

2012-05-28

金融工程(专题报告)

量化选股

考虑风险因子的整合选股模型

--量化选股的线性回归体系构建(四)

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

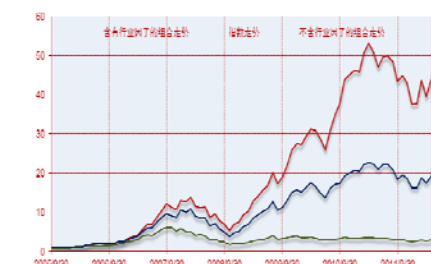
执业证书编号:

联系人: 杨靖凤

☎ (8621)68751636

✉ yangjf@cjsc.com.cn

含有行业因子的组合与普通因子组合走势



资料来源: 长江证券研究部

相关研究

量化选股的线性回归体系构建(一)

量化选股的线性回归体系构建(二)

量化选股的线性回归体系构建(三)

报告要点

■ 风险因子与普通选股因子的区别

普通选股因子是全局有效的,在时间序列上的表现是有效和无效混杂在一起,也就是会在某一个时间有效,然后某个时间又会失效,没有明显的趋势规律。风险选股因子可能是全局无效的,这些因子只会在特定的市场行情下才会有效。风险因子一般都会持续一段时间,形成一波趋势性的投资机会。

■ 风险因子的重要性

- 风险因子包含类别可以简单概括为:风格类风险因子(包括大小盘,估值成长,高低股价),行业收益,主题类以及事件类。相对来说,主题类的较难以量化,风格和行业相对来说更容易量化。
- 以个股对应申万一级行业当月收益率来解释个股当月收益率会非常显著,说明行业的收益率对个股的表现影响非常大。
- 我们之前做的行业多因子择时模型能够对行业实际收益率截面上排序有一定预测作用,使用 Fama-MacBeth 检验也十分显著。
- 我们将行业预期收益率作为一个单独的选股因子做 Fama-MacBeth 检验,结果非常显著,并且检验值比之前所有普通选股因子的绝对值都更大,说明行业预测的准确性对个股的选择影响非常大。

■ 整合选股模型的构建

我们将行业预期收益率这个因子加入之前的按月调仓的选股模型里面,在加入了预期行业收益之后,预期行业收益这个因子的检验值并没有明显的下降,说明这个因子和普通的选股因子的作用是不一样的。而且从解释度来说也增加1%左右,说明这个因子的选股效果较为明显。

考虑行业因子的整合选股模型不是把行业收益不加任何加工就带入选股模型,而是在行业因子前面有回归系数,这个经验系数有两个作用:一是考虑了行业收益预测的不确定性,二是考虑了与其他普通因子的协同性。

从组合表现来看,夏普比率从 38.20% 上升到 45.87%,信息比率从 48.92% 上升到 72.90%,年化收益从 55.32% 上升到 75.44%,改进非常明显。

目录

一、风险因子对选股的启示	3
1、风险因子与普通选股因子的区别	4
2、目前认识到的风险因子	4
二、以行业为例看风险因子的重要性	5
1、当期行业收益率做选股因子	5
2、行业多因子配置模型	5
三、行业因子和普通选股因子的整合选股模型	12
四、总结	13

图表目录

图 1：有色金属行业的备选因子	5
图 2：农林牧渔行业和 PPI 的关系	9
图 3：农林牧渔行业和固定资产投资的关系	9
图 4：农林牧渔行业和贷款余额同比增速的关系	10
图 5：农林牧渔行业和 CPI-食品-蛋-农村的关系	10
图 6：农林牧渔行业因子择时模型效果	11
图 7：两个截面上行业因子模型的预测作用	11
图 8：含有行业因子的组合与普通因子组合走势	13
表 1：单纯 beta 值的 T 检验值	3
表 2：beta 乘以市场收益率的 T 检验值	3
表 3：常见风险因子	4
表 4：当期行业收益率的检验值	5
表 5：上游行业领先多因素模型	6
表 6：中游行业领先多因素模型	6
表 7：下游行业领先多因素模型	7
表 8：农林牧渔行业的多因素模型（样本区间 2005.9-2012.4）	8
表 9：行业预期收益率的预测作用	12
表 10：行业预期收益率的选股作用	12
表 11：包含行业因子的多因子选股模型（样本区间 2005.9-2012.4）	12
表 12：两个组合统计指标比较	13

