

## 量化选股的线性回归体系构建（二） --因子牛熊市区别

### 报告要点

#### 1、研究目的

前一篇报告我们使用 Fama-MacBeth 来建立多因子模型，不过没有区分牛熊市下的表现。正如 CAPM 模型中的 beta 因子，在牛市中高 beta 能获得超额收益，而在熊市中低 beta 股票表现更为出色，投资者在牛熊市中的选股逻辑是有所区别的，所以我们想知道其他因子在牛熊市下的不同表现。

#### 2、牛熊划分

我们使用长江趋势模型来识别牛熊市，主要步骤包括高低点识别和趋势生成。

#### 3、单因子牛熊检验

基于历史数据的估值因子都是在牛市更加显著；

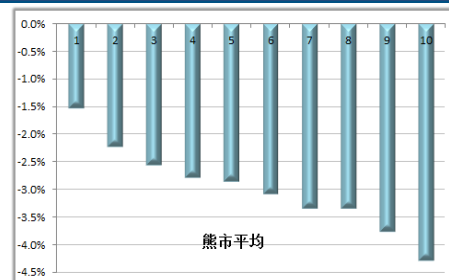
预期因子也是在牛市表现比熊市更好；

增速类的因子在牛市表现更好；

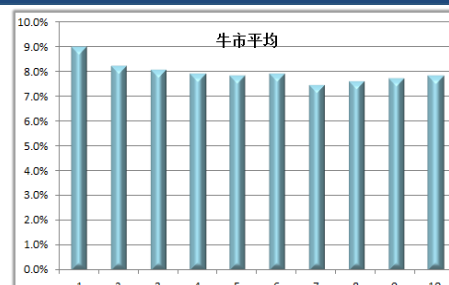
市值因子在熊市表现比牛市更加显著；

熊市中现金比率、偿债能力等因子会体现出选股效果，而牛市中则基本没有。

熊市分组收益（总市值从小到大）



牛市分组收益（总市值从小到大）



相关报告：

[基本面选股因子的单因素效果测试之一](#)

[基本面选股因子的单因素效果测试之二](#)

[基本面选股因子的单因素效果测试之三](#)

[融资融券标的中的有效选股因子](#)

[量化选股的线性回归体系构建（一）](#)

分析师

范辛亭

021-68751859

fanxt@cjsc.com.cn

执业证书编号：S0490510120008

联系人：

袁继飞

021-68751787

yuanjf@cjsc.com.cn

## 正文目录

1. 因子存在极性么 .....	4
2. 牛熊市的划分 .....	4
2.1. 长江趋势模型 .....	4
3. 不同维度的趋势 .....	5
3.1. 样本牛熊划分 .....	5
4. 回归因子模型回顾 .....	6
4.1. 统计工具 .....	6
4.2. 回归的优缺点 .....	7
4.3. 检查的因子 .....	7
4.4. 涉及的细节与处理 .....	7
5. 单变量牛熊检验 .....	8
5.1. 排序与回归的对比 .....	8
5.2. 单因子牛熊市检验值比较 .....	9
6. 多因子组合牛熊检验 .....	10
6.1. 整体流程 .....	10
6.2. 牛市组合 .....	11
6.3. 熊市组合 .....	11
6.4. 组合的组合 .....	12
7. 后续研究方向 .....	13

## 图表目录

图 1: 沪深 300 高低点识别 ( $j=7$ ) .....	5
图 2: 沪深 300 牛熊状态 ( $j=7$ ) .....	5
图 3: 熊市分组收益 (总市值从小到大) .....	8
图 4: 牛市分组收益 (总市值从小到大) .....	8
图 5: 组合的组合走势.....	12
图 6: 组合的组合超额收益走势.....	12
表 1: 牛熊市划分.....	6
表 2: 牛市区分度较高因子 .....	9
表 3: 熊市区分度较高因子 .....	9
表 4: 牛熊市均不显著因子 .....	10

