

# 基于历史状态空间相似性匹配的行业配置 SMIA 模型

## ——相似性匹配量化模型研究之二

安宁宁 资深分析师  
电话: 0755-23948352  
eMail: ann@gf.com.cn  
执业编号: S0260512020003

### 模型基本思想

随机过程可以划分为有记忆性特征和无记忆性特征两种。行业轮动具有记忆性,即当期行业的相对表现会影响下一期行业的相对表现,这是由行业之间错综复杂的经济关系所决定的。我们通过寻找历史状态空间中与当期行业收益率排名相似的一些时间点,观察这些时间点之后一期行业轮动的变化特征,从中选取统计上表现较好的行业,作为当前时间点下一期的推荐超配行业,从而实现相似性匹配行业配置(Similarity Matching Industry Allocation,缩写为 SMIA)量化模型的构建。

### SMIA 量化模型的构建

行业市场排名存在一定的记忆性——每一期行业收益率的排名在一定程度上依赖于上一期的排名,这也正是行业轮动的外在表现。理论上来说,我们可以从历史上找到一些和当期行业排名相同的样本,观察他们下一期行业轮动所产生的变化,然后从统计上找到下一期表现持续较好的一些行业进行超配,并回避一些表现持续较差的行业。但是在实际操作中遇到的问题是,由于 A 股市场成立时间相对较短,对于很多的行业划分来看,不可能在历史上找到足够多的和当期行业收益率具有一致排名的样本。因此,我们只能在有限的历史状态空间中寻找和当前行业排名状态最为相似的几个状态。我们通过序列之间的距离来寻找最相似的历史相对排名,即通过求解当前序列与历史状态空间中序列两两之间的距离,并且经过简单的排序,就可以找到历史数据库中最为匹配的相似序列。只寻找一期最相似的行业市场排序容易造成预测结果波动较大、预测风险较高,因此,我们需要在历史上寻找几期最为相似的排名状态,观察他们之后一期的行业轮动与变化。通过一定的分析,我们采用观察历史相似时点之后一期排名非常靠前的行业,并对他们取并集的方法进行计算。

### SMIA 量化模型在 A 股市场中的实证

按照上述 SMIA 模型构建思路,我们首先选取申万一级行业指数中除“综合”以外的 22 个行业作为预配置标的,每月进行一次行业配置预测与推荐。历史状态空间样本从 2000 年的开始取起,样本数据库窗口的截止时间为预测期的前一期。从实际投资应用的角度,我们采用沪深 300 指数作为比较基准,从 2006 年开始对 SMIA 量化行业配置模型进行回测。从回测结果来看,SMIA 量化资产配置模型可以在统计上有效筛选出相对表现较好的一些行业。在 2005 年 12 月 30 日至 2012 年 7 月 31 日的 80 次预测中,有 54 次预测正确(我们将推荐行业组合平均收益率超过沪深 300 指数基准收益率定义为正确),26 次失败,预测正确率达到 67.5%;并且在绝对值上,盈利平均超额收益率略高于亏损平均超额收益率。从统计上来看,牛市中的预测效果要好于熊市中的预测效果。不过 2012 年以来的 7 次预测中,仍然有 5 次正确。

## 目录索引

一、模型基本思想.....	3
二、行业分类标准.....	4
三、SMIA 量化行业配置模型的构建.....	4
四、SMIA 量化行业配置模型在 A 股市场中的实证.....	7
五、总结.....	12

## 图表索引

图 1: 无记忆性随机过程实例.....	3
图 2: 有记忆性随机过程实例.....	3
图 3: 排名状态相似性匹配实例.....	5
图 4: 进行相似性匹配统计的两种方法.....	7
图 5: SMIA 组合与沪深 300 指数走势对比 (2006 年起) .....	10
图 6: SMIA 模型各年度预测效果 (2006 年起) .....	11
图 7: SMIA 模型各行业推荐次数 (2006 年起) .....	11
表 1: SMIA 回测结果详细列表.....	8
表 2: SMIA 模型回测效果统计 (2006 年起) .....	10
表 3: SMIA 模型各年度预测效果 (2006 年起) .....	10

## 一、模型基本思想

离散动力学系统随时间变化，可能出现的所有状态构成一个状态空间。

例如，我们变换一个正方体魔方，每变换一次，其六个面状态的集合都是相应状态空间的一个元素。又如我们有三枚硬币，每次抛硬币所组成的正反组合状态也是相应状态空间的一个元素。

对于这样的随机过程，我们总体可以分为两类：有记忆性随机过程和无记忆性随机过程。

假设我们有 6 个骰子，进行多次随机抛置。其中第  $i$  ( $i \geq 2$ ) 次抛置的组合状态与第  $i-1$  次的状态是没有任何关系的。这样的随机过程称为无记忆性随机过程。

