

林晓明 执业证书编号：S0570516010001
 研究员 0755-82080134
 linxiaoming@htsc.com

基于通用回归模型的行业轮动策略

华泰行业轮动系列之一

相关研究

- 1 《金工：人工智能选股之 Python 实战》
2017.09
- 2 《金工：人工智能选股之 Boosting 模型》
2017.09
- 3 《金工：人工智能选股之随机森林模型》
2017.08

采用通用回归预测模型挖掘行业背后的交叉预测关系

本文实证了一种基于通用回归预测模型来挖掘行业轮动规律并指导行业配置的方法，其核心思想是利用所有行业当期收益率作为解释因子，对每个行业下期收益率做回归，考察行业收益率之间是否存在显著的交叉预测关系，进而构建轮动策略。该模型的提出主要基于两点考虑：1、各行业之间存在广泛而紧密的联系，通过回归模型，可以定量刻画行业收益率之间的传导规律；2、行业指数本身就是观测宏观经济运行状态的一个很好的窗口，而行业指数的收益率则刻画了经济体的动态变化情况，通过挖掘各行业收益率之间的传导关系，可以用来指导行业配置。

基于主成分回归模型的行业轮动策略构建与回测

基于主成分回归模型的行业轮动策略构建包括数据预处理、主成分合成、回归参数估计等步骤，最终在每个截面处，可以生成下一期各行业收益率的预测值，并指导最终的配置。实证表明，当训练窗口长度为 242 周，全部主成分都参与回归时，该行业轮动策略的多空对冲组合年化收益率达到 28.68%，年化波动率 19.04%，夏普比率为 1.51，而纯多头组合的年化收益率为 25.28%，年化波动率 27.3%，夏普比 0.9，多空对冲策略表现更优。

参与回归主成分越多，训练窗口越长，策略表现越稳定

在基于主成分回归模型的行业轮动策略构建中，最核心的两个参数是训练窗口长度，以及参与回归的主成分个数。我们遍历了参数的各种可能取值，来检验该策略对参数的敏感程度。实证结果表明，当主成分个数大于 3 时，几乎所有训练窗口长度下，多空策略都获得了正收益，而且随着主成分增多，平均收益率也不断抬升；另一方面，当主成分个数固定时，随着训练窗口长度变长，策略表现也更优，说明通过较长训练窗口挖掘出的轮动规律更稳定，外延预测时更可靠。整体而言，该策略鲁棒性较强，对参数取值不是十分敏感。

业绩归因模型显示，主成分回归行业轮动模型获取了显著的 alpha 收益

我们通过构建一个 5 因子模型对行业轮动策略的收益率进行归因，测试其在各个因子上的风险暴露，以此来解释策略收益的来源。其中，5 个因子分别是市场因子、市值因子、账面市值比因子、反转因子、流动性因子。实证结果表明，多空对冲组合在 5% 置信度水平下，获得了 17.32% 的年化超额收益，并且仅在市值因子上有轻微暴露。说明该策略的收益来源主要是 alpha，而不是通过在特定风险因子上持续暴露敞口来获得收益补偿。或者说，该多空策略可能在不同市场环境下把握了占主导地位的风格因子，而不是持续押注于某一类特定的风格。

A 股短期或有回调，但长期上升趋势不变

通过对 2016 年 5 月至今这段慢牛行情进行回归建模，并绘制板块动力图，可以发现周期上游是本轮行情的动力节点，正向推动了其他板块，而进入 9 月以来，市场风格有所切换，作为上涨引擎的周期上游板块陷入调整，整个市场向上动力不足，出于避险的考虑，资金纷纷进入防御板块抱团取暖，所以最近消费、周期中游板块获得显著超额收益，这说明市场目前整体处于防守状态，大概率会陷入调整。但我们认为这个短期回调是为未来的上涨积蓄力量，市场长期向上趋势不变，而防守板块没法带领市场不断冲高，调整到位后，资金大概率会回流到周期上游、大金融等主导板块。

风险提示：通用回归模型是历史经验的总结，存在失效的可能。

正文目录

本文研究导读	5
基于回归预测模型挖掘行业轮动背后规律	6
回归预测模型构建	6
数据采集	7
预处理	7
多重共线性处理	8
显著性检验	9
回归预测模型实证分析	10
2005年1月~2017年10月	10
2013年1月~2015年6月	12
2015年6月~2016年1月	13
2016年5月~2017年10月	14
基于主成分回归的行业轮动策略原理与实证检验	16
主成分回归模型构建	16
分层回测	20
多空回测	22
参数敏感性分析	24
基于主成分回归的行业轮动策略业绩归因	27
业绩归因模型	27
业绩归因实证分析	28
附录	31
CT 统计量推导	31