## 1. 因子存在极性么

我们在上一篇报告中探讨了用线性回归的方法建立多因子模型的方法,对于市场状态并没有进行划分。我们自然会想,各种因子会不会在不同的市场状态下表现也会有差异呢?我们已经知道 CAPM 模型中的  $\beta$  有这种规律:高  $\beta$  的股票在牛市中表现更好,而低  $\beta$  的股票在熊市中表现更佳。

本篇报告我们尝试分析在牛熊的不同市场环境下,不同因子或者组合的表现是否会有明显的差异,也就是看各种因子是否有极性。

## 2. 牛熊市的划分

### 2.1. 长江趋势模型

我们有一套事后识别趋势的量化办法,这套方法利用事后判断的方法来划分识别时间序列是处于上升序列还是处于下降序列。方法中需要考察一个时间序列中每一个点是否是一段时期内的最大值,或者最小值,用到了未来的数据,所以我们称它为事后识别法。尽管这种方法在当时没办法判断趋势状态,但它可以用于对历史的检验。

#### 2.1.1. 高低点标记

对于任何一个时间点  $t$ , 考察时间序列  $x(t)$  在以  $t$  为中心的前后各  $j$  个时间段内的值,若  $x(t)$  是  $x(t-j), \dots, x(t-1), x(t), x(t+1), \dots, x(t+j)$ , 共  $2j+1$  个值中的最大值(最小值), 则  $x(t)$  标记为高点(低点)。若  $x(t-j), \dots, x(t-1), x(t), x(t+1), \dots, x(t+j)$ , 共  $2j+1$  个值超出了时间序列的范围, 则考虑  $x(t)$  是否二者交集的最大值(最小值)。如, 对于终点  $x(n)$ , 如果  $x(n)$  是  $x(n-j), \dots, x(n)$  共  $j+1$  个值中的最大值(最小值), 则标记为高点(低点)。如果连续出现高点(低点), 之间没有低点(高点), 也没有中间间隔点, 则只保留第一个高点(低点), 其他高点(低点)不作标记。在初步标记的结果中, 如果连续出现两个高点(低点), 之间没有低点(高点), 但是有中间间隔点, 则寻找这两个高点之间的最小值点标记为低点(高点)。

#### 2.1.2. 趋势生成

每一个低点与下一个高点之间(含高点)识别为上升趋势, 每一个高点与下一个低点之间(含低点)识别为下降趋势。

如果起始点是高点或者低点, 则此点包含在已经被识别的下降趋势或者上升趋势中。否则, 则进行起始点处的趋势再加工。方法如下: 寻找距离起始点最近的高点或者低点, 若为高点, 则起始点与此高点之间(包含这两个点)识别为下降趋势; 若为低点, 则起始点与此低点之间(包含这两个点)识别为上升趋势。这样做, 我们倾向于认为起始处的趋势不明确, 把它的状态归到了下一个趋势阶段。

如果终点既非高点也非低点, 则进行终点处的趋势再加工。方法如下: 寻找距离终点最近的高点或者低点, 若为高点, 则此高点与终点之间(含终点)识别为上升趋势; 若为低点, 则此低点与终点之间(含终点)识别为下降趋势。我们这样做是相信趋势还会延续, 因为现实中趋势延续的概率比反转的概率稍大一些。