图 1：无记忆性随机过程实例



数据来源：广发证券发展研究中心

另外一种随机过程是有记忆性随机过程。举一个形象一些的例子：一家服装店有四类服装出售——短袖、衬衫、毛衣、羽绒服。某年的四个季度销量排名情况如图 1 所示：

图 2：有记忆性随机过程实例

	春	夏	秋	冬
短袖	2	1	4	4
衬衫	1	2	3	3
毛衣	3	3	1	2
羽绒服	4	4	2	1

数据来源：广发证券发展研究中心

该排名的状态变化也是一种随机过程，但这种随机过程与抛骰子不同，其具有一定的记忆特性。例如，往往在春、夏两季短袖和衬衫销量比较好，但到了秋、季之后，这两个品种的销量开始显著下滑，而毛衣和羽绒衣的销量开始大幅上升。也就是说，每当出现连续两个季度短袖、衬衫（毛衣、羽绒服）销量很好的状态，第三、四个连续的状态中往往都是毛衣、羽绒服（短袖、衬衫）销量较好。这种状态之间具有一定潜在关系的随机过程就是一种有记忆性的随机过程，或者说当前状态与之前发生的状态有关。

在股票二级市场中，行业轮动是机构投资者广泛关注的一个问题。不同行业的股票在经济周期的浪潮中总是展现出不同的相对表现，这种行业轮动具有一定的经济学意义，不同行业之间的供求关系、盈亏周期在一定的经济关系下相互影响、相互联系。这种联系反映到二级市场上，直观来看就是行业整体收益率的相对表现情况。

由于这种行业市场表现的排名背后牵扯着各种各样的经济学逻辑和行业轮动关系，因此行业相对排名随时间的变化应该是一种有记忆性的随机过程。与传统行业配置从行业基本面出发的思路不同，本篇报告则是从行业排名随机过程具有记忆性的角度出发，完全通过市场数据对行业表现做出预测，从而在股票二级市场上实现行业的有效配置。

行业市场排名具有记忆性表现在：当在某个周期内，行业收益率的排名作为一种状态  $S_t$  出现之后，下一个周期的状态  $S_{t+1}$  在一定程度上是与  $S_t$  有关的，或者说在  $t+1$  时刻状态空间中各种状态出现的概率是不相等的。广发相似性匹配行业配置（Similarity Matching Industry Allocation，缩写为 SMIA）量化模型，就是要在历史上找到一些与当前状态最为相似的行业表现排名，观察它们下一期的行业轮动变化，从中找出一些统计性的特征规律，从而对未来一期的行业配置提出预测性建议。

## 二、行业分类标准

对于 A 股市场，行业分类有很多不同的标准。本篇报告中，我们按照国内机构投资者使用较为广泛的一种标准——采用申万一级行业指数，对全 A 股市场的股票进行行业分类。

申万一级行业指数是按照自由流通市值加权计算的派氏指数，采用 A 股全市场样本计算。申万一级行业分类标准参考了中国证监会分类、MSCI 等权威机构的行业分类标准，并结合我国证券市场实际情况调整制定，通过行业划分尽可能地实现行业内市场特征的相似和行业间市场特征的差异，其中包括了 23 个一级行业，分别是农林牧渔、采掘、化工、黑色金属、有色金属、建筑建材、机械设备、电子、交运设备、信息设备、家用电器、食品饮料、纺织服装、轻工制造、医药生物、公用事业、交通运输、房地产、金融服务、商业贸易、餐饮旅游、信息服务和综合。

根据申万行业分类标准确定上市公司行业归属时，以主营业务收入和主营业务利润的构成为主要依据，分别在 6 月、9 月的 25 日根据上市公司主营业务调整的公告，在 7 月、10 月的第 1 个交易日调整行业成份股。

## 三、SMIA 量化行业配置模型的构建

行业轮动背后有着复杂的经济学逻辑，我们这里不对此进行深入研究，我们主要关心的是行业轮动所带来市场数据有规律的变化，并从这些市场数据中去挖掘潜在的优势行业。

之前分析过，行业市场排名存在一定的记忆性——每一期行业收益率的排名在一定程度上依赖于上一期的排名，这也正是行业轮动的外在表现。理论上来说，我们可以从历史上找到一些和当期行业排名相同的样本，观察他们下一期行业轮动所产生的变化，然后从统计上找到下一期表现持续较好的一些行业进行超配，并回避一些表现持续较差的行业。

但是在实际操作中遇到的问题是，由于A股市场成立时间相对较短，对于很多的行业划分来看，不可能在历史上找到足够多的和当期行业收益率具有一致排名的样本。例如，当采用申万一级行业指数时，共包含有23个行业分类，那么行业收益率的排序状态就有 $23! = 25852016738884976640000$ 种可能。显然，历史状态空间中是不可能收录所有行业排名状态的，也就是说，历史状态空间只可能是状态空间的一个子集。