一、风险因子对选股的启示

CAPM 模型把市场收益定义为一个风险因子，如下所示：

$$r_i = r_f + \beta_i * (r_m - r_f)$$

但是 r_m 是一个全局变量，对不同股票来说数值都是一样的，对不同股票的影响体

现在 β_i 上面。我们对 β_i 进行 Fama-MacBeth 检验，步骤和对普通因子进行检验一样，也是分为两步来进行：

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风险因子进行回归：

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

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

在得到每期的回归系数后，我们可以做 Fama-MacBeth 检验：

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

表 1：单纯 beta 值的 T 检验值

选股因子	Fama-MacBeth T 值
β_t	0.97

资料来源：长江证券研究部

可以看到 β_i 本身检验值较小，并没有选股效果，但是如果拿 β_i 乘以当期的市场收益率则结果会有所不同，如下表所示：

表 2：beta 乘以市场收益率的 T 检验值

选股因子	Fama-MacBeth T 值
$\beta_i r_{mt}$	4.87

资料来源：长江证券研究部

将 β_i 乘以市场当期收益率之后，检验值就超过 1.96 的临界值了，并且系数是正的。如果用这个指标来选股，预测大盘上涨的时候就倾向于选择 β 较大的股票；而如果预测大盘下跌的话则倾向于选择 β 较小的股票。

但是如何预测市场收益率 I_m 本身就是个很难的问题，另外一方面，我们之前做过很多普通选股因子的研究，如何将类似于像市场收益率这样的风险因子和普通选股因子结合起来也是一个问题。

1、风险因子与普通选股因子的区别

普通选股因子：

- 是全局有效地，这个全局主要体现在各种行情之下，不管是上涨还是下跌或者还是震荡，不管是大盘股表现较好或者小盘股表现更优；
- 在时间序列上的表现是有效和无效混杂在一起，也就是会在某一个时间有效，然后某个时间又会失效，没有明显的趋势规律。

风险选股因子：

- 风险选股因子可能是全局无效的，这些因子只会在特定的市场行情下才会有效，比如高 β 的股票只有在市场上涨的之后才会有选股效果；
- 风险因子一般都会持续一段时间，形成一波趋势性的投资机会。

当然这两种选股因子可能会有交集，用普通的排序分组法或者打分的办法较难检验风险因子的选股效果，而我们之前介绍的 Fama-MacBeth 则能够解决这样的问题。

2、目前认识到的风险因子

表 3：常见风险因子

风格	大小盘 (SMB)、价值成长 (HML)、动量反转 (UMD)、高低价、周期防御等
行业	申万23个一级行业、中信29个一级行业等
主题	光伏、世博、高铁、传媒、海洋工程、环保等
事件	高分红送转、年报披露信息、异常交易公告、指数成分股调入调出等

资料来源：长江证券研究部

风格和行业的风险因子相对来说容易量化一些，而主题的量化相对来说更难。

二、以行业为例看风险因子的重要性

1. 当期行业收益率做选股因子

我们首先来看看行业收益率对个股收益的影响程度，我们使用个股申万一级行业当月的收益率来解释个股当月的收益率，做 Fama-MacBeth 检验。

表 4：当期行业收益率的检验值

选股因子	Fama-MacBeth T 值
行业	17.55

资料来源：长江证券研究部

从检验结果来看非常显著，这说明行业收益率对个股当期的收益率影响非常大。不过由于这里是使用的当期的行业收益率，所以无法使用。如果我们能有一个行业配置模型，然后能对行业的收益率有一个预测的话，我们估计检验值也许能够比较显著，当然肯定会比使用当月行业收益率检验值会低很多。

