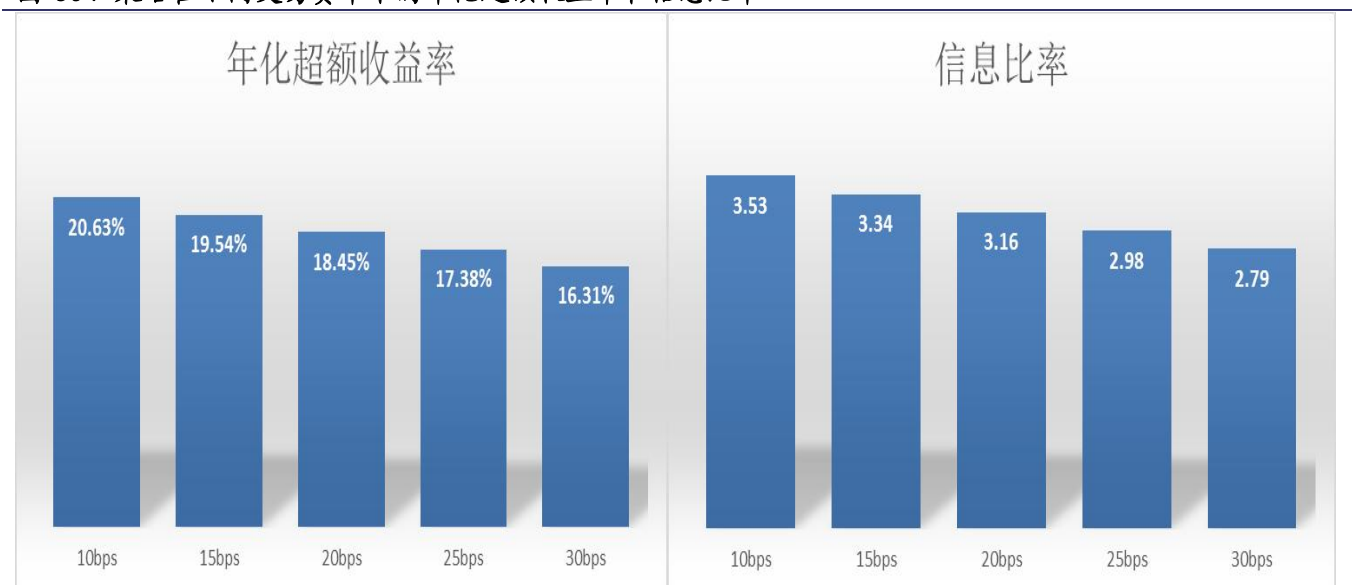
情况。正如我们的预期,随着交易费率的提高,策略的表现渐趋恶化。每提高 0.05% 的交易费率,年化收益下降大约 1.25%, 夏普比率下降大约 0.03, 而年化超额收益下降大约 1.08%, 信息比率下降大约为 0.18。交易费率的对策略的影响还是比较明显, 但即便使用 0.3% 的交易费率, 我们的策略依然有稳定的超额收益。