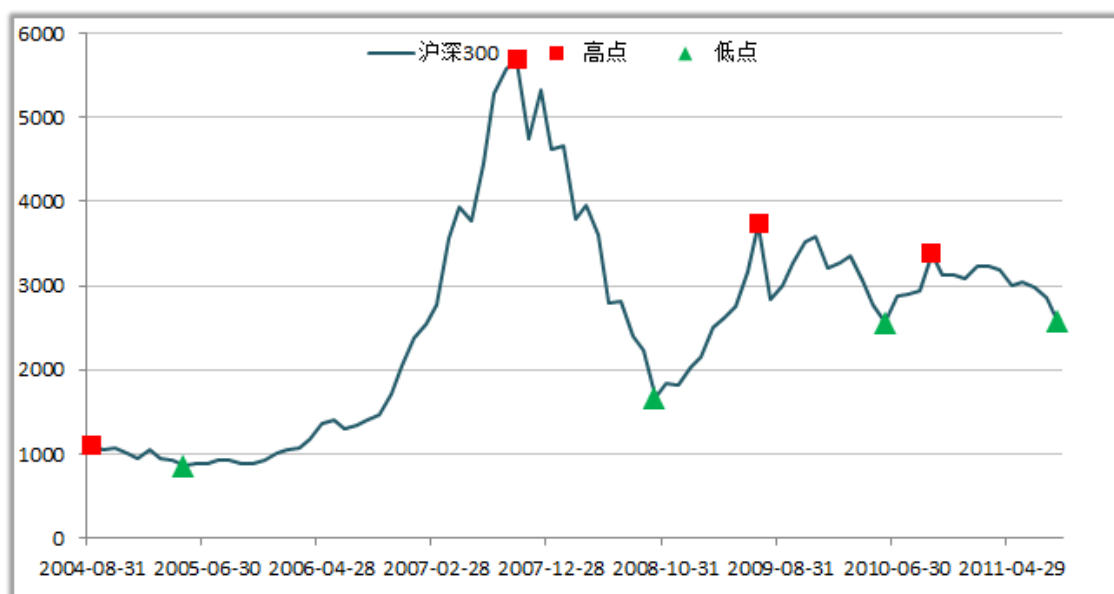
### 3. 不同维度的趋势

我们发现如果  $j$  的长度取得不同，则会识别出不同类型的趋势，取值较大，只识别出长期趋势，取值较小会将短期趋势与长期趋势一并识别出来。不过在我们这个样本中差异不大，所以参数就选了一个普通的 7 个月作为参数。