2. 行业多因子配置模型

我们前期发过三篇关于行业配置的报告：《上游行业的领先多因素模型》，《中游行业的领先多因素模型》，《下游行业的领先多因素模型》，在这三篇报告中我们详细描述了使用行业基本面因子来建立行业择时模型的过程。

对每一个行业，我们都会选择大量的基本面因子作为备选因子库，以有色金属行业为例：

图 1：有色金属行业的备选因子



资料来源：长江证券研究部。

在有色金属行业的模型中，我们选择约 30 多个宏观经济因素和行业本身及其上下游的基本面因子，一共 70 多个指标进行筛选。

对每一个行业我们都进行这样的筛选，然后使用逐步回归的办法建立多因子预测模型，最后的结果如下：

表 5：上游行业领先多因素模型

	行业	领先多因素模型				adj-R ²	正负胜率
上游行业	有色金属	贷款余额同比	库存-LME 铅	PPI-重工业-加工		0.42	0.7
	采掘	积压订单	贷款余额同比	PPI-重工业-加工	库存:钢坯:同比	0.36	0.71

资料来源：长江证券研究部。

表 6：中游行业领先多因素模型

	行业	领先多因素模型				adj-R ²	正负胜率
中游行业	交通运输	m1-m2	贷款余额同比	出口额与进口额增速差		0.32	0.7
	轻工制造	用电量	供应商配送时间	森林		0.27	0.71
	公用事业	PMI	进口	焦炭产量同比	PPI-重工业-采掘 库存:钢坯:同比	0.37	0.68
	电子	PPI	固定资产投资	用电量		0.3	0.78
	化工	用电量	贷款余额同比	库存-LME 铅	纺织业-原材料库存	0.39	0.74
	机械设备	PPI-重工业	库存:钢坯: 同比			0.27	0.75
	信息设备	PPI	固定资产投资	用电量		0.27	0.75
	黑色金属	供应商配送时间	贷款余额同比	库存:钢坯: 同比		0.37	0.7

	建筑建材	汽车产量:乘用车:基本型(轿车):当月同比	PPI-重工业-采掘	房地产自筹资金累计同比	铁矿石原矿量:当月同比	库存:钢坯:同比	0.39	0.75
--	------	-----------------------	------------	-------------	-------------	----------	------	------

资料来源: 长江证券研究部。

表 7: 下游行业领先多因素模型

	行业	领先多因素模型				adj-Rsquare	符号胜率
下游行业	食品饮料	PPI	美元指数	PPI-生活资料-衣着		0.44	0.79
	纺织服装	固定资产投资	贷款余额同比	CPI-家庭设备用品及服务-城市	PPI-纺织业	0.42	0.74
	交运设备	用电量	PPI-重工业-加工	库存:钢坯:同比		0.31	0.71
	家用电器	固定资产投资	贷款余额同比	PPI-生产资料-加工	CPI-家庭设备用品及服务-城市	0.38	0.75
	商业贸易	PPI	消费者预期指数:月	PPI-生活资料-衣着		0.32	0.76
	农林牧渔	PPI	固定资产投资	贷款余额同比	CPI-食品-蛋-农村	0.3	0.74
	餐饮旅游	PPI	用电量	供应商配送时间		0.3	0.71
	综合	原材料库存	进口	贷款余额同比	出口额与进口额增速差	0.31	0.65
	医药生物	PPI	固定资产投资	用电量	出口额与进口额增速差	0.3	0.77
	房地产	用电量	上海:新建住宅价格指数:当月环比	PPI-重工业		0.27	0.74
	信息服务	贷款余额同比	PPI-生活资料-衣着	CPI-家庭设备用品及服务-城		30	0.6