图表目录

图表 1: 时间序列回归示意图.....	6
图表 2: 回归预测模型构建流程.....	6
图表 3: 六大板块构成.....	7
图表 4: 2005 年 1 月至 2017 年 10 月六大板块周频收益率统计.....	7
图表 5: 六大板块相关性分析.....	8
图表 6: 六大板块方差膨胀因子.....	8
图表 7: 板块动力图示例.....	9
图表 8: 2005 年 1 月~2017 年 10 月六大板块净值图.....	10
图表 9: 2005 年 1 月~2017 年 10 月六大板块回归系数 (*, **, ***分别代表双边 1%, 5%, 10%置信度水平).....	11
图表 10: 2005 年 1 月~2017 年 10 月六大板块动力图.....	11
图表 11: 2013 年 1 月~2015 年 6 月六大板块净值图.....	12
图表 12: 2013 年 1 月~2015 年 6 月六大板块回归系数 (*, **, ***分别代表双边 1%, 5%, 10%置信度水平).....	12
图表 13: 2013 年 1 月~2015 年 6 月六大板块动力图.....	12
图表 14: 2015 年 6 月~2016 年 1 月六大板块净值图.....	13
图表 15: 2015 年 6 月~2016 年 1 月六大板块回归系数 (*, **, ***分别代表双边 1%, 5%, 10%置信度水平).....	13
图表 16: 2015 年 6 月~2016 年 1 月六大板块动力图.....	13
图表 17: 2016 年 5 月~2017 年 10 月六大板块净值图.....	14
图表 18: 2016 年 5 月~2017 年 10 月六大板块回归系数 (*, **, ***分别代表双边 1%, 5%, 10%置信度水平).....	15
图表 19: 2016 年 5 月~2017 年 10 月六大板块动力图.....	15
图表 20: 行业轮动模型示意图.....	16
图表 21: 主成分回归流程.....	17
图表 22: 六大板块收益率协方差矩阵的特征值.....	17
图表 23: 六大板块收益率主成分分析.....	18
图表 24: 第一主成分在六大板块上的暴露.....	18
图表 25: 第一主成分与市场基准的净值走势.....	18
图表 26: 第二主成分在六大板块上的暴露.....	19
图表 27: 第二主成分与周期/非周期风格指数的净值走势.....	19
图表 28: 分层回测净值曲线 (分 3 层).....	20
图表 29: 分层回测业绩指标 (分 3 层).....	20
图表 30: 分层回测业绩指标对比图示 (分 3 层).....	21
图表 31: 分层组合净值除以基准组合净值示意图 (分 3 层).....	21
图表 32: 分层回测净值曲线 (分 6 层).....	21
图表 33: 分层回测业绩指标 (分 6 层).....	21
图表 34: 分层回测业绩指标对比图示 (分 6 层).....	22
图表 35: 分层组合净值除以基准组合净值示意图 (分 6 层).....	22

图表 36: 不同分层场景多头策略对比.....	22
图表 37: 多空策略净值曲线	22
图表 38: 多空策略业绩表现	23
图表 39: 多空策略净周收益率及累计收益率	23
图表 40: 多空策略业绩表现	23
图表 41: 参数遍历场景下的多空收益率图示	24
图表 42: 参数遍历场景下的多空收益率统计结果.....	24
图表 43: 参数遍历场景下的多空夏普比率图示	25
图表 44: 参数遍历场景下的多空夏普比率统计结果	26
图表 45: 5 因子累计净值曲线	28
图表 46: 5 因子月度收益率统计	29
图表 47: 5 因子之间相关性分析.....	29
图表 48: 去共线性后 5 因子累计净值曲线	29
图表 49: 基于主成分回归模型的行业轮动策略业绩归因结果.....	30

本文研究导读

行业轮动本质上就是根据不同行业的区间表现差异性进行轮动配置，力求能够抓住区间内表现较好的行业。在整个自上而下的资产配置投资体系中，行业配置占据着重要地位，投资者如果能够把握行业轮动规律，踏准市场节奏，就能够获得显著的超额收益。因而行业轮动也一直是业界研究的热点，在现有文献中，对行业轮动的研究主要集中在以下两个方面：

- 1、从经济周期角度看，不同行业会随着经济状态的周期性变化出现依次上涨、下跌的轮动现象。当经济复苏时，处于产业链下游的汽车、房地产等先导性行业将率先回暖，进而带动中游投资品，制造业的发展，最终传导到上游能源、原材料行业。当经济衰退时，对基本面最为敏感的钢铁、有色等强周期行业将率先下跌，而食品饮料、医药等必需消费品行业则受影响较小。
- 2、从行为经济学角度看，一方面，A股是一个受政策因素影响明显的新兴市场，投资者会受到政策预期的影响，导致投资决策趋同，最终政策受益行业会呈现出“强者恒强”的动量效应；另一方面，投资者先天的比价心理会强化补涨、补跌的认知，在上涨行情中，没有涨的行业就应该补涨，同样下跌行情中，没有跌的行业也应该无条件补跌，这种比价效应会引导市场资金在分化的行业之间流转，造成一种“你方唱罢我登场”的交错涨跌状态。

无论从哪个研究角度，都得到了很多有实践意义的结论，但各自也有不足的地方：

- 1、从经济周期出发，优点是逻辑性强，符合经济学常识，但实践起来较难，因为经济状态的划分本身就是一个难题，实证中甚至发现经济状态会出现频繁跳跃和反复。
- 2、从行为经济学出发，优点是逻辑简单，可操作性强，但模式过于单一，不够灵活，很多时候行业轮动规律是复杂，而且时变的，简单的动量或反转模式难以获得持续、稳定的超额收益。

综上，我们权衡考虑可行性与灵活性，实证了一种基于通用回归预测模型来挖掘行业轮动规律并指导行业配置的方法，其核心思想是利用所有行业当期收益率作为解释因子，对每个行业下期收益率做回归，考察行业收益率之间是否存在显著的交叉预测现象，进而构建轮动策略。该模型的提出主要考虑了以下几点：

- 1、金融市场是一个复杂的整体，各个行业之间存在许多广泛而紧密的联系，一个行业收益率的变化可能引起其他行业的收益率变化，同时该行业也可能受到其它行业收益率变化的影响。比如钢铁行业，既有可能受益于下游需求旺盛而上涨，其本身的产能扩张也会影响到上游能源行业。通过回归模型，可以定量刻画这种行业之间的传导规律，建立起复杂的行业关联关系网络。
- 2、整个金融经济体是一个相互影响，联系紧密的整体，就像宏观经济指标可以用来指导投资，而股市也可以作为经济晴雨表，反映基本面状态。我们认为各行业指数本身就是观测宏观经济运行状态的一个很好的窗口，而行业指数的收益率则刻画了经济体的动态变化情况，通过挖掘各行业收益率之间的传导关系，可以用来指导行业配置。

