

不要迷恋 Alpha

报告要点

■ 是否真的能产生 Alpha?

我们研究发现,很多貌似产生 Alpha 的量化方法找到的其实只是一些市场风格因子的超额收益,如果用多因子模型来归因的话,会发现这些股票其实并没有战胜它们的同类,而只是“幸运”的处在了某一个市场阶段的热门风格中。

■ 怎么用数量化的方法寻找 Alpha?

如果用静态的方法把 Alpha 视为一个“黑箱”,并不能很好的获得超额收益,很容易陷入数量化方法的陷阱。因此我们更倾向于用因子的方法分解收益,并且监控这些因子的变化。

■ 因子的选择

因子的选择必须考虑足够的维度,避免共线性,当然还需要有经济学意义。在这里我们选择了沪深 300,大小盘和高低估值三个因子。

■ 收益表现

该模型的收益率取决于对 Beta 和因子的预测,在完美的预测下,当然可以达到惊人的收益。而在实际应用中,必须对两个决定性要素进行预测,这是我们随后的报告将要深入讨论的内容。

分析师

俞文冰

(8621) 021-68751782

[yuwb@cjsc.com.cn](mailto:yuw@cjsc.com.cn)

执业证书编号: S0490209100276

联系人:

俞文冰

(8621) 021-68751782

[yuwb@cjsc.com.cn](mailto:yuw@cjsc.com.cn)

正文目录

一、Alpha 真的存在么？	3
1、貌似 Alpha 的超额收益	3
2、关于数量化方法寻找 Alpha 的思考	5
二、用多因子模型解释超额收益	5
1、超额收益的来源	6
2、选择合适的因子	6
3、更高的拟合效果，更低的 Alpha	7
4、因子的可预测性	9
三、理想境界	11

图表目录

图 1：一个有超额收益的组合表现	3
图 2：单因子模型和三因子模型 R-Square 直方图比较	8
图 3：单因子模型和三因子模型 Alpha 直方图比较	8
图 4：三因子 Beta 散点图	9
图 5：中证 500 相对中证 100 累计超额收益率图	10
图 6：低市净率指数相对高市净率指数累计超额收益图	10
图 7：“完美预测”下三因子模型收益图	11
图 8：“完美”预测下三因子模型模型超额收益图	12
表 1：单因子归因一个有超额收益的组合	3
表 2：三因子归因一个有“超额收益”的组合	4
表 3：因子相关性表	7

一、Alpha 真的存在么？

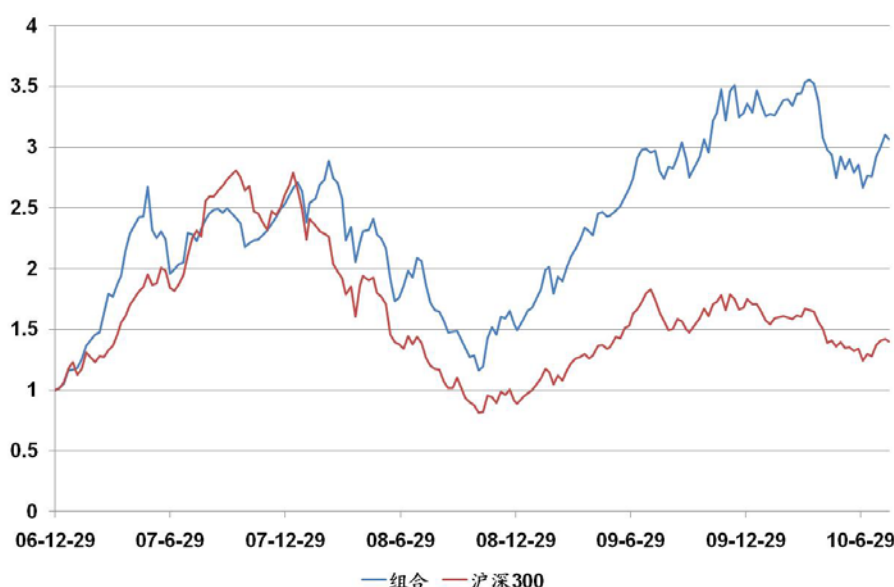
在我们做数量化研究的过程中，数量化选股总是一个很有吸引力的题目，大家都希望能找到某一个或者一组指标，可以筛选出长期跑赢市场的股票。那么是不是真的能找到这样的指标呢？或者说这样的方法适用性到底有多大呢？