市						
其他	金融服 务	PPI	供应商配 送时间	贷款余额 同比	0.32	0.68

资料来源：长江证券研究部。

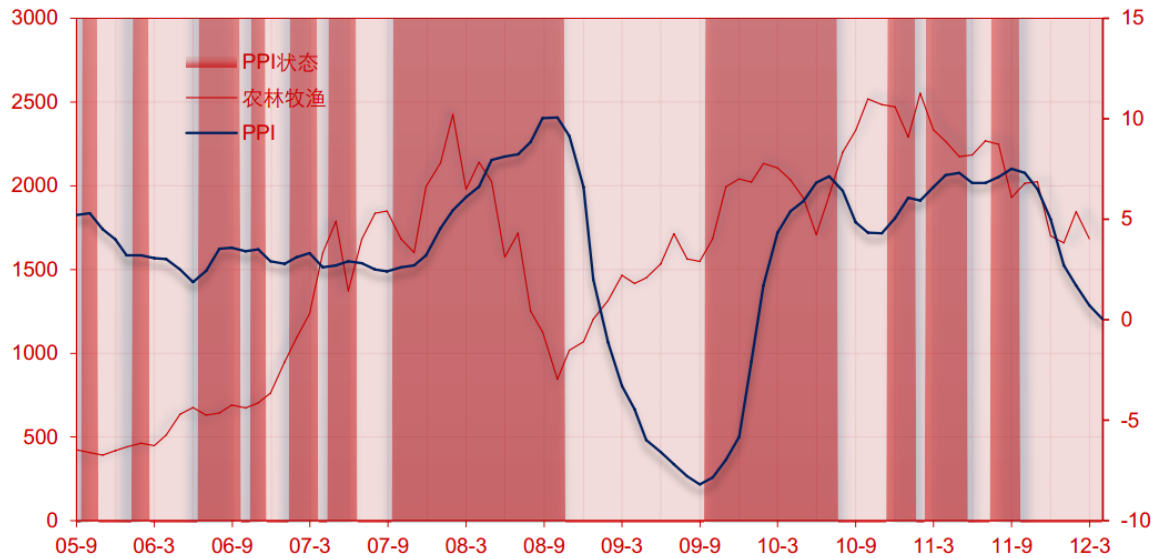
以上三个表是每个行业最后筛选出来的变量，对于每个行业，这些对应的变量都会形成一个多因子模型，所以可以计算每个行业下个月的预期收益率。我们以农林牧渔行业为例，最后的模型为：

表 8：农林牧渔行业的多因素模型（样本区间 2005.9-2012.4）

	PPI 固定资产投资 贷款余额同比增速 CPI-食品-蛋-农村				
j	1 4 4 2				
k	-2 -2 -1 -2				
beta	0.29	-6.13	-5.60	7.68	5.03
t_test	0.08	-2.42	-2.28	2.65	2.04
Rsquare	0.30				
正负符号胜率	0.75				