后文中，我们主要从三个方面来实证该通用回归模型的表现：首先，我们对历史行情中比较典型的几个时间段建立回归模型，以板块动力图的形式可视化展示行业收益率之间的相互传导关系，并与板块实际表现相互印证；然后，我们基于主成分回归模型构建定价方程，指导行业轮动配置，考察策略的净值回测结果，并进行参数敏感性分析；最后，我们对策略净值进行收益归因，分析策略的主要收益来源。

基于回归预测模型挖掘行业轮动背后规律

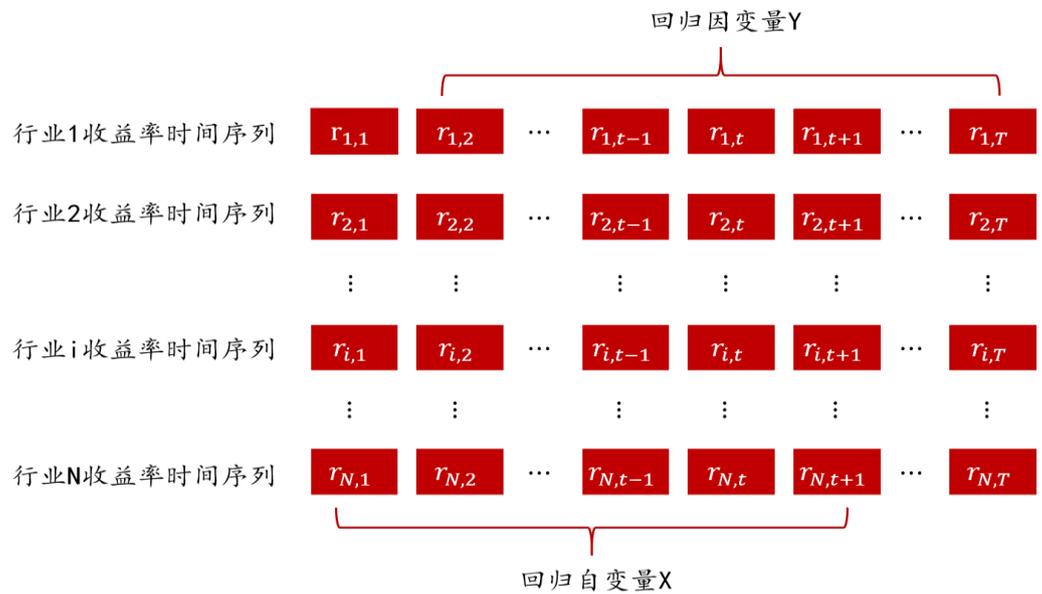
回归预测模型构建

为了考察行业收益率之间是否存在显著的交叉预测现象，我们利用所有行业当期收益率作为解释因子，对每个行业下期收益率做回归（参考图表1），构建如下时间序列回归模型：

$$r_{i,t+1} = \alpha_i + \sum_{j=1}^N \beta_{i,j} r_{j,t} + \varepsilon_{i,t+1} \text{ for } t = 1, 2 \dots T - 1$$

其中，N为行业个数， $r_{j,t}$ 代表行业j第t期收益率。上述回归方程定量描述了一个给定行业当期收益率受所有行业上期收益率的影响程度，在所有行业上建立这种回归模型后，就能刻画整个行业集群之间错综复杂的联系了。

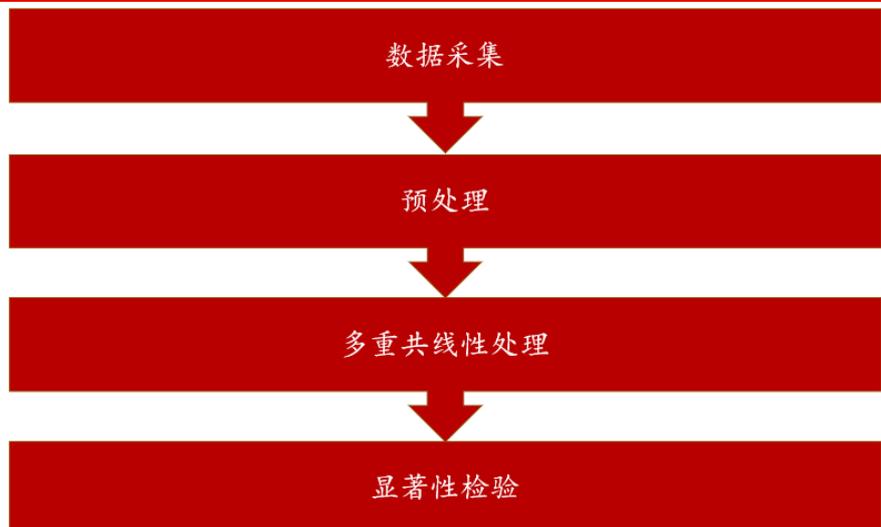
图表1：时间序列回归示意图



资料来源：华泰证券研究所

整个回归预测模型的构建流程如图表2所示，主要分为数据采集、预处理、多重共线性处理、回归方程显著性检验几部分，下面分别介绍。

图表2：回归预测模型构建流程



资料来源：华泰证券研究所

数据采集

数据采集工作分为两步：

1. 选取 29 个一级行业中除国防军工和综合之外的行业指数构建六大板块：周期上游、周期中游、周期下游、大金融、消费、成长。板块内，各行业指数采用等权方法合成板块指数。后文中，我们将统一基于六大板块进行分析。这里之所以不直接用一级行业指数进行回归，一方面是考虑到细分行业之间相关性更强，共线性更显著，影响回归效果；另一方面是考虑到行业数目越多，建立的行业关联网络越复杂，难以把握最核心的规律。
2. 得到六大板块指数后，计算其周频对数收益率，采用对数处理是为了让收益率序列更平稳，削弱数据的异方差和共线性，而且有利于计算。