其实这是一个很难证伪的命题，因为哪怕我们可以证明 N 种选股模型或者指标无效，也不能证明这个方法完全无效。但是从目前来看，选股策略带来的超额收益是有局限性的。

1、貌似 Alpha 的超额收益

假如某天，我们通过指标筛选和实证检验，找到了一个投资组合，其表现如图 1：

图 1：一个有超额收益的组合表现



资料来源：Wind 资讯，长江证券研究部

从图中我们发现这个组合从 07 年开始一直到 2010 年中期都相当不错，稳定跑赢大盘。如果我们对其中的股票进行单因子(沪深 300)的归因分析，发现组合的 Beta 是在 1 附近的，而平均的年化 Alpha 超过 8%。

表 1：单因子归因一个有超额收益的组合

交易代码	股票简称	年化 Alpha	Beta
000925	众合机电	24.5583%	0.8670
600093	禾嘉股份	35.9479%	0.9704
600553	太行水泥	21.3387%	1.2273
600074	中达股份	-2.9525%	0.9752
000909	数源科技	11.3182%	1.1101

000532	力合股份	9.9229%	1.1123
600108	亚盛集团	5.7168%	0.9612
000665	武汉塑料	11.5995%	1.0450
600856	长百集团	13.7033%	1.2808
600746	江苏索普	2.8638%	1.0647
600745	中茵股份	3.4202%	0.9909
600731	湖南海利	22.8887%	1.0024
600185	格力地产	4.1822%	1.1596
000663	永安林业	15.2237%	1.0483
600734	实达集团	-33.6115%	1.0029
600248	延长化建	6.7971%	1.1492
000150	宜华地产	8.2883%	1.2240
600165	宁夏恒力	8.8051%	0.9855
000975	科学城	2.2180%	1.0286
600225	天津松江	-2.5417%	1.1444
均值		8.4844%	1.0675

注：收益率以周为单位

资料来源：Wind 资讯，长江证券研究部

看上去是一个不错的结果，但是如果我们做一个三因子的模型来分解收益，如表 2 所示：

表 2：三因子归因一个有“超额收益”的组合

交易代码	股票简称	单因子模型		多因子模型			
		年化 Alpha	Beta	年化 Alpha	市场因子	大小盘因子	高低估值因子
000925	众合机电	24.5583%	0.8670	-7.0722%	0.8670	-1.4703	-0.5910
600093	禾嘉股份	35.9479%	0.9704	-0.8227%	0.8605	-1.2803	-0.3898
600553	太行水泥	21.3387%	1.2273	-0.9431%	1.1283	-1.0140	-0.6511
600074	中达股份	-2.9525%	0.9752	-10.6612%	0.8582	-1.0030	-0.6284
000909	数源科技	11.3182%	1.1101	-5.1756%	1.0183	-1.1414	-0.4609
000532	力合股份	9.9229%	1.1123	-4.9990%	1.0389	-1.2542	-0.3305
600108	亚盛集团	5.7168%	0.9612	-6.5441%	0.9154	-1.1454	-0.4356
000665	武汉塑料	11.5995%	1.0450	-9.1756%	0.9268	-0.9233	-0.6504
600856	长百集团	13.7033%	1.2808	-8.4538%	1.1003	-1.0392	-0.5325
600746	江苏索普	2.8638%	1.0647	-2.2495%	0.9396	-0.9770	-0.5913
600745	中茵股份	3.4202%	0.9909	-13.3286%	0.8684	-0.8482	-0.7151
600731	湖南海利	22.8887%	1.0024	-4.8014%	0.8553	-1.0387	-0.5243
600185	格力地产	4.1822%	1.1596	-5.4014%	1.0703	-0.7256	-0.8266
000663	永安林业	15.2237%	1.0483	-1.6289%	0.9873	-0.8875	-0.6605