#### 3.1. 样本牛熊划分

我们将上述趋势识别办法用到沪深 300 指数 2004 年 8 月份以来的月度时间序列上，首先识别高低点。如下图所示：

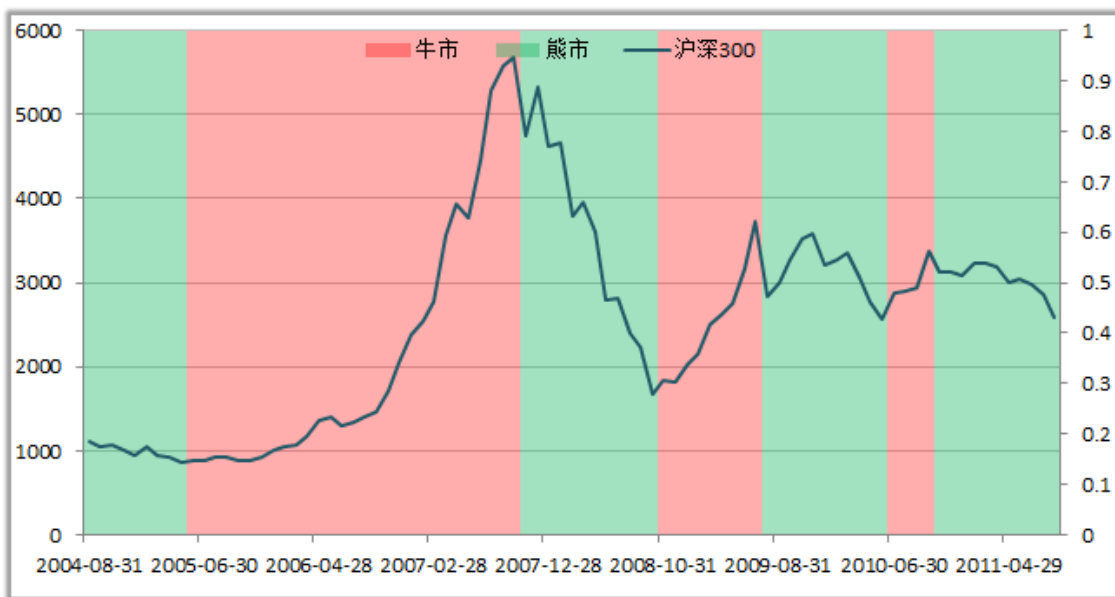
图 1：沪深 300 高低点识别（ $j=7$ ）



资料来源：长江证券研究部。

获得高低点之后，只需要将高低点连起来就可以区分牛熊状态。

图 2：沪深 300 牛熊状态（ $j=7$ ）



资料来源：长江证券研究部。

牛熊市各阶段如下所示：

**表 1：牛熊市划分**

时间段	牛市	熊市
2004-8-31 到 2005-5-31		✓
2005-5-31 到 2007-10-31	✓	
2007-10-31 到 2008-10-31		✓
2008-10-31 到 2009-7-31	✓	
2009-7-31 到 2010-6-30		✓
2010-6-30 到 2010-10-29	✓	
2010-10-29 到 2011-9-30		✓

资料来源：天软科技，长江证券研究部

## 4. 回归因子模型回顾

### 4.1. 统计工具

我们使用的工具是 Fama-MacBeth 检验。Fama-MacBeth 回归方法是 1973 年 Eugene F. Fama 和 James D. MacBeth 在检验 CAPM 模型时提出的一种回归方法，能够降低组合内差异的影响而捕捉组合之间的特征。其主要步骤分为两阶段：1) 根据横截面数据估计单个因子的回归系数；2) 从时间维度计算回归系数序列 t 统计量。本文为了检验各因子对下一个月收益的解释，即比较横截面上因子的差异，因此采用这种方法能够很好的发现各种量化指标的统计显著性。

在每一期（这里指每个月底），我们用所有股票的收益率对检验的因子进行回归：

$$y_{t+1} = a_t + b_t x_t + e_t$$

这里的  $y_{t+1}$  是  $t+1$  期的股票收益率， $x_t$  表示第  $t$  期末尾的因子值。

在得到每期的回归系数后，我们可以做 Fama-MacBeth 检验：

$$t(b_i) = \frac{\mu(b_i)}{\sigma(b_i)} \sqrt{T}$$

其中  $T$  是指时间长度，我们这里是按照月度调仓，所以  $T$  就指整个样本的长度。

## 4.2. 回归的优缺点

回归的方法优点很多，使用回归的方法构建多因子模型的优点可以归结为以下几条：

- 1、 在多因子的建立过程中对比信息更方便，因为多因子回归的时候能够控制一个变量来比较其他变量的影响是否显著，这样就能发现在控制某些因子之后一些显著的因子，也可以对比筛选一些比较类似的指标；
- 2、 用线性回归的方法可以避免打分方法带来的给某个因子权重过大的弊端；
- 3、 通过回归的方法构建的组合能容纳更多的因子，这样可以提高超额收益的幅度和稳定性。

不过，回归的弊端也较明显，回归的方法相对来说更复杂，处理起来更繁琐，而且不容易让人理解；回归方法所建立的参数还涉及到在样本外的适用性的问题，当然我们本质上只是想要揭示这样的一种规律性，至于具体的参数，各人构建的方法可以差异很大。

## 4.3. 检查的因子

这里因子我们利用了前期单因子系列报告的结论，测试了几个效果较好的因子，包括：

- 估值类：EP、BP、SP、CFP、PEG；
- 增速及盈利类：主营收入当季同比增速、总资产周转率、股东权益周转率；
- 市值类：对数总市值、对数流通市值；
- 盈利预测类：预期 EP、预期增长率、预期 PEG、预期短期主营收入变化
- 技术类：换手率、换手率变化。

## 4.4. 涉及的细节与处理

使用回归的方法其实需要处理的细节较多，主要有以下几个方面：

- 缺省数据处理：由于涉及到盈利预测的数据，而在较早期，盈利预测的数据并不多，所以我们需要面对如何处理早期样本不多的情况。当然我们可以只处理那些有盈利预测的股票，不过在多因子的情况下会导致能够处理的样本较少。我们也可以补全其他没有这个指标的股票一个特定的值（比如指标的中位数），尽量让其不影响回归所体现的规律性。具体步骤是：