六大板块划分如下表所示（具体的板块划分方式请参见华泰金工研报《行业指数频谱分析及配置模型》）。可以看到，行业聚类结果与传统定性行业分类结果一致，说明基本的产业经济逻辑是深深的融入到金融市场的运行过程之中的。

图表3：六大板块构成

板块	板块包含行业
周期上游	石油石化、煤炭、有色金属
周期中游	电力及公用事业、钢铁、基础化工、交通运输
周期下游	建筑、建材、汽车、机械
大金融	银行、非银行金融、房地产
消费	轻工制造、商贸零售、餐饮旅游、家电、纺织服装、医药、食品饮料、农林牧渔
成长	计算机、传媒、通信、电力设备、电子元器件

资料来源：Wind，华泰证券研究所

六大板块周频收益率序列的统计结果如下表，收益率最高的两个板块是大金融和消费，波动率最低的两个板块是周期中游和消费，夏普比最高的是大金融和消费，整体而言，消费板块在风险收益比上要优于其他板块。

图表4：2005年1月至2017年10月六大板块周频收益率统计

	均值	标准差	最大值	最小值	年化夏普比
周期上游	0.31%	4.62%	20.56%	-20.31%	0.48
周期中游	0.26%	4.14%	14.66%	-22.38%	0.46
周期下游	0.40%	4.57%	23.75%	-21.27%	0.63
大金融	0.45%	4.52%	20.21%	-16.29%	0.71
消费	0.41%	4.30%	17.71%	-19.41%	0.69
成长	0.39%	4.73%	17.10%	-20.06%	0.59

资料来源：Wind，华泰证券研究所

预处理

为了提升回归效果，在得到六大板块周频对数收益率序列之后，要进行去极值、中心化、标准化等预处理操作。本文采用“中位数去极值法”对极端值进行处理

$$\tilde{x}_i = \begin{cases} x_M + n * D_{MAD}, & \text{if } x_i > x_M + n * D_{MAD} \\ x_M - n * D_{MAD}, & \text{if } x_i < x_M - n * D_{MAD} \\ x_i, & \text{else} \end{cases}$$

x_M : 序列 x_i 的中位数

D_{MAD} : 序列 $|x_i - x_M|$ 的中位数

\tilde{x}_i : x_i 去极值修正后的值

数据去极值后的序列再进行中心化、标准化：

$$\tilde{x}_i = \frac{x_i - u}{\sigma}$$

u : 序列 x_i 的均值

σ : 序列 x_i 的标准差

\tilde{x}_i : 序列 x_i 标准化之后的值

多重共线性处理

共线性是指解释变量之间有精确的或者近似的线性关系。在我们的回归模型中，自变量是各个行业的收益率序列，而大部分时间里，各行业同涨齐跌的现象是比较显著的，导致其收益率序列之间相关程度较高，如图表5所示，消费、周期中游、周期下游三个板块的相关系数最高，均超过0.9，大金融和成长板块的相关系数最低，也超过了0.6。共线性的存在不会影响回归系数的唯一性和无偏性，但它会造成回归系数估计值的方差变大，使得回归系数的置信区间变得很宽，这样一方面会降低估计量的精准度，甚至改变了系数的符号；另一方面也可能使估计量的 T 值变小，导致一些对因变量有显著影响的解释变量通不过假设检验。所以，共线性与否直接关系到回归方程的优良性，必须加以处理。

图表5：六大板块相关性分析

	周期上游	周期中游	周期下游	大金融	消费	成长
周期上游	--	0.89	0.88	0.79	0.84	0.78
周期中游	0.89	--	0.95	0.77	0.93	0.88
周期下游	0.88	0.95	--	0.77	0.95	0.90
大金融	0.79	0.77	0.77	--	0.70	0.63
消费	0.84	0.93	0.95	0.70	--	0.94
成长	0.78	0.88	0.90	0.63	0.94	--

资料来源：Wind，华泰证券研究所

常见的判断共线的方法有条件数判别法和方差膨胀因子判别法：

- 条件数 κ 是指解释变量相关系数矩阵 R 的最大特征值与最小特征值之比。在应用中一般认为，若 $\kappa < 100$ ，则共线性的程度较小；若 $\kappa > 1000$ ，则存在严重的共线性。
- 方差膨胀因子（VIF）是指是指解释变量之间存在多重共线性时的方差与不存在多重共线性时的方差之比，VIF越大，共线性越严重。经验表明：当 $0 < VIF < 10$ ，不存在多重共线性；当 $10 \leq VIF < 100$ ，存在较强的多重共线性；当 $VIF \geq 100$ ，存在严重多重共线性。

这里我们采用方差膨胀因子法来检测共线性，对于已经通过中心化、标准化的收益率矩阵 X ，计算其相关系数矩阵 R ，并求逆，得到如下结果：

$$C = (c_{ij}) = R^{-1}$$

则称其主对角元素 $VIF_j = c_{jj}$ 为自变量 x_j 的方差膨胀因子，记 R_j^2 为自变量 x_j 对其余 $N-1$ 个自变量回归的拟合优度，容易证明：

$$VIF_j = c_{jj} = \frac{1}{1 - R_j^2}$$

也就是说，某个自变量 x_j 的方差膨胀因子越大，与其余 $N-1$ 个自变量的线性相关程度越强，其解释能力也越容易被其他变量替代。

六大板块方差膨胀因子计算结果如下，可以看到周期中游、周期下游、消费板块的方差膨胀因子均大于10，共线性比较明显，这与前面做相关分析的结论是一致的。

图表6：六大板块方差膨胀因子

	周期上游	周期中游	周期下游	大金融	消费	成长
方差膨胀因子	5.83	13.70	17.21	3.09	16.35	8.54

资料来源：Wind，华泰证券研究所

由于共线性的存在会严重影响回归效果，所以必须要进行相关处理。最直接的思路就是把共线性显著的自变量剔除，比如利用方差膨胀因子法，每次将VIF最大的自变量剔除，计算剩余自变量的VIF，直到所有自变量的VIF都在10以内，然后进行回归。实证结果显示，如果按照该方法处理，最后保留的自变量只剩下周期上游、周期中游、大金融、成长四大板块，被剔除的周期下游、消费板块对其他板块的影响就无法得到了，所以我们采用另外两种思路：