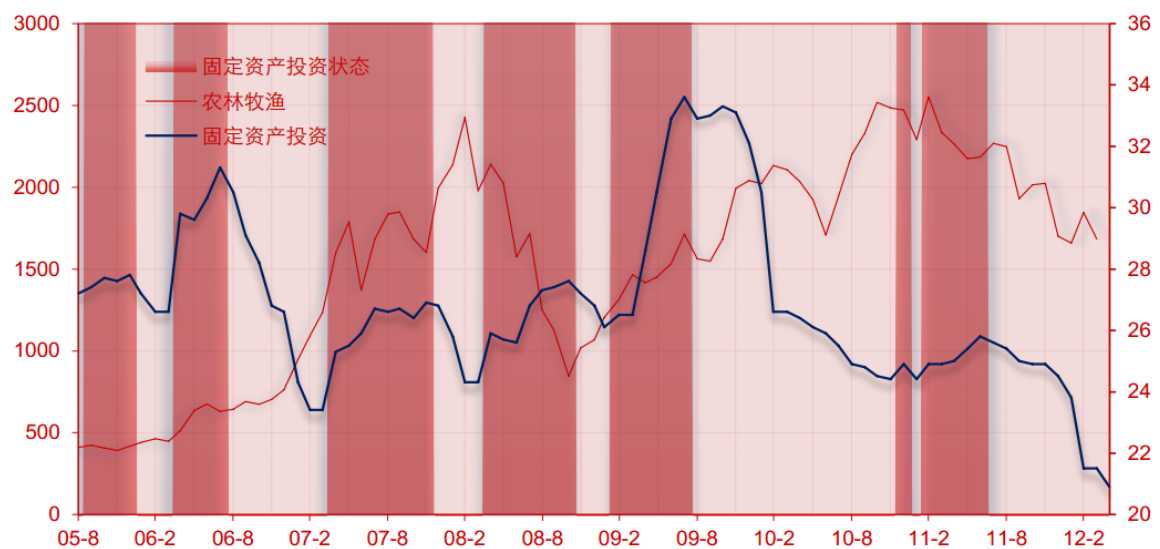
资料来源：长江证券研究部。

单纯从这个公式难以看出各个变量和行业指数的关系，我们可以通过画图来直观地看看单个变量与行业指数的关系。

图 2：农林牧渔行业和 PPI 的关系


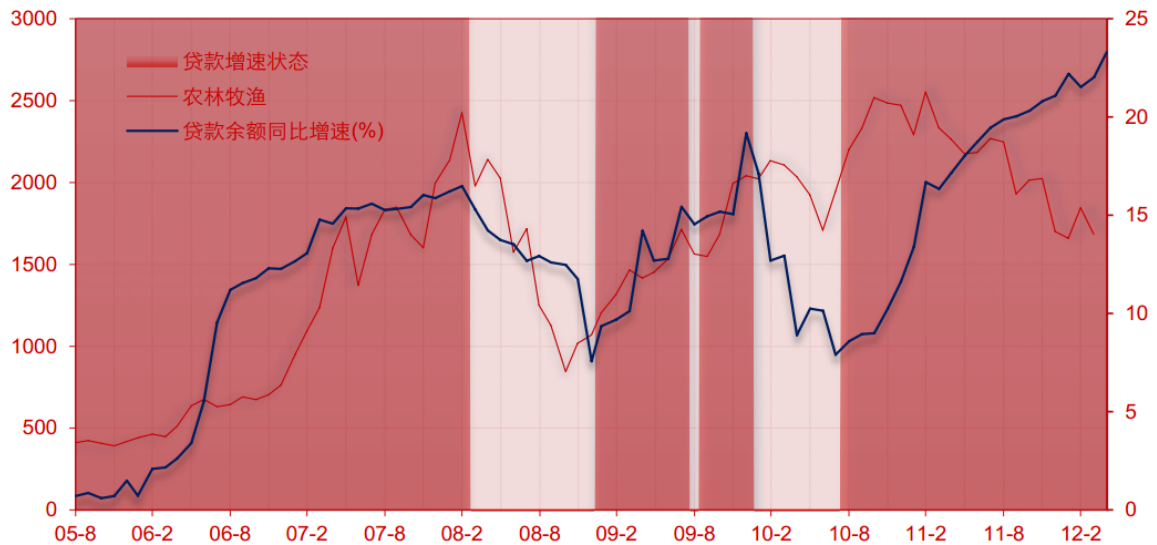
资料来源：长江证券研究部。

农林牧渔和 PPI 是反向的关系，这可以从回归方程看出来，因为其 β 值是负的。从上图也能看出来，在 PPI 处于上升状态的红色区域，农林牧渔指数却是在下降。