在历史数据库中找到几个相同的行业排名状态是不现实的，因此，我们只能在有限的历史状态空间中寻找和当前行业排名状态最为相似的几个状态。

如何寻找最相似的行业市场排名状态呢？我们先来定性地看一个例子。在图3中，我们给出一个排名状态s，那么在A、B、C、D四个选项中，哪一个和状态s最为相似？

图3：排名状态相似性匹配实例

状态s	1	2	3	4	5	6	7	8
A	5	4	6	7	1	3	2	8
B	6	4	3	1	5	8	7	2
C	1	3	2	4	5	6	8	7
D	8	7	3	5	1	6	4	2

数据来源：广发证券发展研究中心

通过仔细观察，绝大多数人会选择C这个状态。为什么会选择C呢？直观感觉的



背后总是存在理性思考的。假设两个长度相等的序列  $\{x_i\}$  和  $\{y_i\}$ ，它们之间相似性的度量可以通过两序列之间的距离进行度量。

两序列距离的定义有很多种，我们接下来介绍几类常用的距离。

### 1、欧氏距离

欧氏距离 (Euclid Distance) 也称欧几里得度量、欧几里得距离，是最为常用的距离定义：

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

### 2、曼哈顿距离

曼哈顿距离 (Manhattan Distance) 是由闵可夫斯基在十九世纪所创立的，用以标明两个点上在空间坐标上的绝对轴距总和：

$$d(x, y) = \sum_i |x_i - y_i| \quad (2)$$

### 3、闵可夫斯基距离

$$d(x, y) = \left[ \sum_i |x_i - y_i|^m \right]^{1/m} \quad (3)$$

欧式距离和曼哈顿距离可以看作闵可夫斯基距离在  $m=2$  和  $m=1$  情况下的特例。而当  $m \rightarrow \infty$  时，所定义的距离被称之为切比雪夫距离 (Chebyshev distance)，为两个序列之间的元素绝对值差值的最大值：

$$d(x, y) = \max |x_i - y_i| \quad (4)$$

无论采用上述哪种距离的定义，我们定量去计算上述相似性匹配实例 (图 3)，都可以得到序列 C 是序列 S 的最相似匹配选项。也就是说，通过求解当前序列与历史状态空间中序列两两之间的距离，并且经过简单的排序，就可以找到历史数据库中最为匹配的相似序列。

只寻找一期最相似的行业市场排序容易造成预测结果波动较大、预测风险较高，因此，我们需要在历史上寻找几期最为相似的排名状态，观察他们之后一期的行业轮动与变化。这里又会出现两种统计的方法：以选取超配行业为例，一种是观察这些相似期之后一期排名相对靠前的行业，对它们取交集；另一种是观察这些相似期之后一期排名非常靠前的行业，对他们取并集。示意图如图 4 所示，其中每个数字代表申万一级行业指数中的一个行业。

通过分析，我们认为前一种方法所取到的是表现可能相对较好的一些行业；而后一种方法取到的是表现可能非常好的一些行业。因此我们认为后一种方法的效果可能更好，这一点也在我们的回测计算中得到了肯定。因此，在后面的实证中，我们全部按照方法 (b)，即通过寻找相似期之后一期少量排名非常靠前的行业，对他们取并集，得到推荐超配的优势行业组合。

图 4：进行相似性匹配统计的两种方法

方法 (a) 行业收益率由大到小排序→

1	2	15	7	3	11	23	16	10	17	5	6	12	4	21	8	20	18	9	22	13	14	19
8	1	6	23	9	22	14	20	19	17	5	16	4	13	3	12	21	7	11	10	18	15	2
1	17	9	18	8	15	6	22	2	16	21	3	14	13	11	19	5	20	10	23	12	4	7
11	17	9	3	8	22	10	18	12	23	6	5	20	13	7	21	4	14	16	2	19	15	1

取交集

方法 (b) 行业收益率由大到小排序→

1	2	15	7	3	11	23	16	10	17	5	6	12	4	21	8	20	18	9	22	13	14	19
8	1	6	23	9	22	14	20	19	17	5	16	4	13	3	12	21	7	11	10	18	15	2
1	17	9	18	8	15	6	22	2	16	21	3	14	13	11	19	5	20	10	23	12	4	7
11	17	9	3	8	22	10	18	12	23	6	5	20	13	7	21	4	14	16	2	19	15	1

取并集

数据来源：广发证券发展研究中心

## 四、SMIA量化行业配置模型在A股市场中的实证

按照上述 SMIA 模型构建思路，我们首先选取申万一级行业指数中的行业作为预配置标的。由于申万一级行业中“综合”行业的个股对于主营业务界定不是非常清晰，并且个股走势分化程度较大，因此我们认为其在行业收益率上的排名具有较多随机成分，容易对其他行业的相似性匹配造成较大影响，因此我们首先剔除掉该行业，保留剩余的 22 个行业进行统计与计算。