图 38、策略多头在不同交易费率下的年化收益率和 Shape 比率



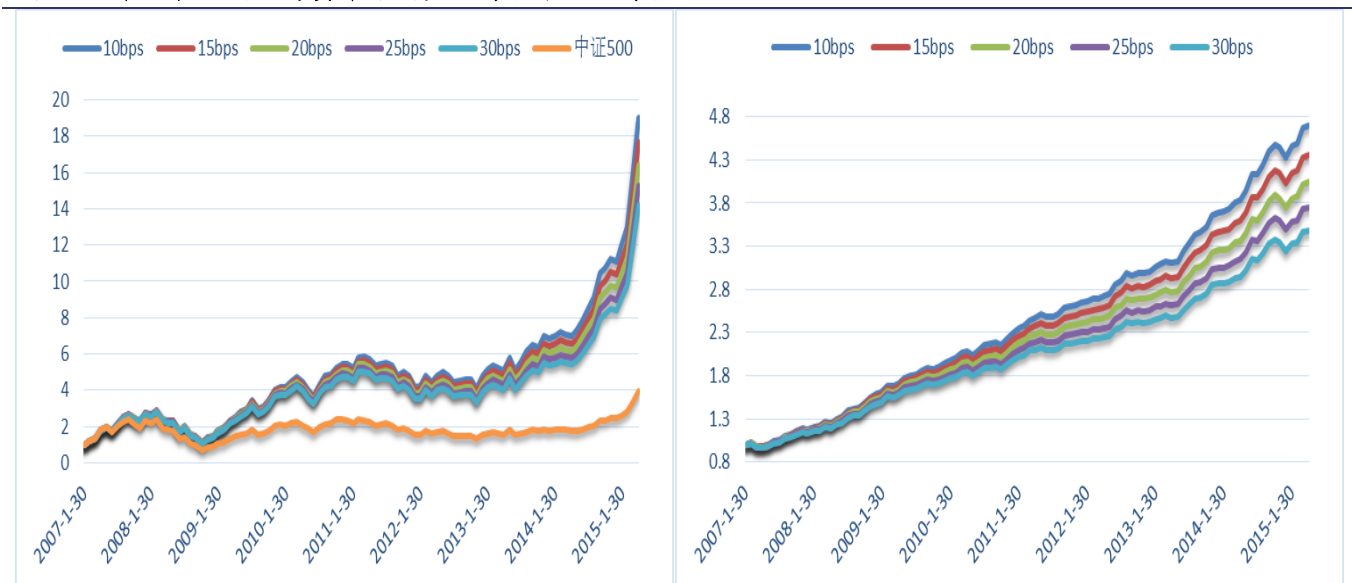
资料来源: 兴业证券研究所

图 39、策略在不同交易费率下的年化超额收益率和信息比率



资料来源: 兴业证券研究所

图 40、策略在不同交易费率下的多头净值和超额净值



资料来源：兴业证券研究所

### 5.3、策略滞后成交表现

为了更真实的考察策略的表现，我们在每个月月底形成投资组合的调仓信号，但用下一个交易日的开盘价进行调仓。在这一设定下，策略的表现如下表所示。对比用当天收盘价成交的结果，可以看到策略的表现并没有明显的下降，这充分说明我们的策略稳健性较好。

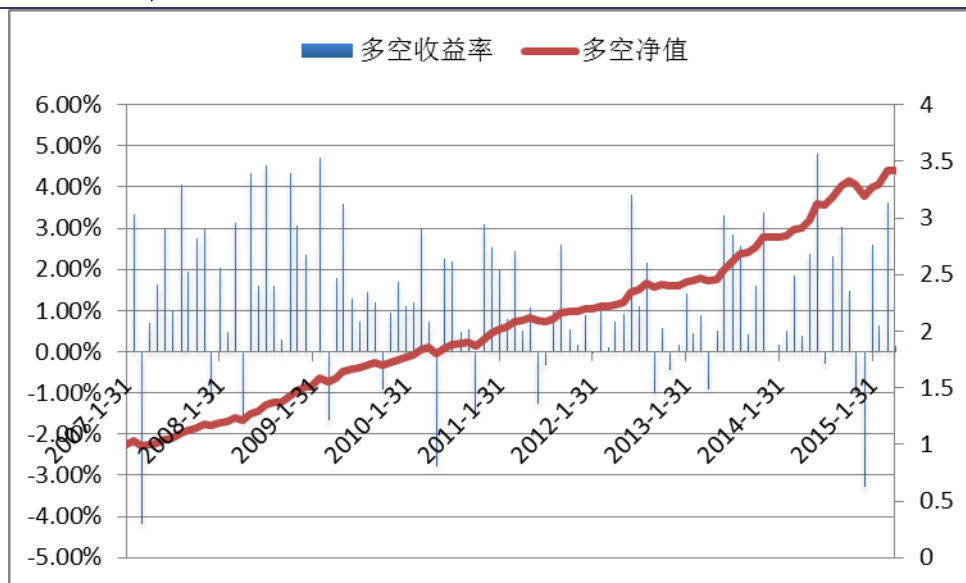
表 8、SA-12M-5SM 因子策略滞后成交统计数据

统计指标	策略	中证 500	多空
年化收益率	37.53%	18.13%	16.07%
年化波动率	36.72%	36.93%	5.96%
Sharpe 比率	1.02	0.49	2.70
最大回撤率	61.98%	69.27%	4.23%
最大回撤开始日期	20080229	20070928	20141031
最大回撤结束日期	20081031	20081031	20141231
胜率	63.64%	62.63%	83.84%

资料来源：兴业证券研究所



图 41、SA-12M-5SM 因子策略滞后成交超额收益率和净值（2007-1-31 至 2015-4-30）



资料来源：兴业证券研究所

## 6、总结

本文是兴业证券金融工程团队《猎金：Alpha 因子探索系列报告》的第六篇。在这篇文章中，我们基于机器学习中的 AdaBoost 算法，深入挖掘传统因子的选股能力，构建了非线性的多因子模型，得到了兴业 Smart Alpha 动态选股模型，并构建了相应的 SA 因子。该因子巧妙的利用了已有的传统因子，极大地改善了它们效果，为多因子选股模型的建立提供了一个有意义的新方向。