- 1、 计算  $t$  期所有奇异值股票的平均收益率  $r$ ；

- 2、根据  $t$  期回归方程计算对应因子值  $x$ ;
- 3、计算  $t$  期因子值  $x$  对应的分位数;
- 4、计算所有期的分位数平均值  $f$ ;
- 5、在每一期将奇异值股票的因子值替换为当期因子值的  $f$  分位数的值。

- 极端数据的处理：对于增速类型的因子，由于存在极端的很大的值，所以需要考虑极端值对整个回归结果的影响。我们通过比较剔除了极端值之后的检验结果和没剔除时候的结果发现差异较大，所以我们对奇异值的处理类似缺省值。
- 部分指标可能并不是简单的线性，比如市值这样的指标，我们对其取对数，而像换手率这样的指标，通过观察几期散点图，我们增加了二次项。

## 5. 单变量牛熊检验

我们按照前面对沪深 300 指数趋势的划分来检查因子在牛熊市下的不同表现。计算 Fama-MacBeth 值的时候分为牛市和熊市的情况，看看是否有明显的差异。为了和直观比较，我们首先检查了一下排序的结果和回归结果的对比。

### 5.1. 排序与回归的对比

排序和回归都可以对因子的选股效果做出评测，理论上两者应该殊途同归，从结果来看，排序中选股效果较好的因子回归的时候 Fama-MacBeth 检验值也较高。我们以总市值为例，下图中为不同市场中的分组收益情况的选股效果展示，具体算法见[单因子报告](#)：

图 3：熊市分组收益（总市值从小到大）

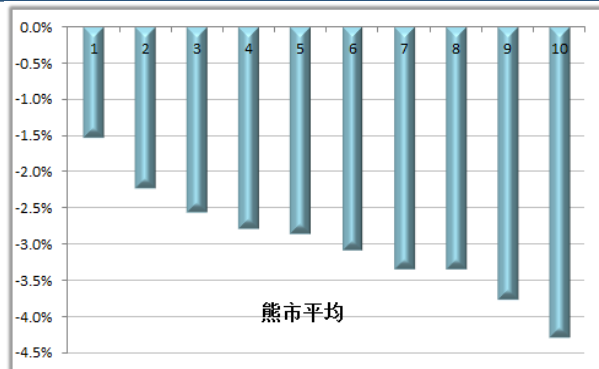
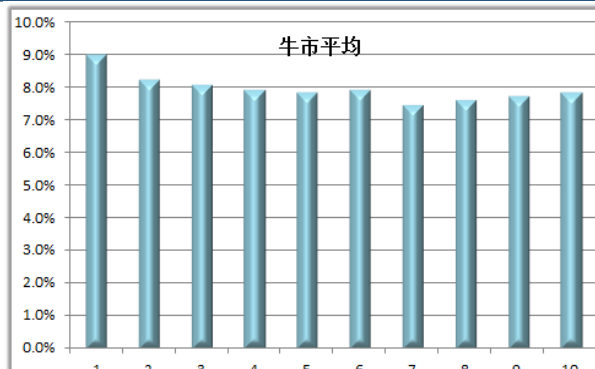


图 4：牛市分组收益（总市值从小到大）



资料来源：天软科技，长江证券研究部

从排序结果看，在熊市中总市值这个因子区分度更好，而在牛市中不太明显。

从回归检验结果来看，和排序结果一致：

因子	牛市检验值	熊市检验值
对数总市值	-0.52752109	-2.224906366



可以看到，在熊市中 Fama-MacBeth 检验值明显高于牛市中的检验值，和排序的结果一致。

## 5.2. 单因子牛熊市检验值比较

我们按照因子在牛熊市下的表现分为三个表，下表是在牛市下区分度较高的因子：

**表 2：牛市区分度较高因子**

因子	牛市检验值	熊市检验值
预期 EP	2.49238781	1.03690521
SP	2.79059285	1.46454075
CFP	2.55761537	1.01072182
预期 G	1.96804437	1.10048783
主营增速	2.29911147	1.83171346
预期 PEG	-5.6014564	-3.1552681
股东权益周转率	2.7194144	1.06279794
历史 PEG	-2.104106252	0.079748744
TTM 主营收入增长率	2.319352643	-0.403225262
NPM(销售净利率)	-2.352339132	-0.202609252
股东权益比	-2.476141781	1.537456722
Beta	3.746567806	-2.934349151