申万一级行业指数基期设定在 1999 年 12 月 31 日，基点为 1000 点。因此，我们的历史状态空间样本从 2000 年的开始取起，样本数据库窗口的截止时间为预测期的前一期。

对于预测周期，我们考虑到按照随机游走理论，各个行业收益率的波动与时间具有正相关关系；而如果以周为预测单位，由于各行业表现区分度不大、收益率较为接近，市场表现排名中可能会出现较多的噪声成分；因此，我们采用按月为周期进行预测，即通过对当月各行业收益率排名的情况，在行业收益率月度排名的历史数据库中进行相似性匹配，找到几期最为相似状态，观察这几期后一期的行业表现，取少量排名靠前行业的并集，作为推荐超配的优势行业。

从实际投资应用的角度，我们采用沪深300指数作为比较基准，从2006年开始对SMIA量化行业配置模型进行回测，结果如表1和图5所示。

表 1: SMIA 回测结果详细列表

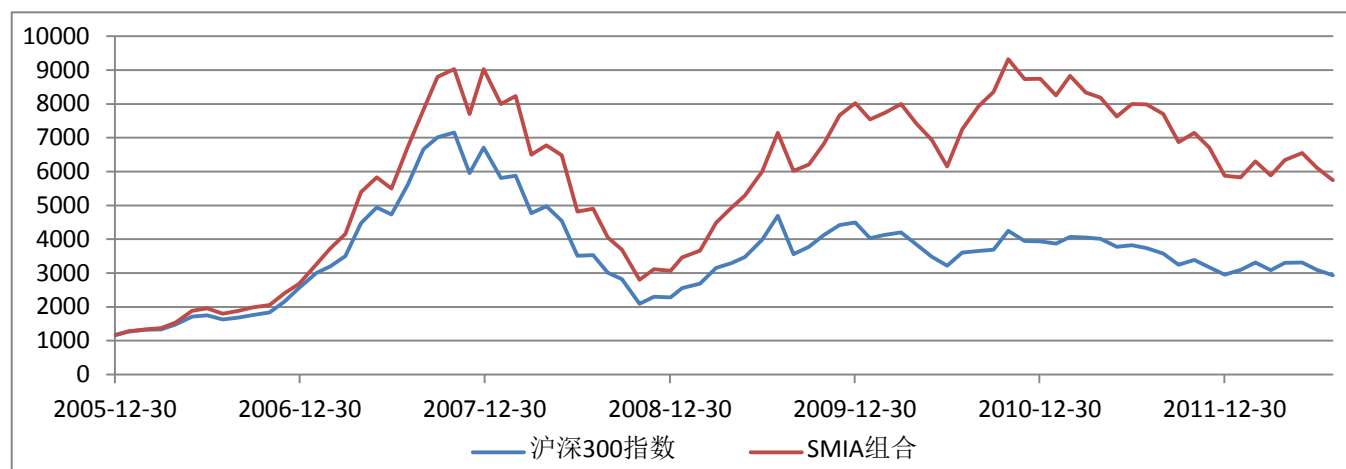
预测时间	推荐行业	推荐行业下期平均收益率 $\bar{r}_i$	$\sigma_{r_i}$	$r_{HS300}$ 下期	超额收益率	预测结果
2005-12-30	黑色金属, 电子, 房地产, 金融服务	9.83%	2.03%	9.33%	0.50%	正确
2006-1-25	采掘, 化工, 黑色金属, 餐饮旅游	4.42%	2.27%	4.30%	0.12%	正确
2006-2-28	黑色金属, 有色金属, 金融服务, 餐饮旅游	2.93%	5.53%	0.77%	2.16%	正确
2006-3-31	采掘, 机械设备, 电子, 金融服务	11.82%	4.06%	10.49%	1.33%	正确
2006-4-28	农林牧渔, 交运设备, 食品饮料, 轻工制造	23.15%	5.56%	16.47%	6.68%	正确
2006-5-31	化工, 有色金属, 轻工制造, 交通运输	4.07%	4.03%	2.09%	1.98%	正确
2006-6-30	农林牧渔, 采掘, 有色金属	-8.22%	3.04%	-7.15%	-1.07%	错误
2006-7-31	交运设备, 信息设备, 轻工制造, 商业贸易	4.75%	1.28%	3.43%	1.32%	正确
2006-8-31	采掘, 医药生物, 金融服务	5.36%	6.35%	4.82%	0.54%	正确
2006-9-29	采掘, 有色金属	3.10%	1.10%	4.36%	-1.26%	错误
2006-10-31	有色金属, 房地产, 餐饮旅游	17.09%	8.22%	17.06%	0.03%	正确
2006-11-30	黑色金属, 有色金属, 餐饮旅游	11.88%	4.88%	19.06%	-7.18%	错误
2006-12-29	农林牧渔, 采掘, 黑色金属	21.30%	3.19%	16.87%	4.43%	正确
2007-1-31	采掘, 黑色金属, 信息设备, 轻工制造	14.54%	2.95%	6.68%	7.86%	正确
2007-2-28	有色金属, 食品饮料, 轻工制造, 医药生物	11.34%	3.83%	9.32%	2.02%	正确
2007-3-30	农林牧渔, 电子, 信息设备, 交通运输	29.87%	3.36%	27.93%	1.94%	正确