600734	实达集团	-33.6115%	1.0029	-46.8496%	0.9820	-1.1159	-0.4204
600248	延长化建	6.7971%	1.1492	-0.5736%	0.9328	-0.9790	-0.5469
000150	宜华地产	8.2883%	1.2240	-6.0976%	1.0800	-0.8865	-0.6378
600165	宁夏恒力	8.8051%	0.9855	-5.4609%	0.9016	-1.0679	-0.4517
000975	科学城	2.2180%	1.0286	-2.6807%	0.9560	-1.1414	-0.3718
600225	天津松江	-2.5417%	1.1444	-0.8467%	1.1351	-0.5709	-0.9416
均值		8.4844%	1.0675	-7.1883%	0.9711	-1.0255	-0.5679

资料来源：Wind 资讯，长江证券研究部

我们发现三因子模型的结果与单因子有很大的区别，用三因子归因以后，发现这个组合的 Alpha 其实是负的。同时这个组合的股票在大小盘因子和高低估值因子上都有很明显的负的 Beta。这说明这些股票都是典型的小盘+低估值的股票，但是组合中的股票并没有达到同类型股票的平均水平。

因此，这个组合的“超额收益”实际上来自于小盘+低估值的风格属性，而无巧不巧的是，在样本模拟期内，这个风格的确表现出了明显的强势。而如果我们用这个时间段作为建模区间，去筛选财物或者估值指标，很容易就找出类似的模型，但是这样的模型其实并不靠谱，在其他时间，这个组合并没有表现出明显的超额收益。

2、关于数量化方法寻找 Alpha 的思考

正是由于上述的顾虑，我们就要反思用数量化的方法寻找超额收益的途径。显然数量化研究不同于行业研究员的基本面研究，我们无法调研，拿不到除了定期报表之外的任何信息。而显然报表的短期效应在其公布之前和公布之后的很短时间很有可能会得到大部分甚至全部的体现，这就使得数量化筛选基本面指标面临严重的滞后问题。

当然数量化分析还有一个优势，就是可以建立模型，对全市场进行横向和纵向的比较，以此寻找市场可能出现的“被忽略”的机会。

但是这个方法依然是存在隐患的，毕竟量化分析的根基是实证，实证就有样本数据依赖的问题，而样本外数据规律是否能够保持永远是量化分析无法解答的问题。就算样本外的数据依然遵循某些规律，会实现其蕴含的投资价值，其周期依然是未知数。就目前的资本市场而言，超过半年的“价值发现”周期可能是大部分资金很难忍受的。

二、用多因子模型解释超额收益

正是因为上述的这些原因，我们一直在思考，到底超额收益从何处来？或者应该用什么样的数量化方法去获得超额收益？

1、超额收益的来源

在我们做研究的过程中，我们会倾向于用更多的金融学中定义的收益异常或者其他市场无效的方法来寻找超额收益，比如：小盘股效应，低 PE 效应，月份效应等等。

当然还有一种方法是，我们去寻找一些数据上有规律，原理上似乎有道理的模式，比如：低市盈率、高现金流、高分红的股票有吸引力。

但是问题是，我们其实并不知道，这真的是会带来超额收益，还只是市场在某一个阶段对某一个类型或者风格的股票的宠爱。更无法知道这些定价偏差会不会消失，何时消失。

换一个角度来想，我们把小盘溢价，低估值溢价等分别用大小盘因子和高低估值因子来刻画，如果小盘股溢价或者低估值溢价出现的话，就意味着这些因子都会有相应趋势性的表现。而如果这些溢价现象是间歇性的出现的话，那么这些因子就会出现回复性质的表现。

相比简单的静态模型，这个方式可以更清楚的监控因子的变化，更明确的找到超额收益的来源。

那么剩下的事情就是要找到合适的因子。

2、选择合适的因子

在寻找因子的时候我们必须考虑以下一些因素：

(1) 因子的维度：我们在寻找多因子的时候，这些因子最好要包容市场尽可能多的维度。比如我们如果从来没有考虑过大小盘的维度，那么在 2009-2010 年的时候就会非常的被动。当然这些维度最好能包容足够多的市场情况，或者我们可以在一种新的市场现象出现的时候尽快的添加这一维度的因子。

(2) 共线性。显然这是困扰所有统计模型的一个问题。而且我们不推荐用主成分回归等方法去处理这个问题，如果你不想加工出一堆自己都解释不清楚的东西的话。而且这个问题是和因子的维度缠绕出现的。因为某几个因子在某一段时间内可能出现很强的共线性，但是也许在未来他们又会出现差异。在足够多的维度，以及较低的共线性以维持统计意义之间，我们需要做一个好的平衡。