本系列研究的目的是为广大机构投资者在构建自己的 alpha 模型时提供不同的想法和思路，从而提升模型的竞争力。希望这次关于 alpha 模型的一点思考也能给您带来启发和帮助，同时我们也将继续探索的进程，向着更高的目标前进！

## 附录 1: 兴业因子池

表 9、兴业因子池

类别	因子名称	因子定义	方向
价值	DividendYield_FY0	预测每股股利-最近预测年度 / 收盘价	降序
价值	DividendYield_FY1	预测每股股利-最近预测年度次年 / 收盘价	降序
价值	Price2EPS_LYR	收盘价 / 基本每股收益-最新年报	升序
价值	EP_LYR	净利润(不含少数股东损益)-最新年报 / 总市值	降序
价值	EP_TTM	净利润(不含少数股东损益)-TTM / 总市值	降序
价值	EP_Fwd12M	净利润-未来 12 个月 / 总市值	降序
价值	EP_FY0	净利润-最近预测年度 / 总市值	降序
价值	EP_FY1	净利润-最近预测年度次年 / 总市值	降序
价值	SP_TTM	营业收入-TTM / 总市值	降序
价值	CashFlowYield_LYR	经营活动产生的现金流量净额-TTM / 总市值	降序
价值	CashFlowYield_FY0	预测每股现金流-最近预测年度 / 收盘价	降序
价值	CashFlowYield_TTM	经营活动产生的现金流量净额-TTM / 总市值	降序
价值	FreeCashFlowYield_TTM	(经营活动产生的现金流量净额-TTM - 购建固定资产、无形资产和其他长期资产支付的现金-TTM) / 总市值	降序
价值	BP_LR	股东权益合计(不含少数股东权益)-最新财报 / 总市值	降序
价值	BP_FY0-Median	预测每股净资产中值-最近预测年度 / 收盘价	降序
价值	Sales2EV	营业收入-TTM / (总市值 + 非流动负债合计-最新财报 - 货币资金-最新财报)	降序
成长	SaleEarnings_SQ-YoY	单季度营业利润同比增长率	降序
成长	Earnings_SQ-YoY	单季度净利润同比增长率	降序
成长	Sales_SQ-YoY	单季度营业收入同比增长率	降序
成长	Earnings_LTG	净利润长期(5 年)历史增长率	降序
成长	Sales_LTG	营业收入长期(5 年)历史增长率	降序
成长	Earnings_STG	净利润短期(1 年)历史增长率	降序
成长	Sales_STG	营业收入短期(1 年)历史增长率	降序
成长	Earnings_LFG	净利润长期(4 年)预期增长率	降序
成长	Sales_LFG	营业收入长期(4 年)预期增长率	降序
成长	Earnings_SFG	净利润短期(1 年)预期增长率	降序
成长	Sales_SFG	营业收入短期(1 年)预期增长率	降序
成长	Asset_STG	总资产短期(1 年)历史增长率	升序
质量	ROE_LR	净利润(不含少数股东损益)-TTM / 股东权益合计(不含少数股东权益)-最新财报	降序
质量	ROA_LR	净利润(不含少数股东损益)-TTM / 资产总计-最新财报	降序
质量	GrossMargin_TTM	(营业收入-TTM - 营业成本-TTM) / 营业收入-TTM - 1	降序