2007-4-30	黑色金属, 有色金属, 信息设备, 金融服务	7.95%	3.17%	10.38%	-2.43%	错误
2007-5-31	黑色金属, 交运设备, 金融服务	-5.71%	5.99%	-4.17%	-1.54%	错误
2007-6-29	采掘, 建筑建材, 轻工制造, 金融服务	22.40%	7.35%	18.50%	3.90%	正确
2007-7-31	黑色金属, 食品饮料, 交通运输	16.33%	5.82%	18.75%	-2.42%	错误
2007-8-31	采掘, 化工, 家用电器, 餐饮旅游	12.55%	12.51%	5.36%	7.19%	正确
2007-9-28	化工, 黑色金属, 金融服务	2.53%	8.68%	1.93%	0.60%	正确
2007-10-31	黑色金属, 食品饮料, 纺织服装, 金融服务	-14.66%	2.79%	-16.72%	2.06%	正确
2007-11-30	采掘, 黑色金属, 商业贸易	17.14%	8.03%	12.68%	4.46%	正确
2007-12-28	电子, 医药生物, 金融服务	-11.43%	5.65%	-13.45%	2.02%	正确
2008-1-31	黑色金属, 有色金属, 食品饮料, 房地产	3.05%	4.51%	1.17%	1.88%	正确
2008-2-29	采掘, 有色金属, 轻工制造	-21.11%	4.04%	-18.91%	-2.20%	错误
2008-3-31	电子, 食品饮料, 金融服务	4.30%	11.42%	4.45%	-0.15%	错误
2008-4-30	采掘, 黑色金属, 信息设备, 餐饮旅游	-4.34%	5.95%	-8.78%	4.44%	正确
2008-5-30	黑色金属, 机械设备, 交运设备, 食品饮料	-25.72%	4.42%	-22.69%	-3.03%	错误
2008-6-30	黑色金属, 有色金属, 房地产	1.80%	3.02%	0.48%	1.32%	正确
2008-7-31	采掘, 机械设备, 轻工制造, 金融服务	-17.29%	7.56%	-14.74%	-2.55%	错误
2008-8-29	采掘, 电子, 金融服务, 餐饮旅游	-8.90%	3.98%	-6.19%	-2.71%	错误
2008-9-26	有色金属, 纺织服装, 医药生物, 公用事业	-24.09%	7.69%	-25.85%	1.76%	正确
2008-10-31	采掘, 房地产	11.11%	7.25%	9.99%	1.12%	正确
2008-11-28	黑色金属, 有色金属, 金融服务	-1.52%	4.37%	-0.67%	-0.85%	错误
2008-12-31	医药生物, 房地产, 金融服务	12.97%	3.40%	11.83%	1.14%	正确
2009-1-23	信息设备, 纺织服装, 房地产	5.60%	2.49%	5.30%	0.30%	正确
2009-2-27	采掘, 房地产, 金融服务	22.56%	3.96%	17.16%	5.40%	正确
2009-3-31	采掘, 信息设备, 金融服务, 商业贸易	9.95%	5.35%	4.59%	5.36%	正确