1. 逐步回归法，其基本思想是将变量逐个引入模型，每引入一个解释变量后都要进行F检验，并对已经选入的解释变量逐个进行t检验，当原来引入的解释变量由于后面解释变量的引入变得不再显著时，则将其删除，以确保每次引入新的变量

之前回归方程中只包含显著性变量。经过逐步回归，最后保留在模型中的解释变量既是重要的，又没有严重多重共线性。后文中，我们在对历史行情典型区间段进行建模时，将采用逐步回归法。

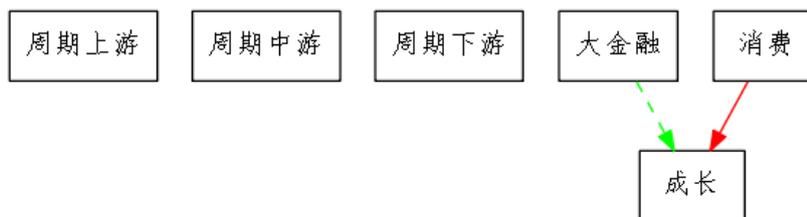
- 主成分回归法，其原理是将原来的解释变量通过线性变换重新组合成一组新的相互无关的变量，来替代原来的解释因子，构建回归方程。其优点是可以消除原始多元线性回归模型中的共线性影响，缺点是新合成的解释因子不一定有直观的经济学含义。我们在构建行业轮动策略时，将基于主成分回归法。

显著性检验

选取合适的回归模型后，就可以进行参数估计和显著性检验了。这里我们通过 t 检验挑选出显著不为 0 的回归系数，来刻画行业之间的相互联系。举个例子，假设针对成长板块构建的回归预测方程中，消费板块的回归系数显著不为 0 且是正值，则说明消费板块的本期收益率对成长板块的下期收益率有显著的正向推动作用，同理，如果大金融板块的回归系数显著不为 0 且为负值，则说明大金融板块的本期收益率与成长板块的下期收益率有一个负相关关系。

我们采用板块动力图的形式来可视化展示这种行业之间错综复杂的交叉预测关系，假设六大板块是有向图中的六个节点，显著不为 0 的正回归系数用红色实线边来表示，显著不为 0 的负回归系数用绿色虚线边来表示，则上述例子构建的板块动力图如下所示。

图表7：板块动力图示例



资料来源：华泰证券研究所

该有向图能非常直观、全景式的展示六大板块收益率之间的交叉预测关系，方便我们把握行业轮动的微观规律。但理解时需要强调以下几点：

- 回归方程展示的是一种统计显著关系，而非因果关系，比如上图中，我们能得到的是在回归建模的区间段内，消费板块延迟一期的收益率序列和成长板块的当期收益率序列之间有显著的正相关关系，而不能说消费板块的涨跌和成长板块的涨跌互为因果。当然，大概率情况下，通过模型构建出的板块动力图都能和符合经济学逻辑的结论相互印证。
- 通过某条边连接的两个节点之间并非你涨我跌或你跌我涨的绝对关系，而是一个相对概念，比如上图中，并非大金融板块上周涨了，成长板块这周就会跌，因为成长板块还受到了消费板块的影响，两者谁的贡献更大可能是动态改变的。

正确的理解方式是从全局上看板块动力图，比如某个板块正向影响了多个板块，说明该板块可能有先导效应，推动了市场的演进；又比如某个板块负向影响了多个板块，说明该板块走势可能和其他板块存在一定负相关关系，大概率是防御性板块。后文介绍中，我们将结合案例详细描述。

回归预测模型实证分析

本节我们将针对历史行情中比较典型的几个区间段进行回归建模，并绘制板块动力图，考察该回归预测模型的实际表现。

2005年1月~2017年10月

首先，我们针对整个历史区间进行分析，得到六大板块的净值走势如下图所示。整体而言，各行业之间同涨齐跌的现象还是比较明显的，但也有一些值得挖掘的细节：一方面，不同阶段的主导板块是不一样的，06年至07年的大牛市明显是在大金融，周期上游等大盘板块的引领下不断冲高的，而14年至15年的大牛市主导板块则变成了成长板块；另一方面，不同局部区间也能看到各大板块的走势分化，比如2010年下半年，周期上游和消费板块获得了显著的超额收益，而大金融板块则处于震荡下跌的行情中，同样的分化发生在2013年下半年，消费、成长板块明显跑赢周期上游。投资者如果能把握住这种风格上的切换，或者轮动的规律，就能获得显著的超额收益。

图表8：2005年1月~2017年10月六大板块净值图



资料来源：Wind，华泰证券研究所

进一步，我们对区间内的行业收益率做回归建模，得到回归系数估计如图表9所示。这里需要说明两点：

1. 为了克服共线性影响，我们采用了逐步回归的方式，对每个板块，都只保留了解释能力最强的几个因子，因而有些因子的回归系数为空。
2. 最终构建的回归方程中，拟合优度 R^2 的量级都比较低，这是正常的，因为行业收益率的影响因子是广泛而复杂的，往往含有不可预测的成分，我们仅仅挑选各板块上期收益率作为解释因子，容易出现欠拟合现象。但这里 R^2 较低并不代表回归模型没有意义，Campbell and Thompson (2008) 提出了一种计量方法，能够从定量角度刻画回归方程在经济学意义上的预测能力，其计算公式如下（具体推导请参见附录）：

$$CT = \left(\frac{R_i^2}{1 - R_i^2} \right) \left(\frac{1 + S_i^2}{S_i^2} \right)$$