(3) 经济意义云云，参见任意一篇定量分析的经济类论文。

我们找出通常会考虑的几个因子，观察他们的相关性，如表 3 所示：

我们会发现，其中中证 500-中证 100 和 上证周期-上证非周期，这两个因子之间的相关性非常高。其实我们如果仔细看他们的成分股，会发现中证 100 和上证周期重合度很高。因此他们之间的高度相关性是必然的，这两个因子也可以认为可以互相替代。

剩下的一堆因子，互相之间相关关系也很明显，我们最后从中挑出了高低 PB 因子，该因子与中证 500-中证 100 的相关度不高，但是与价值成长，高低 PE 之间的相关性非常高。之所以选择了高低 PB 而非高低 PE，也是考虑可能净资产的数据会比盈利更稳定，划分效果更明显。

当我们找出最后的三因子：沪深 300，中证 100-中证 500(大小盘)，高市净率-低市净率，我们发现其实基本上就是 Fama 三因子模型。

表 3：因子相关性表

	沪深 300	中证 500-100	周期- 非周期	中证 100-中 小板	价值- 成长	高 PB- 低 PB	活跃股- 沪深 300	大盘指数 -小盘指 数	高 PE- 低 PE
沪深 300	1.0000	-0.1852	-0.1678	-0.0708	0.1493	0.1455	0.1187	0.1590	-0.0393
中证 500-100	-0.1852	1.0000	-0.6446	-0.0646	-0.0923	0.1230	0.0344	0.0549	0.1212
周期- 非周期	-0.1678	-0.6446	1.0000	0.2244	0.0639	-0.0797	-0.1613	-0.0160	-0.1836
中证 100- 中小板	-0.0708	-0.0646	0.2244	1.0000	0.8308	0.8044	-0.7780	0.8532	-0.8236
价值- 成长	0.1493	-0.0923	0.0639	0.8308	1.0000	0.7970	-0.5932	0.7741	-0.6902
高 PB- 低 PB	0.1455	0.1230	-0.0797	0.8044	0.7970	1.0000	-0.5842	0.8975	-0.6380
活跃股- 沪深 300	0.1187	0.0344	-0.1613	-0.7780	-0.5932	-0.5842	1.0000	-0.6546	0.6048
大盘- 小盘	0.1590	0.0549	-0.0160	0.8532	0.7741	0.8975	-0.6546	1.0000	-0.7030
高 PE- 低 PE	-0.0393	0.1212	-0.1836	-0.8236	-0.6902	-0.6380	0.6048	-0.7030	1.0000