2009-4-30	采掘, 黑色金属, 医药生物	7.28%	7.27%	5.22%	2.06%	正确
2009-5-27	食品饮料, 纺织服装, 轻工制造, 金融服务	13.41%	8.22%	14.74%	-1.33%	错误
2009-6-30	采掘, 电子, 交运设备, 纺织服装	19.13%	6.91%	17.94%	1.19%	正确
2009-7-31	有色金属, 纺织服装, 医药生物	-15.86%	9.49%	-24.22%	8.36%	正确
2009-8-31	采掘, 黑色金属, 交运设备, 食品饮料	3.16%	2.90%	6.17%	-3.01%	错误
2009-9-30	农林牧渔, 采掘, 黑色金属, 餐饮旅游	10.03%	3.85%	9.17%	0.86%	正确
2009-10-30	农林牧渔, 有色金属, 信息设备	12.26%	4.40%	7.05%	5.21%	正确
2009-11-30	农林牧渔, 黑色金属, 交运设备	4.64%	1.70%	1.82%	2.82%	正确
2009-12-31	采掘, 信息设备, 金融服务, 餐饮旅游	-5.93%	8.16%	-10.39%	4.46%	正确
2010-1-29	黑色金属, 电子, 金融服务, 信息服务	2.53%	1.34%	2.42%	0.11%	正确
2010-2-26	采掘, 建筑建材, 轻工制造, 房地产	3.35%	0.90%	1.95%	1.40%	正确
2010-3-31	黑色金属, 有色金属, 金融服务	-7.06%	3.78%	-8.32%	1.26%	正确
2010-4-30	有色金属, 建筑建材, 医药生物, 信息服务	-6.69%	4.26%	-9.59%	2.90%	正确
2010-5-31	采掘, 黑色金属, 电子	-11.27%	2.21%	-7.58%	-3.69%	错误
2010-6-30	采掘, 建筑建材, 房地产, 餐饮旅游	17.80%	2.23%	11.93%	5.87%	正确
2010-7-30	电子, 交运设备, 食品饮料, 轻工制造	9.40%	2.60%	1.20%	8.20%	正确
2010-8-31	有色金属, 医药生物, 金融服务	5.35%	6.10%	1.12%	4.23%	正确
2010-9-30	交运设备, 轻工制造, 金融服务	11.62%	3.31%	15.14%	-3.52%	错误
2010-10-29	农林牧渔, 黑色金属, 有色金属, 轻工制造	-6.27%	5.53%	-7.19%	0.92%	正确
2010-11-30	农林牧渔, 采掘, 黑色金属, 纺织服装	0.03%	2.59%	-0.28%	0.31%	正确
2010-12-31	黑色金属, 有色金属, 食品饮料, 医药生物	-5.60%	5.49%	-1.65%	-3.95%	错误
2011-1-31	采掘, 有色金属, 房地产, 金融服务	7.02%	3.21%	5.30%	1.72%	正确
2011-2-28	农林牧渔, 电子, 交运设备, 交通运输	-5.57%	1.90%	-0.50%	-5.07%	错误
2011-3-31	有色金属, 家用电器, 食品饮料, 金融服务	-1.79%	3.27%	-0.95%	-0.84%	错误
2011-4-29	建筑建材, 房地产, 金融服务, 商业贸易	-6.88%	1.35%	-5.99%	-0.89%	错误
2011-5-31	交运设备, 纺织服装, 轻工制造, 房地产	4.84%	2.16%	1.42%	3.42%	正确
2011-6-30	黑色金属, 机械设备, 家用电器, 医药生物	-0.05%	3.25%	-2.37%	2.32%	正确
2011-7-29	采掘, 家用电器, 餐饮旅游	-3.62%	4.57%	-4.22%	0.60%	正确
2011-8-31	黑色金属, 电子, 纺织服装, 金融服务	-10.85%	2.37%	-9.32%	-1.53%	错误
2011-9-30	黑色金属, 有色金属, 金融服务, 餐饮旅游	3.99%	3.54%	4.41%	-0.42%	错误
2011-10-31	有色金属, 房地产, 商业贸易, 餐饮旅游	-6.07%	2.95%	-6.45%	0.38%	正确
2011-11-30	有色金属, 家用电器, 食品饮料, 商业贸易	-12.44%	4.11%	-6.97%	-5.47%	错误
2011-12-30	食品饮料, 金融服务, 信息服务	-0.70%	6.17%	5.05%	-5.75%	错误
2012-1-31	黑色金属, 有色金属, 交通运输	8.12%	1.05%	6.89%	1.23%	正确
2012-2-29	农林牧渔, 交运设备, 医药生物, 餐饮旅游	-6.63%	2.51%	-6.80%	0.17%	正确
2012-3-30	有色金属, 交运设备, 食品饮料, 金融服务	7.72%	0.37%	6.98%	0.74%	正确
2012-4-27	电子, 信息设备, 医药生物	3.22%	2.54%	0.22%	3.00%	正确
2012-5-31	采掘, 有色金属, 医药生物	-6.54%	8.46%	-6.48%	-0.06%	错误
2012-6-29	有色金属, 电子, 食品饮料, 餐饮旅游	-6.14%	3.95%	-5.23%	-0.91%	错误
2012-7-31	采掘, 交运设备, 家用电器, 餐饮旅游	-3.01%	1.47%	-5.49%	2.48%	正确
平均值	-	-	-	-	0.95%	-
标准差	-	-	-	-	3.19%	-