其中， R_i^2 代表板块*i*回归方程的拟合优度， S_i^2 代表板块*i*的夏普比率（周频收益率均值除以标准差）。CT统计量的含义是：相比于完全忽略所有对超额收益有影响的解释变量，当投资者从风险项中剥离出一部分解释变量时，组合预期超额收益的增加幅度。更通俗一点的说，假设有两个采用均值方差模型的投资者A和B，投资者A在对风险资产的预期超额收益做预测时，仅仅利用历史的平均超额收益，不依赖任何预测模型，而投资者B则剥离出了一些对超额收益率有显著解释能力的因子，那么相比于A，投资者B能获得的预期超额收益率的增长率就是CT统计量。所以我们构建的回归模型虽然 R^2 较低，但只要这些因子的解释能力在下一期能保持，那期望超额收益率都能得到显著提升。从下表能看到，解释能力最弱的大金融板块也能获得89%的提升。

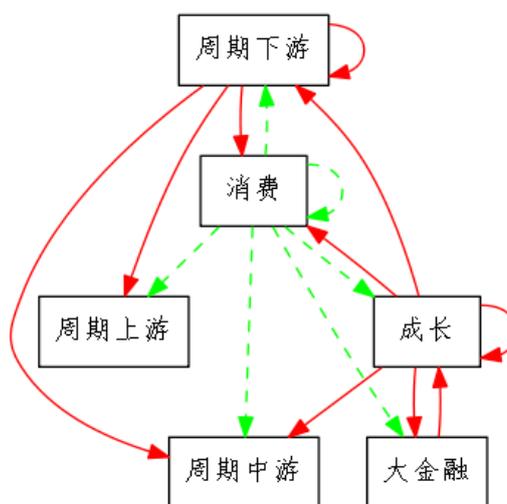
图表9： 2005年1月~2017年10月六大板块回归系数（*，，***分别代表双边1%，5%，10%置信度水平）**

	周期上游	周期中游	周期下游	大金融	消费	成长
周期上游						
周期中游						
周期下游	0.42***	0.36***	0.34***		0.33***	
大金融						0.15***
消费	-0.33***	-0.58***	-0.57***	-0.25**	-0.66***	-0.49***
成长		0.34***	0.32***	0.26**	0.44***	0.47***
R2	2.39%	3.61%	3.06%	0.85%	4.39%	3.86%
CT	5.57	9.22	4.14	0.89	5.03	6.06

资料来源：Wind，华泰证券研究所

得到回归方程后，我们筛选出置信度水平在5%以上的回归系数，并构建板块动力图（图表10），可以看到：

1. 成长板块作为动力节点，正向影响了除周期中游之外的其他板块，这意味着成长板块当期收益率对这些受影响板块的下期收益率有一个显著的正向推动作用，成长板块当期涨得越多，受影响板块下期也大概率涨得越多，或者跌得越少，反之亦然。那么我们可以认为成长板块大多数时间内就是市场的一个先导板块，这与我们的直观印象是一致的，因为05年至17年这段时间中，小盘风格更占优，成长板块大部分时间内都在引领市场的涨跌。
2. 消费板块作为防御节点，和其他所有板块的下期收益率都有一个显著的负相关关系，这意味着消费板块当期涨得越多，其他板块下期大概率涨得越少，或者跌得越多。这种轮动逻辑在市场上经常能看到，比如经济衰退，大盘向下时，消费板块中的医药、食品饮料行业由于其必需消费品属性，比较抗跌；又比如大盘经过一轮上涨行情，有调整需求时，消费板块也往往成为避险资金的集中地，所以消费板块的收益率往往和其他板块有一定的背离。
3. 周期下游对周期中游，以及周期上游的正向传导关系，这和我们前文所说的基本产业链逻辑是相互印证的。当经济复苏时，处于周期下游的汽车、建筑等先导性行业将率先回暖，进而带动中游交通运输，制造业的发展，最终传导到上游能源、原材料行业。这种传导机制在实体经济中可能需要一个比较长的过程，但资本市场上炒的是预期，逻辑兑现可能更快。

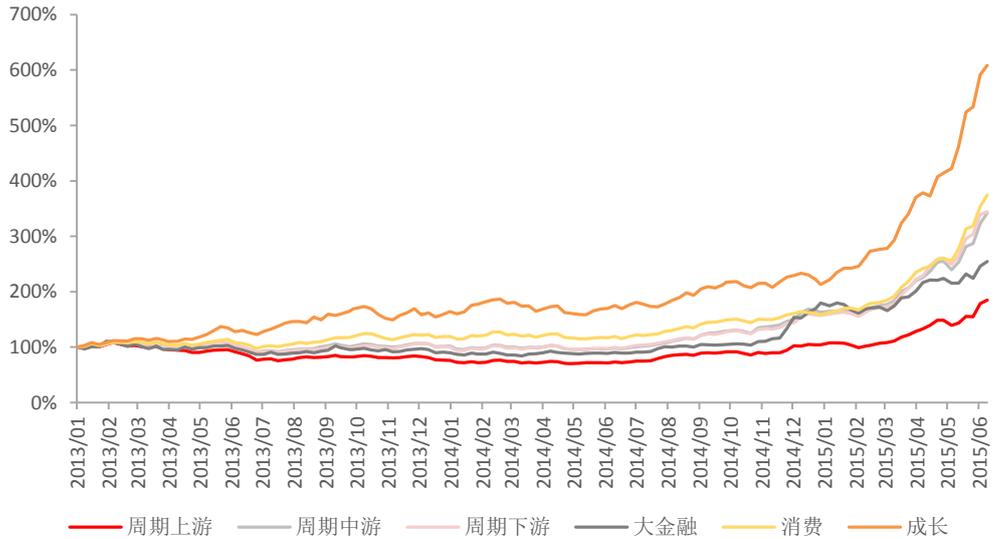
图表10： 2005年1月~2017年10月六大板块动力图

资料来源：Wind，华泰证券研究所

2013年1月~2015年6月

图表11~图表13是我们针对2013年1月~2015年6月这段大牛市行情的回归建模结论。从净值图能看到，成长板块主导了这轮牛市，而周期上游和大金融板块的涨幅则远远落后，这是因为自08年全球金融危机之后，宏观经济基本面一直处于下行趋势中，通缩预期促使政府采取降息等宽松政策，但基本面的恶化导致绝大部分企业盈利严重受损，充裕的资金并没有合适的去向，只能流入炒预期、看未来的成长板块。从板块动力图也能看出，成长板块的收益率，正向推动了周期中游、周期下游以及消费板块的收益率，也印证了成长板块是这一轮行情的动力节点、主导板块。