注 1：因子算法是用两类指数的周收益率相减；

注 2：相关系数计算的是 2006 年 1 月 1 日到 2010 年中

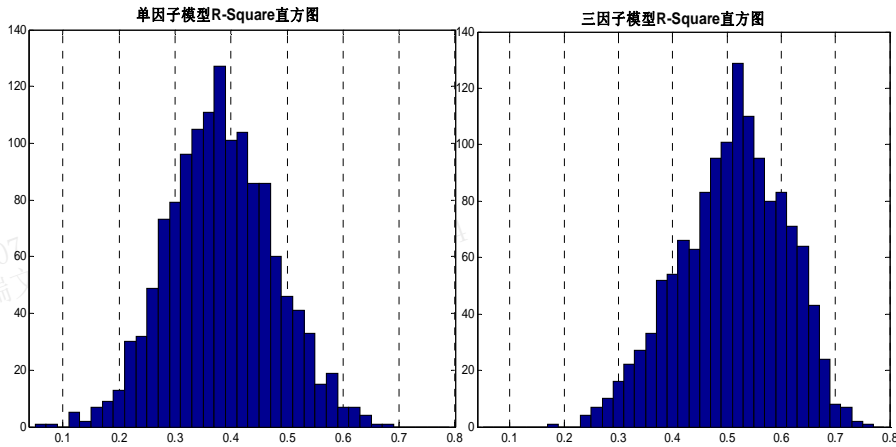
资料来源：Wind 资讯，长江证券研究部

3、更高的拟合效果，更低的 Alpha

那么我们费尽心思找出来的三因子，究竟可以带来什么呢？

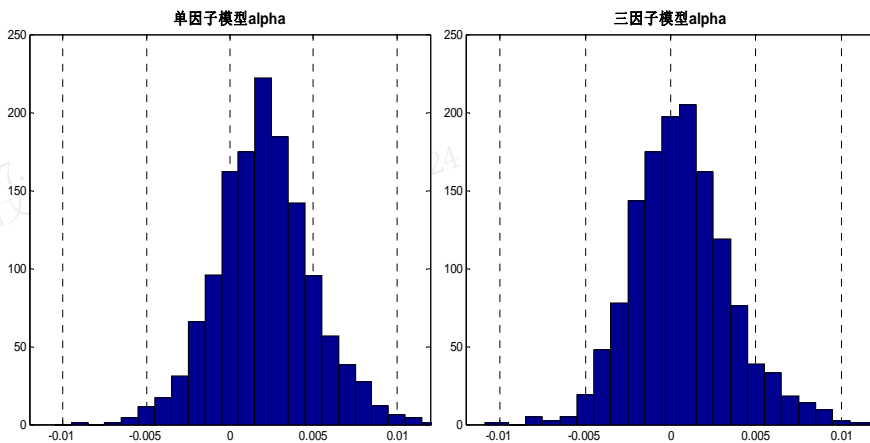
我们对所有股票分别做了单因子和三因子模型，然后比较它们的 R-Square 和 Alpha，如图 2 和图 3 所示：

图 2：单因子模型和三因子模型 R-Square 直方图比较



资料来源：Wind 资讯，长江证券研究部

图 3：单因子模型和三因子模型 Alpha 直方图比较



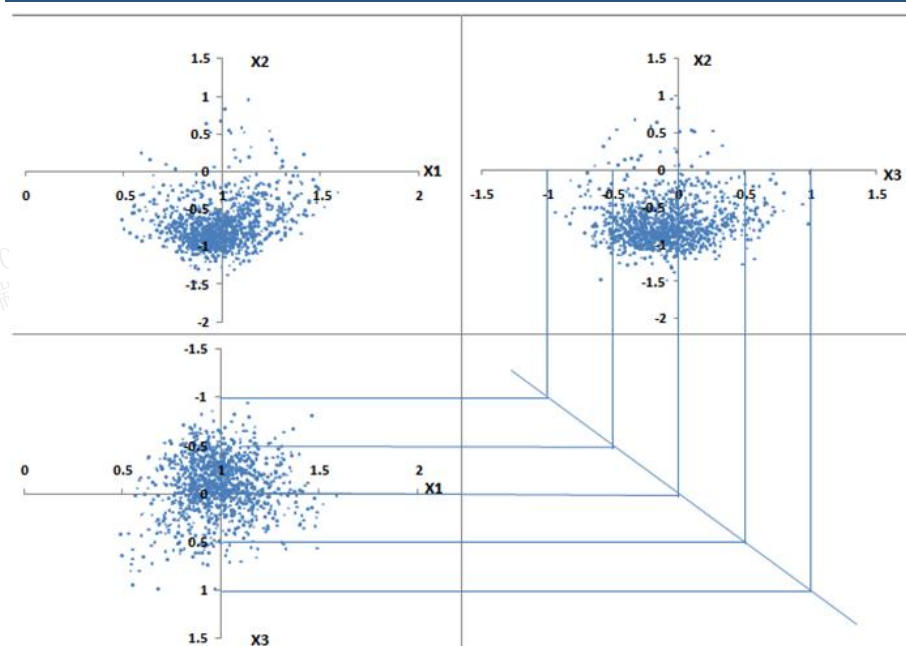
资料来源：Wind 资讯，长江证券研究部

显然从图中可以看出，对于绝大部分股票三因子模型的 R-Square 都有较大幅度提高，同时 Alpha 有比较明显的下降。这就说明很多单因子模型原来无法解释的超额收益被三因子模型解释了。

如果我们观察三个因子的 Beta，如图 4 所示，可以发现三因子的 Beta 基本上都是散布的。这说明三个因子之间的相关性很弱，或者说在三个坐标轴划分的八个象限中，我们都可以根据我们的需求任意的找出我们想要的股票。