数据来源：广发证券发展研究中心

图 5: SMIA 组合与沪深 300 指数走势对比 (2006 年起)



数据来源: 广发证券发展研究中心

从上述回测结果来看, SMIA 量化资产配置模型可以在统计上有效筛选出相对表现较好的一些行业。在 2005 年 12 月 30 日至 2012 年 7 月 31 日的 80 次预测中, 有 54 次预测正确(我们将推荐行业组合平均收益率超过沪深 300 指数基准收益率定义为正确), 26 次失败, 预测正确率达到 67.5%; 并且在绝对值上, 盈利平均超额收益率略高于亏损平均超额收益率。

表 2: SMIA 模型回测效果统计 (2006 年起)

总预测次数	80
正确次数	54
错误次数	26
盈利平均超额收益率	2.594%
亏损平均超额收益率	-2.455%
成功率	67.5%

数据来源: 广发证券发展研究中心

另外从图 4 中可以更为直观地看出, 牛市中的预测效果要好于熊市中的预测效果。不过 2012 年以来的 7 次预测中, 仍然有 5 次正确。各年度预测效果的详细情况可以参见表 3 和图 6。

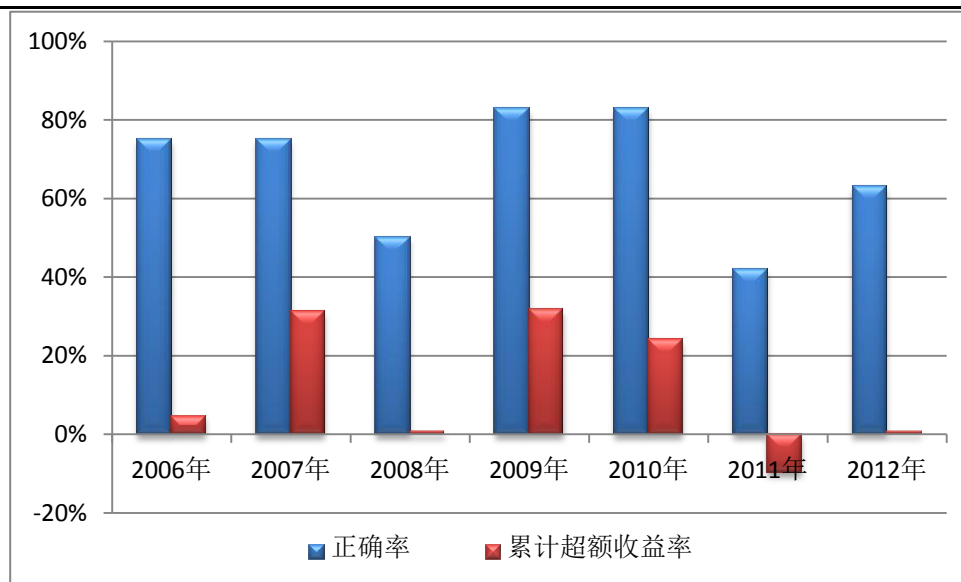
表 3: SMIA 模型各年度预测效果 (2006 年起)

年度	预测次数	正确次数	正确率	累计超额收益率
2006 年	12	9	75%	4.70%
2007 年	12	9	75%	31.17%
2008 年	12	6	50%	0.74%
2009 年	12	10	83%	31.63%
2010 年	12	10	83%	24.09%
2011 年	12	5	42%	-9.72%

2012 年	8	5	63%	0.64%
平均值	-	-	67%	11.89%
标准差	-	-	16%	16.73%

数据来源：广发证券发展研究中心

图 6：SMIA 模型各年度预测效果（2006 年起）

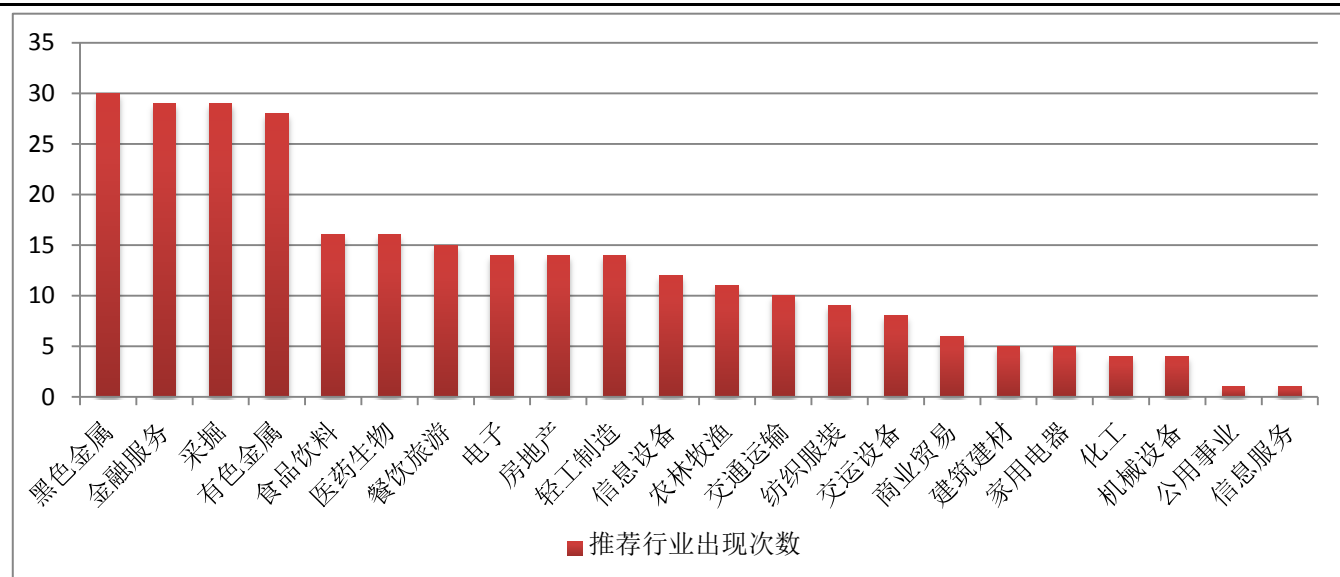


数据来源：广发证券发展研究中心

可以看出，总体表现最差的年度分别为 2008 年、2011 年和 2012 年，均为熊市年份。特别是 2011 年，出现了 2006 年以来唯一的一次负超额收益，判断正确率也出现了唯一的一次低于 50%。