图表11： 2013年1月~2015年6月六大板块净值图



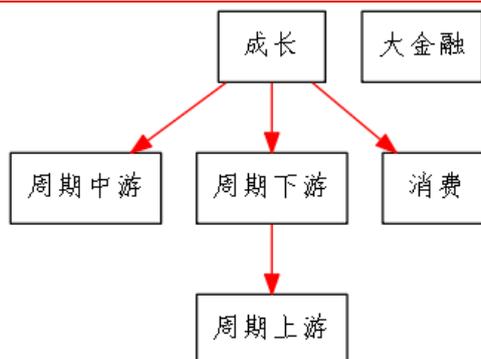
资料来源：Wind，华泰证券研究所

图表12： 2013年1月~2015年6月六大板块回归系数 (*, **, *分别代表双边1%, 5%, 10%置信度水平)**

	周期上游	周期中游	周期下游	大金融	消费	成长
周期上游						
周期中游						
周期下游	0.22**					
大金融						
消费						
成长		0.24***	0.18**		0.23***	0.15*
R2	5.03%	5.58%	3.15%	0.00%	5.40%	2.22%

资料来源：Wind，华泰证券研究所

图表13： 2013年1月~2015年6月六大板块动力图

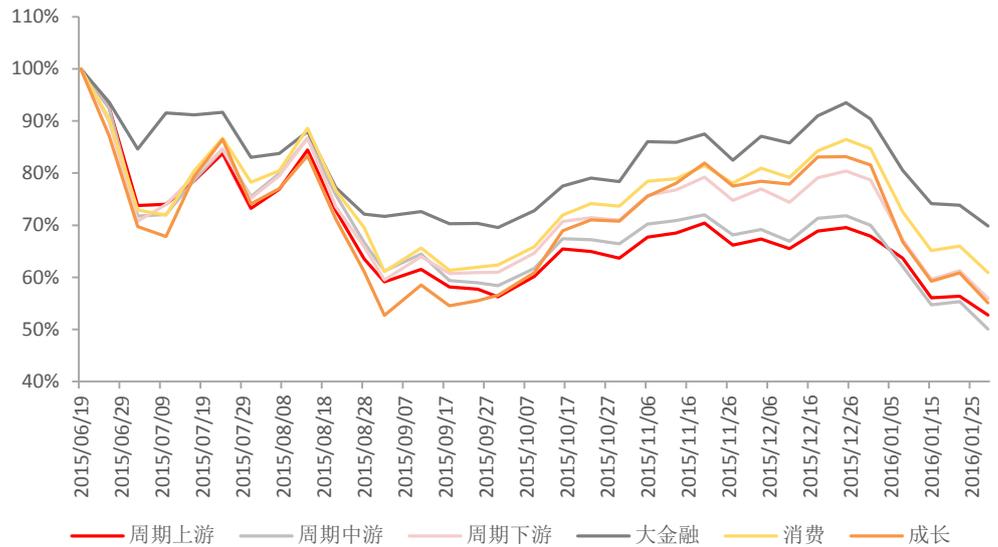


资料来源：Wind，华泰证券研究所

2015年6月~2016年1月

图表14~图表16是我们针对2015年6月~2016年1月这段大跌行情的回归建模结论。彼时，大盘从5178点的高位连续杀跌，恐慌情绪释放过后，大金融板块最先止跌企稳，在整个下跌行情中也最抗跌。从板块动力图也能看出，大金融板块正向影响了其他板块，说明大金融板块跌速放缓后，其他板块也逐渐企稳，所以说大金融板块是这一轮行情的中心节点。

图表14： 2015年6月~2016年1月六大板块净值图



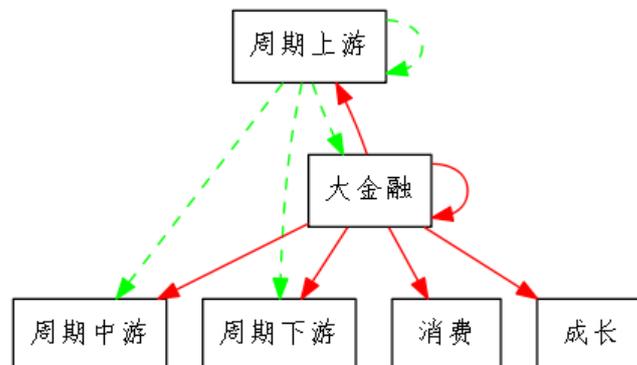
资料来源：Wind，华泰证券研究所

图表15： 2015年6月~2016年1月六大板块回归系数 (*, **, *分别代表双边1%, 5%, 10%置信度水平)**

	周期上游	周期中游	周期下游	大金融	消费	成长
周期上游	-1.17**	-1.11**	-1.21**	-1.45***	-0.91*	-0.86*
周期中游						
周期下游						
大金融	0.68**	0.73**	0.73**	0.7**	0.7**	0.84**
消费						
成长	0.77*	0.67	0.72	0.83*	0.55	0.39
R2	24.97%	24.04%	25.33%	27.42%	23.55%	27.55%