当然很明显的发现 X2(大小盘因子)上，Beta 小于 0 的要远远多于 Beta 大于 0 的，这说明小盘股的数量还是要远远多于大盘股。

图 4: 三因子 Beta 散点图



资料来源: Wind 资讯, 长江证券研究部

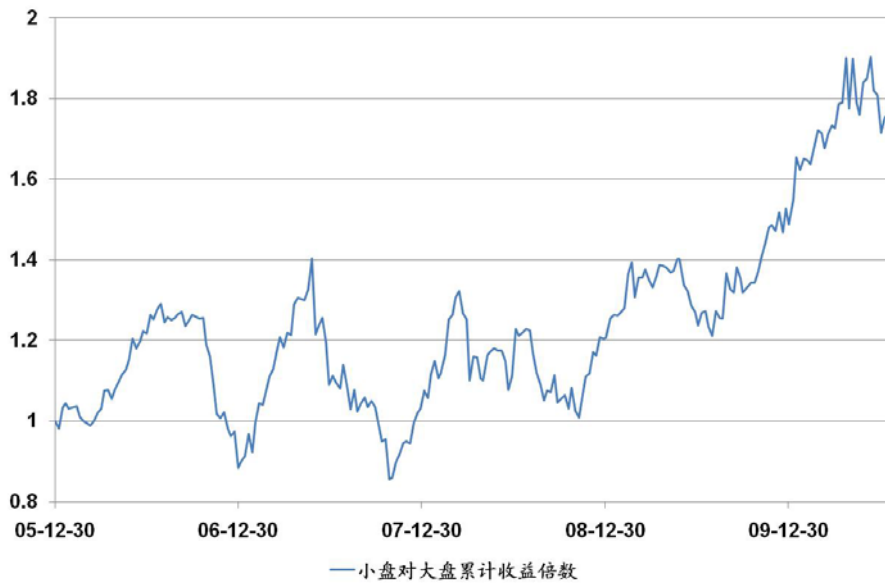
4、因子的可预测性

如果我们只是做一下收益率分解, 那么其实我们是在做绩效归因。但是我们的实际目的是要用这个思想来做数量化投资, 所以必然要遇到的问题是, 我们希望我们的这两个因子可预测性强一点。关于因子预测的问题, 我们在之后的报告中还会详细阐述, 在这里我们先看一下因子的大致行为特征。

第一个因子是沪深 300, 这是大家都在做的市场判断, 我们也有自己的趋势跟踪和上升指数的方法进行判断, 这里不做赘述了。

第二个因子是大小盘因子。简单的看, 我们可以用图 5 中中证 500 相对于中证 100 的累计超额收益(计算方法: 在原点归一, 然后用中证 500 的点位除以中证 100 的点位)来表达。我们可以看到, 在 2009 年底之前, 这个规律一直是比较规律的均值回复特征。但是从 2009 年四季度开始, 忽然呈现出一种明显的强趋势。其实这一现象在 1999-2001 年也发生过, 只不过当时中证系列指数还没有公布, 我们没有办法列出。对于这个因子, 我们更倾向于用趋势或者类似趋势的方法来预测, 可以得到较高的准确度。如果用均值回复, 可能会犯比较大的错误。