最后，我们对所有行业被推荐的次数进行了简单统计，如图 7 所示。

图 7：SMIA 模型各行业推荐次数（2006 年起）



数据来源：广发证券发展研究中心

从图 7 中可以看出，2006 年以来推荐次数最多的行业是黑色金属、金融服务、采掘和有色金属，平均推荐次数在 25-30 次；推荐次数最少的是公用事业与信息服务，都仅推荐过 1 次。按照我们的统计方式来看，这样的推荐次数分布是和历史上实际行业排名靠前的次数具有紧密联系的——历史上排名靠前的次数越多，从统计上来讲，该行业被推荐概率也就越大。另外从时间维度总体来看，SMIA 模型在牛市环境中推荐的周期类行业次数较多，而在熊市环境中推荐的防御类行业次数较多，这也是符合金融市场变化特征的。

## 五、总结

本篇量化行业配置报告所提供的模型与一般的行业配置策略不同，我们抛开行业基本面的变化，完全从市场数据的角度出发，构建了 SMIA 量化模型。报告的出发点是：行业轮动具有记忆性，即当期的行业收益率排名会影响下一期行业相对表现的情况，这是由行业之间错综复杂的经济关系所决定的。我们通过寻找历史状态空间中与当期行业排名相似的一些时间点，观察这些时间点之后一期行业轮动的变化特征，从中选取统计上表现较好的行业，作为当前时间点下一期的推荐超配行业。

通过回测 2006 年起至今的月度数据，SMIA 模型表现良好，80 个月推荐的 80 次行业组合中，有 54 个月判断正确（获取超额收益），正确率 67.5%；并且在绝对值上，盈利平均超额收益率略高于亏损平均超额收益率。

对于 SMIA 模型，同我们之前发布的另类交易策略研究报告《基于时域分形的相似性匹配日内低频交易策略（SMT）》一样，都是依赖于历史数据库所构建的量化模型，因此随着历史数据库样本的不断增长，该类相似性匹配策略的预测准确性和投资参考价值都将有望不断提高。这让我们不由地想起一句话：历史从我们的脚印里走出来。

## 广发金融工程研究小组

罗军，首席分析师，华南理工大学理学硕士，2010 年进入广发证券发展研究中心。

俞文冰，首席分析师，CFA，上海财经大学统计学硕士，2012 年进入广发证券发展研究中心。

叶涛，资深分析师，CFA，上海交通大学管理科学与工程硕士，2012 年进入广发证券发展研究中心。

安宁宁，资深分析师，暨南大学数量经济学硕士，2011 年进入广发证券发展研究中心。

胡海涛，分析师，华南理工大学理学硕士，2010 年进入广发证券发展研究中心。

夏满阳，分析师，上海交通大学金融工程硕士，2012 年进入广发证券发展研究中心。

汪鑫，分析师，中国科学技术大学金融工程硕士，2012 年进入广发证券发展研究中心。

李明，分析师，伦敦城市大学卡斯商学院计量金融硕士，2010 年进入广发证券发展研究中心。

蓝昭钦，分析师，中山大学理学硕士，2010 年进入广发证券发展研究中心。

史庆盛，研究助理，华南理工大学金融工程硕士，2011 年进入广发证券发展研究中心。

张超，研究助理，中山大学理学硕士，2012 年进入广发证券发展研究中心。

## 相关研究报告

基于时域分形的相似性匹配日内低频交易策略（SMT）	张超	2012-09-17
---------------------------	----	------------

	广州市	深圳市	北京市	上海市
地址	广州市天河北路 183 号 大都会广场 5 楼	深圳市福田区民田路 178 号华融大厦 9 楼	北京市西城区月坛北街 2 号 月坛大厦 18 层	上海市浦东新区富城路 99 号震旦大厦 18 楼
邮政编码	510075	518026	100045	200120
客服邮箱	gfyf@gf.com.cn			
服务热线	020-87555888-8612			

## 免责声明

广发证券股份有限公司具备证券投资咨询业务资格。本报告只发送给广发证券重点客户，不对外公开发布。

本报告所载资料的来源及观点的出处皆被广发证券股份有限公司认为可靠，但广发证券不对其准确性或完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券或其附属机构的立场。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且不予通告。

本报告旨在发送给广发证券的特定客户及其它专业人士。未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。