质量	LTD2Equity-LR	非流动负债合计-最新财报 / 股东权益合计(不含少数股东权益)-最新财报	升序
质量	BerryRatio	(营业收入-TTM - 营业成本-TTM) / 销售费用-TTM	降序
质量	AssetTurnover	营业收入-TTM / 资产总计-最新财报	降序
质量	CurrentRatio	流动资产合计-最新财报 / 流动负债合计-最新财报	降序
质量	EPS_FY0-Dispersion	预测每股收益-最近预测年度标准差 / 预测每股收益-最近预测年度	升序
动量	Momentum-1M	复权收盘价 / 复权收盘价-1月前 - 1	升序
动量	Momentum-3M	复权收盘价 / 复权收盘价-3月前 - 1	升序
动量	Momentum-12M	复权收盘价 / 复权收盘价-12月前 - 1	降序
动量	Momentum-12M-1M	(复权收盘价 / 复权收盘价-12月前 - 1) - (复权收盘价 / 复权收盘价-1月前 - 1)	降序
动量	Momentum-60M	复权收盘价 / 复权收盘价-60月前 - 1	升序
情绪	EPSChange_FY0-1M	预测每股收益-最近预测年度 1 个月的变化	降序
情绪	EPSChange_FY0-3M	预测每股收益-最近预测年度 3 个月的变化	降序
情绪	CFPSChange_FY0-3M	预测每股现金流-最近预测年度 3 个月的变化	降序
情绪	DPSChange_FY0-1M	预测每股股利-最近预测年度 1 个月的变化	降序
情绪	DPSChange_FY0-3M	预测每股股利-最近预测年度 3 个月的变化	降序
情绪	ROEChange_FY0-1M	预测净资产收益率-最近预测年度 1 个月的变化	降序
情绪	ROEChange_FY0-3M	预测净资产收益率-最近预测年度 3 个月的变化	降序
情绪	EarningsChange_FY0-1M	预测净利润-最近预测年度 1 个月的变化	降序
情绪	EarningsChange_FY0-3M	预测净利润-最近预测年度 3 个月的变化	降序
情绪	SalesChange_FY0-1M	预测主营业务收入-最近预测年度 1 个月的变化	降序
情绪	SalesChange_FY0-3M	预测主营业务收入-最近预测年度 3 个月的变化	降序
情绪	RatingChange-3M	分析师综合评级 3 个月的变化	降序
情绪	综合评级分值	分析师综合评级	降序
技术	LnFloatCap	流通市值的自然对数	升序
技术	AmountAvg-1M	过去一个月日均成交额	升序
技术	NormalizedAbnormalVolume	过去一个月日均成交量 / 12 个月日均成交量	升序
技术	TurnoverAvg-1M	过去一个月日均换手率	升序
技术	TurnoverAvg-3M	过去三个月日均换手率	升序
技术	TurnoverAvg-1M-3M	过去一个月日均换手率 / 过去三个月日均换手率	升序
技术	TSKEW	过去一年日收益率数据计算的偏度	升序
技术	ILLIQ	AVERAGE(日涨跌幅绝对值 / 日成交金额)	降序
技术	SmallTradeFlow	月散户量差(仅主动) / 月成交量	升序
技术	MACrossover	(15 周指数移动平均线 - 36 周指数移动平均线) / 36 周指数移动平均线	降序
技术	RealizedVolatility-1Y	过去一年日收益率数据计算的标准差	升序

## 附录 2: 算法流程

1. 计算训练样本集  $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$ , 其中:

$$y_i = \begin{cases} 1, & r_i > \text{median}(\mathbf{r}) \\ -1, & r_i < \text{median}(\mathbf{r}) \end{cases}, \quad \mathbf{r} = (r_1, K, r_i, K, r_N), \quad r_i \text{ 是股票 } i \text{ 下一期收益率};$$

$\mathbf{x}_i = (x_{i1}, K, x_{ik}, K, x_{iK})$ ,  $x_{ik}$  是股票  $i$  在因子  $k$  上正规化后的暴露,  $N$  为股票数,

$K$  为因子数, 初始化权重向量为等权向量  $\mathbf{w} = (w_1, K, w_i, K, w_N)$ ,  $w_i = \frac{1}{N}$ 。

2. 构建第  $l=1, K, L$  层弱分类器:

- 1) 按照因子  $k$  的值, 以其分位数将训练样本分成  $Q$  组:  $X_1, K, X_j, K, X_Q$ ,  
在每一个分位数组里, 分别计算强势子和弱势子的总权重:

$$W_k^{j\pm} = \sum_{y_i = \pm 1, x_{ik} \in X_j} w_i$$

- 2) 计算该因子的区分度指标:

$$Z_l^k = f_d(W_k^+, W_k^-)$$

其中,  $f_d$  是区分度计算函数。

- 3) 对于所有的因子, 比较区分度指标的大小, 选择使该指标为最小的因子生成该层的弱分类器, 即:

$$h_l(x) = h_k(x), k = \arg \min_{k \in F} \{Z_l^k\}$$

$$h_k(x) = f_w(W_k^{j+}, W_k^{j-})$$

其中,  $f_w$  是弱分类器的得分函数。

- 4) 更新权重:

$$w_i = g_w(w_i, y_i, h)$$

其中,  $g_w$  是权重的调整函数。将权重向量  $\mathbf{w}$  归一化。

3. 最终的强分类器定义为所有弱分类器的和:

$$H(x) = \sum_{l=1}^L h_l(x)$$

注: 本文中, 我们弱分类器的个数设为 12 个, 分为数组的个数设为 5 个。

## 投资评级说明

**行业评级** 报告发布日后的12个月内行业股票指数的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

- 推荐: 相对表现优于市场;  
中性: 相对表现与市场持平  
回避: 相对表现弱于市场

**公司评级** 报告发布日后的12个月内公司的涨跌幅度相对同期上证综指/深圳成指的涨跌幅为基准,投资建议的评级标准为:

- 买入: 相对大盘涨幅大于15%;  
增持: 相对大盘涨幅在5%~15%之间  
中性: 相对大盘涨幅在-5%~5%;  
减持: 相对大盘涨幅小于-5%

## 机构销售经理联系方式

机构销售负责人			邓亚萍	021-38565916	dengyp@xyzq.com.cn
上海地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
罗龙飞	021-38565795	luolf@xyzq.com.cn	盛英君	021-38565938	shengyj@xyzq.com.cn
杨 忱	021-38565915	yangchen@xyzq.com.cn	王 政	021-38565966	wangz@xyzq.com.cn
冯 诚	021-38565411	fengcheng@xyzq.com.cn	王 溪	021-20370618	wangxi@xyzq.com.cn
顾 超	021-20370627	guchao@xyzq.com.cn	李远帆	021-20370716	liyuanfan@xyzq.com.cn
地址: 上海市浦东新区民生路1199弄证大五道口广场1号楼20层(200135) 传真: 021-38565955					
北京地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
朱圣诞	010-66290197	zhusd@xyzq.com.cn	李 丹	010-66290223	lidan@xyzq.com.cn
肖 霞	010-66290195	xiaoxia@xyzq.com.cn	郑小平	010-66290223	zhengxiaoping@xyzq.com.cn
刘晓浏	010-66290220	liuxiaoliu@xyzq.com.cn	吴 磊	010-66290190	wulei@xyzq.com.cn
何 嘉	010-66290195	hejia@xyzq.com.cn	陈 杨	010-66290195	chenyangjg@xyzq.com.cn
地址: 北京市西城区武定侯街2号泰康国际大厦6层609(100033) 传真: 010-66290200					
深圳地区销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
朱元贱	0755-82796036	zhuyy@xyzq.com.cn	李 昇	0755-82790526	lisheng@xyzq.com.cn
杨 剑	0755-82797217	yangjian@xyzq.com.cn	邵景丽	0755-23836027	shaojingli@xyzq.com.cn
王维宇	0755-23826029	wangweiyu@xyzq.com.cn			
地址: 福田区中心四路一号嘉里建设广场第一座701(518035) 传真: 0755-23826017					
海外销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
刘易容	021-38565452	liuyirong@xyzq.com.cn	徐 皓	021-38565450	xuhao@xyzq.com.cn
张珍岚	021-20370633	zhangzhenlan@xyzq.com.cn	陈志云	021-38565439	chanchiwan@xyzq.com.cn
曾雅琪	021-38565451	zengyaqi@xyzq.com.cn	龚学敏	021-38565982	gongxuemin@xyzq.com.cn
地址: 上海市浦东新区民生路1199弄证大五道口广场1号楼20层(200135) 传真: 021-38565955					
私募及企业客户负责人			刘俊文	021-38565559	liujw@xyzq.com.cn
私募销售经理					
姓名	办公电话	邮 箱	姓名	办公电话	邮 箱
徐 瑞	021-38565811	xur@xyzq.com.cn	杨雪婷	021-20370777	yangxueting@xyzq.com.cn
地址: 上海市浦东新区民生路1199弄证大五道口广场1号楼20层(200135) 传真: 021-38565955					



**【信息披露】**

本公司在知晓的范围内履行信息披露义务。客户可登录 [www.xyzq.com.cn](http://www.xyzq.com.cn) 内幕交易防控栏内查询静默期安排和关联公司持股情况。

**【分析师声明】**

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

**【法律声明】**

兴业证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告仅供兴业证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。客户应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司 <http://www.xyzq.com.cn> 网站刊载的完整报告为准，本公司接受客户的后续问询。

本公司的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。

本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

本公司系列报告的信息均来源于公开资料，我们对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。

在法律许可的情况下，兴业证券股份有限公司可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到兴业证券股份有限公司及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

若本报告的接收人非本公司的客户，应在基于本报告作出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。未经授权的转载，本公司不承担任何转载责任。