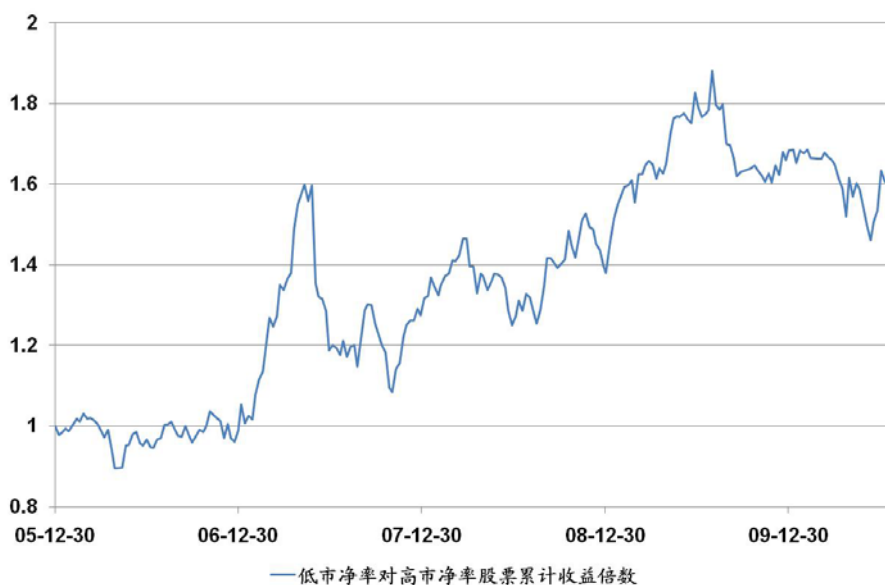
图 5：中证 500 相对中证 100 累计超额收益率图



资料来源：Wind 资讯，长江证券研究部

在图 6 中我们可以看到，高低市净率风格的切换要比大小盘难预测的多，之间的纠缠和震荡也更多。

图 6：低市净率指数相对高市净率指数累计超额收益图



资料来源：Wind 资讯，长江证券研究部

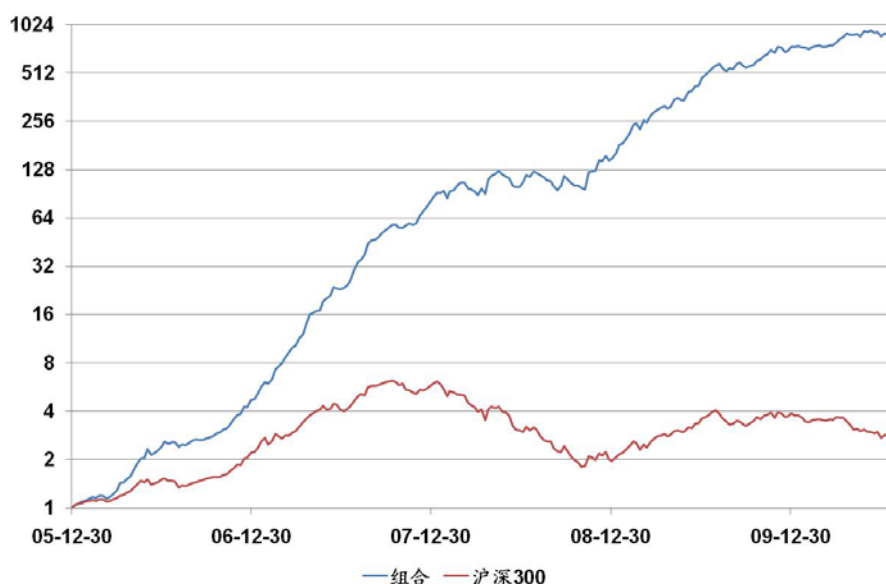
三、理想境界

如果三因子(或者多因子)模型可以用来分析股票的收益来源,那么我们有没有可能用这个方法在投资中获得满意的收益呢?我们通过这个三因子模型,对股票打分,选择得分最高的N个股票构建组合。

我们在以下假设条件下做模拟:

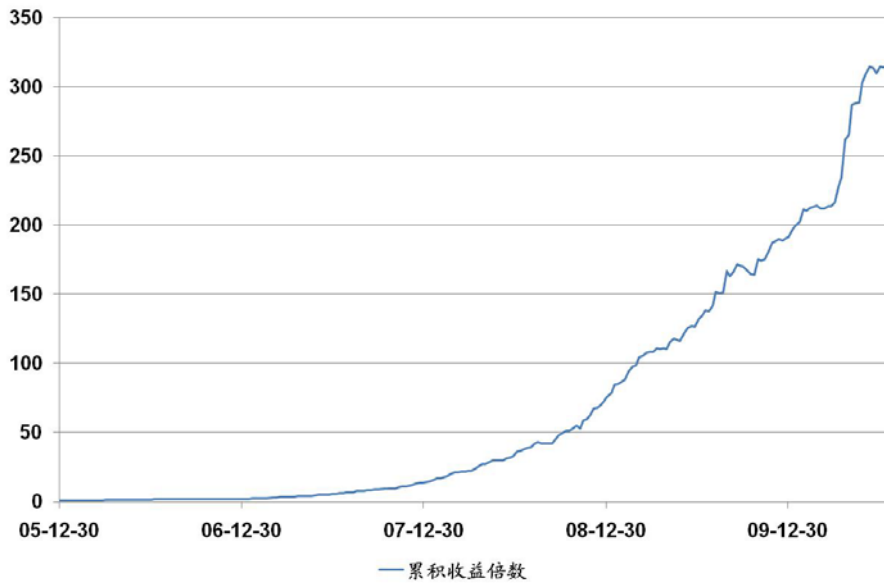
- (1) 模拟时间: 2006年1月-2010年8月;
- (2) 周策略;
- (3) Beta的估计: 所有A股对上述三个因子的全部历史数据普通最小二乘回归,这里会有“未来信息”流入;
- (4) 因子每周的走势都可以完美预测,这里有很多“未来信息”混入;
- (5) 动态的构建一个包含20只股票的组合,我们也尝试了其他组合数量,效果相仿;
- (6) 构建了40只股票的缓冲池,周换仓策略,加上不低的交易成本假定,会很大的影响收益,因此我们建立“缓冲池”的概念。加入我们每周选出前20只股票进入组合,剩下的21-40名的股票就进入缓冲池。如果上一周组合内的股票,到了本周不在前20,但是在缓冲池中,我们依然保留在组合中,如果掉出缓冲池,我们就剔除组合。然后选择新的综合得分最高的股票纳入组合。
- (7) 交易成本按1%计算

图 7: “完美预测”下三因子模型收益图



资料来源: Wind 资讯, 长江证券研究部

图 8: “完美”预测下三因子模型模型超额收益图



资料来源: Wind 资讯, 长江证券研究部

非常不幸(还是幸运), 这个收益率在实际中是达不到的, 因为我们假定了 Beta 和因子的“完美预测”。那么实际中能达到多高的收益呢?

这取决于 Beta 和因子的预测。我们在随后的报告中将展开详细的讨论。

分析师介绍

俞文冰, CFA, 同济大学数学系统计学专业本科, 上海财经大学统计学研究生, 曾在中海基金负责产品设计, 风险控制和金融工程研究工作, 目前长江证券研究部担任金融工程资深研究员。

对本报告的评价请反馈至长江证券机构客户部

姓名	分工	电话	E-mail
伍朝晖	副主管	(8621) 68752398	13564079561 wuzh@cjsc.com.cn
甘露	华东区总经理	(8621) 68751916	13701696936 ganlu@cjsc.com.cn
王磊	华东区总经理助理	(8621) 68751003	13917628525 wanglei3@cjsc.com.cn
鞠雷	华南区副总经理	(8621) 68751863	13817196202 julei@cjsc.com.cn
程杨	华北区总经理助理	(8621) 68753198	13564638080 chengyang1@cjsc.com.cn
李劲雪	上海私募总经理	(8621) 68751926	13818973382 lijx@cjsc.com.cn
张晖	深圳私募总经理	(0755) 82766999	13502836130 zhanghui1@cjsc.com.cn
沈方伟	深圳私募总经理助理	(0755) 82750396	15818552093 shenfw@cjsc.com.cn

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沪深 300 指数的涨跌幅为基准, 投资建议的评级标准为: 看好: 相对表现优于市场 中性: 相对表现与市场持平 看淡: 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沪深 300 指数的涨跌幅为基准, 投资建议的评级标准为: 推荐: 相对大盘涨幅大于 10% 谨慎推荐: 相对大盘涨幅在 5% ~ 10% 之间 中性: 相对大盘涨幅在 -5% ~ 5% 之间 减持: 相对大盘涨幅小于 -5% 无投资评级: 由于我们无法获取必要的资料, 或者公司面临无法预见结果的重大不确定性事件, 或者其他原因, 致使我们无法给出明确的投资评级。

重要声明

长江证券系列报告的信息均来源于公开资料, 本公司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。本公司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, 不代表对证券价格涨跌或市场走势的确定性判断。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。

本公司及作者在自身所知情的范围内, 与本报告中所评价或推荐的证券没有利害关系。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头寸并进行交易, 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。

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、复制和发布。如引用、刊发, 需注明出处为长江证券研究部, 且不得对本报告进行有悖原意的引用、删节和修改。