资料来源：天软科技，长江证券研究部

表 2 是在熊市中区分度更好的因子：

**表 3：熊市区分度较高因子**

因子	牛市检验值	熊市检验值
换手率变化	-3.13452	-4.4262099
对数流通市值	-0.5464988	-1.9599884
对数总市值	-0.5275211	-2.2249064
现金比率	-0.9017017	1.60893237
总资产周转率	1.305967511	1.811294604
ROE 当季同比	1.208060746	2.698146544
股息率(>0)	-1.650143598	-3.458254294
速动比率	-1.669582587	3.018827357
流动比率	-1.208203393	1.864524744
预期 PEG	-5.6014564	-3.1552681
超速动比率	-1.631972607	2.74474041
Beta	3.746567806	-2.934349151

资料来源：天软科技，长江证券研究部

表 3 是在牛熊市中都不显著的因子：

**表 4：牛熊市均不显著因子**

因子	牛市检验值	熊市检验值
EP	1.0403376	-0.0944459
BP	1.21739203	1.18929928
预期主营变化率	-0.9196673	-0.1319043
ROA	-0.1229288	0.17524982
ROE	-0.3637055	0.3480461
经营现金流/EPS	-0.174384066	0.192790479
营业外收入增长率	-0.519638747	-1.130415613
GPM(销售毛利率)	-1.416976381	1.529976062
OPM(营业利润率)	-0.84979495	-0.284459136
利息保障倍数	-0.295241026	-0.57221502
10 大流通股占总变动	0.033712023	0.479377642

资料来源：天软科技，长江证券研究部

其中 Beta 这个因子我们使用各只股票在样本区间内对沪深 300 指数做回归获得。

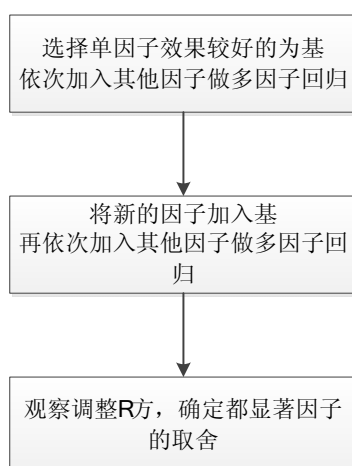
从结果来看有以下几点规律：

- 基于历史数据的估值因子都是在牛市更加显著或者根本不显著；
- 预期因子也是在牛市表现比熊市更好；
- 增速类的因子在牛市表现更好；
- 市值因子在熊市表现比牛市更加显著；
- 熊市中现金比率、偿债能力因子会体现出选股效果，而牛市中则基本没有。

## 6. 多因子组合牛熊检验

### 6.1. 整体流程

这个流程和之前第一篇报告里面所述基本一致，不过我们这里是区分了牛市和熊市而已，也就是我们在做 Fama-MacBeth 检验的时候分为牛市和熊市来考虑。



## 6.2. 牛市组合

	CFP	主营增速	HSLBH	历史 PEG	销售净利率	预期 EP	预期 G	预期 PEG	AdjR-sq:
系数	0.039950857	5.95238E-06	-0.022833548	-0.000109929	-0.000326524	0.199912786	0.000109452	-0.00359881	0.0435
检验值	2.317287672	2.670387713	-3.420363986	-1.818201564	-1.84228482	3.069030619	1.783079979	-3.221924015	

考虑到个股本身的  $\beta$  的影响，我们加入了  $\beta$  这个因子一起进行回归，结果如下所示：

	CFP	主营增速	HSLBH	历史 PEG	销售净利率	预期 EP	预期 G	预期 PEG	beta 因子	AdjR-sq
系数	0.032936	0.000006	-0.021603	-0.000080	-0.000358	0.183953	0.000081	-0.004024	0.056016	0.063982
检验值	1.974050	2.751542	-3.263004	-1.373926	-1.954768	2.894481	1.319270	-3.685191	2.903251	