图 3：农林牧渔行业和固定资产投资的关系


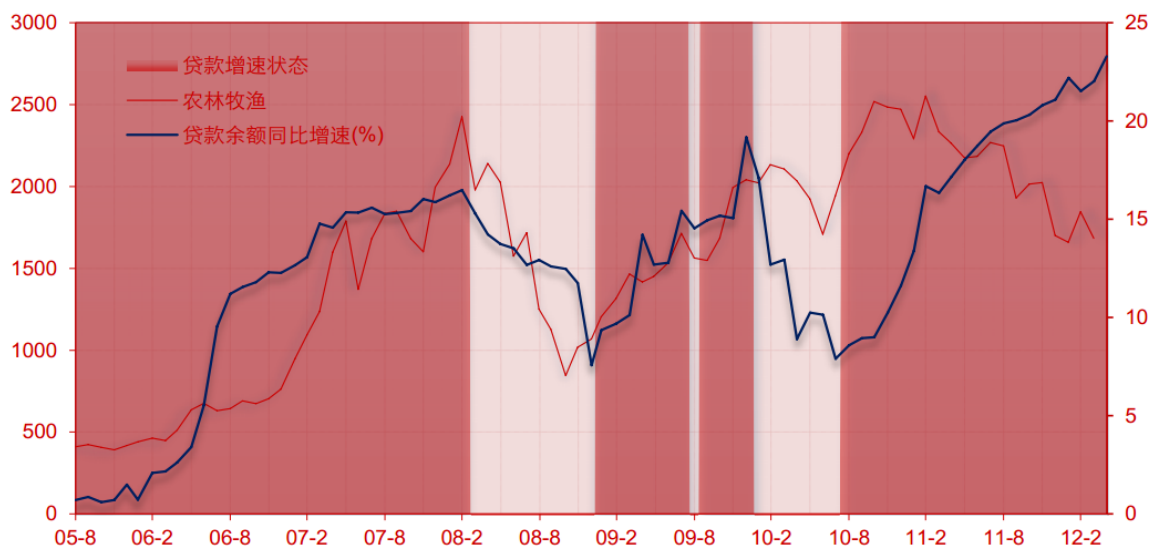
资料来源：长江证券研究部。

农林牧渔和固定资产投资仍然是反向的关系，而且从前面的表里面可以看到，很多非周期的行业都是和固定资产投资是反向的关系，这也比较符合逻辑，当固定资产投资增长较快的时候，经济运行健康，一般说来是周期性行业表现更好。从上表可以看到，纺织服装和家用电器行业都体现了这样的规律。