可以看到，加入  $\beta$  因子之后历史 PEG 和预期 G 都不再显著，所以我们将这两个指标剔除。最后的表如下：

	CFP	主营增速	HSLBH	销售净利率	预期 EP	预期 PEG	beta 因子	AdjR-sq:
系数	0.032573	0.000006	-0.021694	-0.000360	0.174962	-0.004688	0.056662	0.062921
检验值	1.943949	2.772203	-3.261681	-1.974617	2.877938	-4.900497	2.936449	

## 6.3. 熊市组合

	主营增速	HSLBH	对数总市值	roe 当季同比	速动比率	AdjR-sq:
系数	0.000002	-0.021650	-0.005491	0.000012	0.000822	0.049559
检验值	1.847336	-5.183284	-2.143528	3.190436	3.012449	

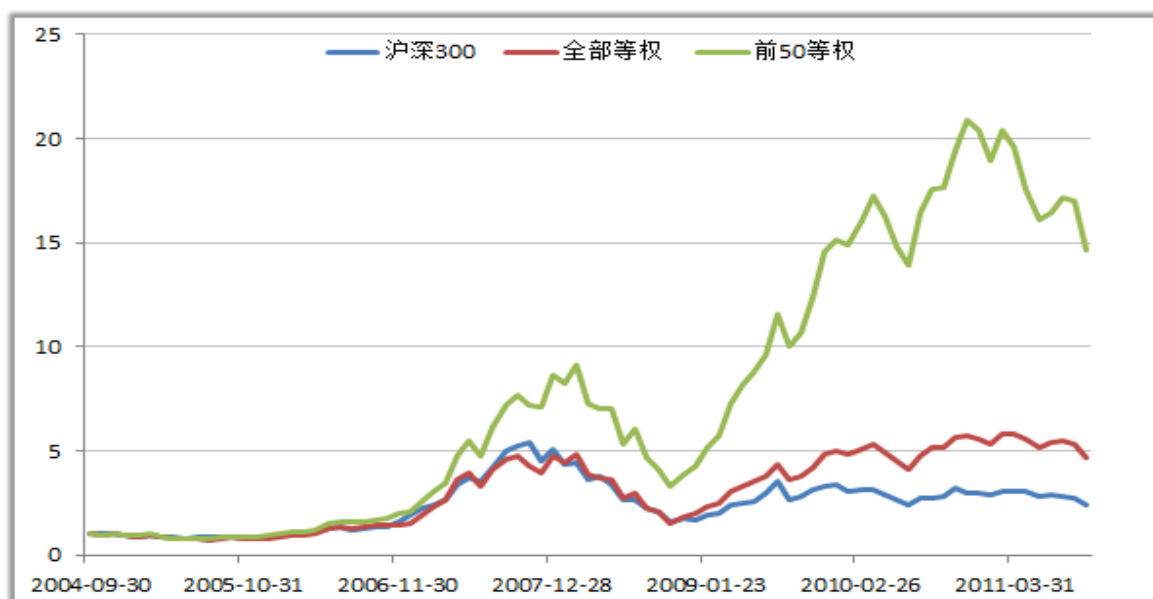
加入  $\beta$  之后的组合入下表所示：

	主营增速	HSLBH	对数总市值	roe 当季同比	速动比率	beta	AdjR-sq:
系数	0.000002	-0.021205	-0.005230	0.000013	0.000621	-0.031196	0.067682
检验值	1.659324	-5.020375	-2.062305	3.585018	2.411028	-2.168446	