图 4：农林牧渔行业和贷款余额同比增速的关系


资料来源：长江证券研究部。

贷款余额同比增速作为流动性的度量，出现在了多个行业里面。农林牧渔行业也不例外，并且关系是正向的关系。

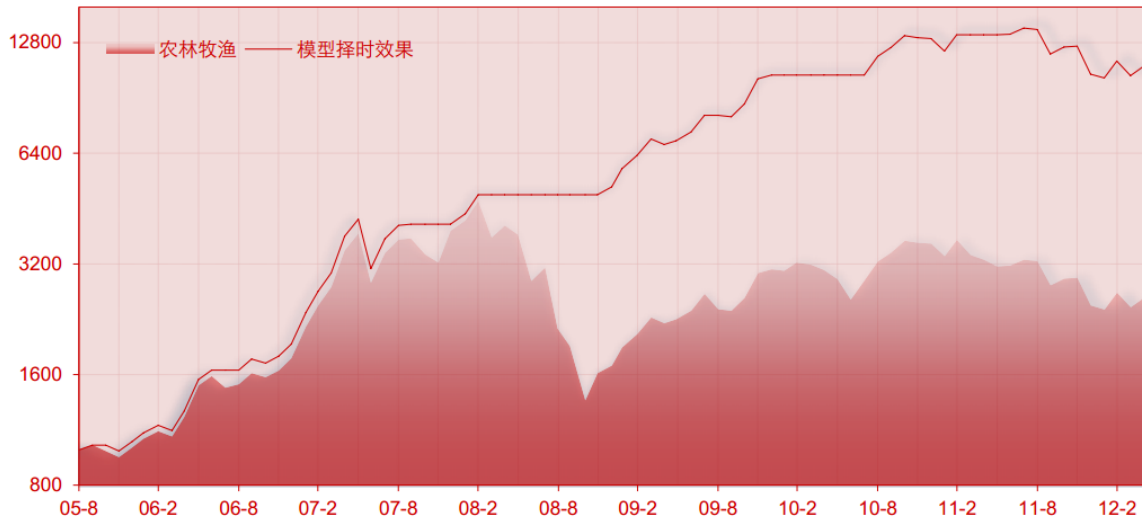
图 5：农林牧渔行业和 CPI-食品-蛋-农村的关系


资料来源：长江证券研究部。

CPI-食品-蛋-农村是CPI的一个分项，它衡量了鸡蛋的价格。可以看到这个指标和农林牧渔行业是显著地正相关，从逻辑上来说，鸡蛋价格越高，养殖类的企业盈利情况更好，因为鸡蛋是一种刚性需求，并不会因为价格的升高而减少消费量。

图 6：农林牧渔行业因子择时模型效果

农林牧渔行业因子模型择时效果



资料来源：长江证券研究部。

上图是根据每个月因子择时模型计算出来的预期收益率进行择时的结果，如果预期收益率为正，那么就做多；如果为负，那么就空仓。

对每个行业在每个月做这样的预测，我们就可以在每个月对所有行业有一个截面的预期收益率，如果这个预期收益率还比较靠谱的话，那么就有可能对我们的选股有所改进。

图 7：两个截面上行业因子模型的预测作用

2010年1月			2011年3月		
行业	预期收益率	实际收益率	行业	预期收益率	实际收益率
餐饮旅游	0.06	0.03	综合	0.08	0.00
信息设备	0.04	0.01	黑色金属	0.06	0.06
房地产	0.04	-0.13	化工	0.04	-0.02
医药生物	0.03	0.01	金融服务	0.04	0.02
电子元器件	-0.01	-0.01	有色金属	0.03	0.03
建筑建材	-0.03	-0.07	采掘	0.02	0.04
信息服务	-0.03	0.02	公用事业	0.01	-0.01
机械设备	-0.03	-0.06	食品饮料	0.01	-0.01
商业贸易	-0.04	-0.04	家用电器	0.00	-0.01
综合	-0.04	-0.01	农林牧渔	0.00	-0.09
家用电器	-0.04	-0.09	信息服务	-0.01	-0.05
轻工制造	-0.04	-0.03	交通运输	-0.01	-0.03
交通运输	-0.05	-0.01	餐饮旅游	-0.02	-0.04
交运设备	-0.05	-0.13	房地产	-0.03	0.05
农林牧渔	-0.05	-0.01	建筑建材	-0.03	-0.01
公用事业	-0.05	-0.06	机械设备	-0.03	-0.05
金融服务	-0.06	-0.12	商业贸易	-0.04	-0.03
食品饮料	-0.06	-0.03	轻工制造	-0.04	-0.02
纺织服装	-0.07	-0.02	纺织服装	-0.04	0.02
黑色金属	-0.08	-0.17	交运设备	-0.05	-0.05
化工	-0.10	-0.14	医药生物	-0.06	-0.04
有色金属	-0.10	-0.15	电子元器件	-0.08	-0.05
采掘	-0.11	-0.18	信息设备	-0.08	-0.07

资料来源：长江证券研究部。

这是我们选择两个截面上预测效果的结果展示，如果我们的行业因子择时模型在截面上能够对实际的收益率排行作出一定的预测，那么就可以将其纳入我们的选股体系中进行测试。

我们在每期用行业的预测收益率解释实际的收益率，还是使用 Fama-MacBeth 模型进行检验，结果如下：

表 9：行业预期收益率的预测作用	
样本区间	2005 年 9 月-2012 年 4 月
Fama—MacBeth T 值	8.90
资料来源：长江证券研究部	

可以看到检验值是相当显著的，这说明行业的预期收益率对行业实际收益率解释性较强，因此可以将行业的预期收益率作为风险因子纳入选股模型。

我们将行业预期收益率作为一个单独的选股因子进行检验，结果如下：

表 10：行业预期收益率的选股作用	
样本区间	2005 年 9 月-2012 年 4 月
Fama—MacBeth T 值	7.74
资料来源：长江证券研究部	

虽然相对于当月行业收益率的检验值低了很多，但是仍然非常显著。

三、行业因子和普通选股因子的整合选股模型

我们将行业预期收益率纳入之前的每月调仓的模型里面，结果如下：

表 11：包含行业因子的多因子选股模型（样本区间 2005.9-2012.4）										
	预期 EP	SP	预期 G	主营增速	预期 PEG	20 日换手率	60 日换手率	20 日换手率平方	60 日换手率平方	
beta	0.1040	0.0015	0.0001	0.0000	-0.0024	-0.0101	0.0111	0.0003	-0.0007	
t_test	2.6327	1.7006	1.8435	3.1443	-3.2102	-3.3118	3.7403	1.4484	-2.9142	
R ²	0.0490									
	预期 EP	SP	预期 G	主营增速	预期 PEG	20 日换手率	60 日换手率	20 日换手率平方	60 日换手率平方	预期行业收益
beta	0.1028	0.0014	0.0001	0.0000	-0.0023	-0.0104	0.0112	0.0003	-0.0007	0.1875
t_test	2.7441	1.5745	1.7717	3.1791	-3.0404	-3.4356	3.8222	1.4914	-2.9320	7.5267
R ²	0.05785									
资料来源：长江证券研究部										

可以看到，在加入了预期行业收益之后，预期行业收益这个因子的检验值并没有明显的下降，说明这个因子和普通的选股因子的作用是不一样的。而且从解释度来说也增加 1% 左右，说明这个因子的选股效果较为明显。

我们按照上面两个模型将个股的下个月预期收益率计算出来，然后取前 50 只股票做成组合，其表现如下：

图 8：含有行业因子的组合与普通因子组合走势



资料来源：长江证券研究部。

从结果可以看到，加入了行业因子的组合表现显著优于没有行业因子的组合，我们对其统计一些指标，结果如下：

表 12：两个组合统计指标比较

	含有行业预期	不含行业预期
胜率	80.25%	74.07%
超额收益	3899.75%	1408.13%
年化收益	75.44%	55.32%
平均月收益率	5.48%	4.38%
夏普比率	45.87%	38.20%
月平均超额收益	2.70%	1.60%
信息比率	72.90%	48.92%

资料来源：长江证券研究部。

四、总结

行业因子的模型不是把行业收益不加任何加工就带入选股模型，而是在行业因子前面有回归系数，这个经验系数有两个作用：

- 考虑了行业收益预测的不确定性；

- 考虑了与其他普通因子的协同性。

本文主要以行业预期收益率作为风险因子为例，讲述了如何将风险因子和普通选股因子结合起来实现整合选股模型思路。后续我们可能会推出基于其他风险因子的整合选股模型，比如将大小盘轮动作为一个风险因子纳入到选股模型里面。

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。		



研究部/机构客户部

上海

浦东新区世纪大道 1589 号长泰国际金融大厦 21 楼
(200122)

电话: 021-68751100

传真: 021-68751151

武汉

武汉市新华路特 8 号长江证券大厦 9 楼
(430015)

传真: 027-65799501

北京

西城区金融大街 17 号中国人寿中心 606 室
(100032)

传真: 021-68751791

深圳

深圳市福田区福华一路 6 号免税商务大厦 18 楼
(518000)

传真: 0755-82750808

0755-82724740

重要声明

长江证券股份有限公司具有证券投资咨询业务资格，经营证券业务许可证编号：Z24935000。

本报告的作者是基于独立、客观、公正和审慎的原则制作本研究报告。本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，也不保证所包含信息和建议不发生任何变更。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，不包含作者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。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司及作者在自身所知情形范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

本报告版权仅仅为本公司所有，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。如引用须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。刊载或者转发本证券研究报告或者摘要的，应当注明本报告的发布人和发布日期，提示使用证券研究报告的风险。未经授权刊载或者转发本报告的，本公司将保留向其追究法律责任的权利。