## 6.4. 组合的组合

我们将牛市组合和熊市组合放在一起，在牛市的时候选择牛市组合，在熊市的时候选择熊市组合，当然实际上很难做到，我们只是放在这里展示一下。

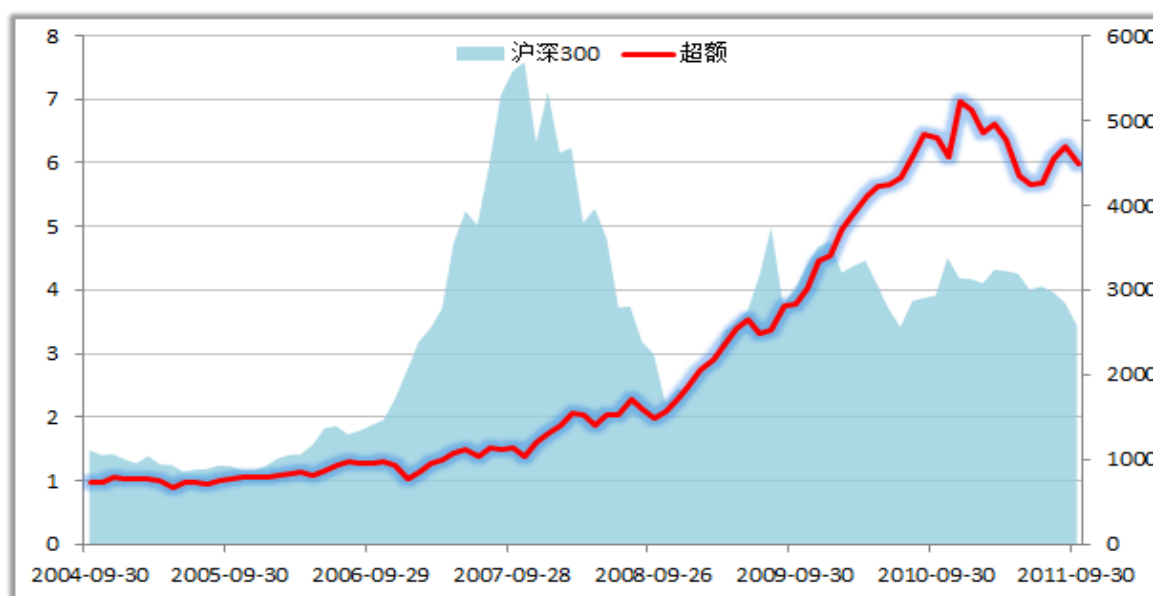
图 5：组合的组合走势



资料来源：长江证券研究部。

从超额收益来看，最大的一段回撤发生在 2010 年到 2011 年之间的几个月份，之前的第一篇报告中多数组合也是在这个区间段发生了较大回撤，看来这段时间并不是靠量化组合能够胜出的舞台。

图 6：组合的组合超额收益走势



资料来源：长江证券研究部。

## 7. 后续研究方向

- 测试不同因子的最优持仓时间：我们现在持仓时间都是 1 个月，但是实际上可能有些因子的效果需要持有更长时间才能体现出来，所以我们需要测试各种因子的最优持仓时间。而持有时间拉长又会涉及到自相关的问题，所以还需要用一些办法来消除这种关系。
- 控制行业层面的属性，这个可以通过行业中性的方法实现，尽量让策略的组合在每个行业没有太大的倾斜。也可以加入一些行业因子属性的方式来做。比如测试 EP 的时候，测试相对于行业平均 EP 水平的 EP，这样能避免某些行业本身就是 EP 较大或者较小的情况。

### 分析师介绍

范辛亭，中国科学技术大学博士，香港中文大学博士后，中山大学副教授，长江证券金融工程首席分析师。

袁继飞，同济大学计算机科学与技术专业本科，软件与理论研究生，长江证券金融工程分析师，主要研究领域为高频数据挖掘和股指期货。

### 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
周志德	主管	(8621) 68751807	13681960999 Zhouzd1@cjsc.com.cn
甘 露	华东区总经理	(8621) 68751916	13701696936 ganlu@cjsc.com.cn
鞠 雷	华南区总经理	(8621) 68751863	13817196202 julei@cjsc.com.cn
程 杨	华北区总经理	(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

### 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。