资料来源：Wind，华泰证券研究所

图表16： 2015年6月~2016年1月六大板块动力图



资料来源：Wind，华泰证券研究所

2016年5月~2017年10月

本节我们将针对最近一年半的板块走势进行回归建模，并发表预测观点。

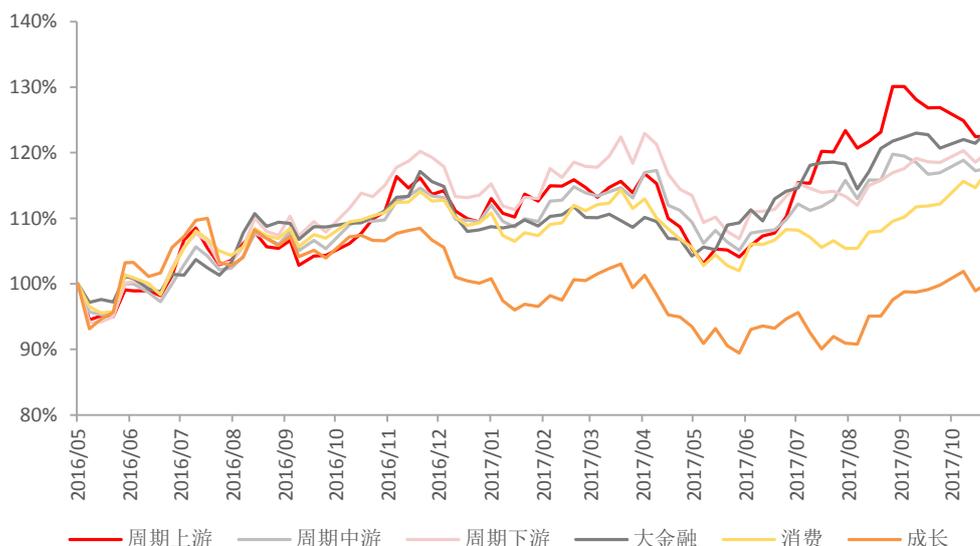
2016年2月底，“华泰周期研究系列”的第一篇报告《市场的轮回》提出：全球股票市场（同比）普遍存在42个月周期，且于2016年初同步见底，建议关注全球股票市场反转。

2016年5月初，“华泰周期研究系列”的第二篇报告《市场的频率》指出：大宗商品长周期上行，PPI、CPI数据同步反转，经济在好转，具有实际盈利增长基础的股票将再次获得资金青睐，周期性板块会迎来新一轮的投资机会，未来价值型周期股将作为主导板块，长期领跑市场。

图表17~图表19是我们针对2016年5月至今这段慢牛行情的回归建模结果，从六大板块净值走势图来看，显然周期板块的涨幅要大于消费和成长板块，是这轮行情的主导板块，这印证了我们之前的预测观点。同时，从板块动力图也能看出，周期上游是本轮行情的动力节点，正向推动了其他板块，这意味着周期上游当期涨得越多，其他板块下期也大概率涨得越多，或者跌得越少，反过来，当周期上游板块开始回调时，其他板块也大概率增速放缓，甚至下跌。

可以注意到，自进入9月以来，市场风格有所切换，作为动力节点的周期上游板块陷入调整。上涨引擎一旦熄火，整个市场向上动力就不足，出于避险的考虑，资金纷纷进入防御板块抱团取暖，所以最近消费、周期中游板块获得显著超额收益，这说明投资者的投资情绪有所降温，市场目前整体处于防守状态，短期做多动能不强，大概率会陷入调整。但我们认为这个短期回调是为未来的上涨积蓄力量，如同出拳前的蓄力，市场长期向上趋势不变。根据周期模型测算，当前市场处于上涨周期的后半程，A股到明年二季度仍然处于上升期。而防守板块显然没法带领市场不断冲高，调整到位后，资金大概率会回流到周期上游、大金融等主导板块，市场才会重拾升势。

图表17： 2016年5月~2017年10月六大板块净值图



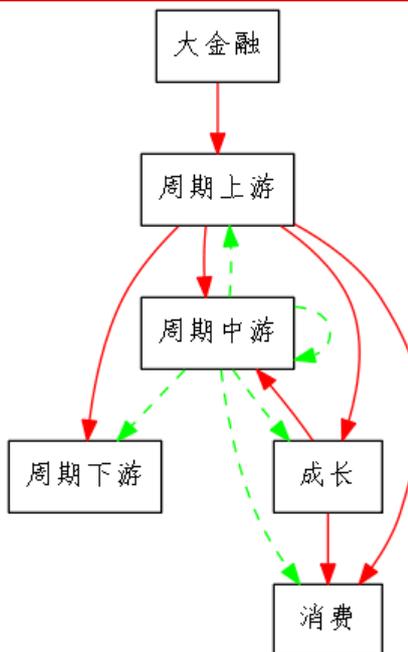
资料来源：Wind，华泰证券研究所

图表18: 2016年5月~2017年10月六大板块回归系数(*, **, *分别代表双边1%, 5%, 10%置信度水平)**

	周期上游	周期中游	周期下游	大金融	消费	成长
周期上游	0.37*	0.61***	0.58***		0.52**	0.64***
周期中游	-0.58***	-0.9***	-0.84***		-0.87***	-0.99***
周期下游						
大金融	0.35**					
消费						
成长		0.4**	0.35*		0.42**	0.39*
R2	14.45%	12.69%	11.20%	0.00%	11.22%	14.96%

资料来源: Wind, 华泰证券研究所

图表19: 2016年5月~2017年10月六大板块动力图



资料来源: Wind, 华泰证券研究所

基于主成分回归的行业轮动策略原理与实证检验

主成分回归模型构建

前文描述中，我们通过对多个典型历史区间进行建模分析，验证了通用回归模型可以挖掘到板块背后的交叉预测关系，而且这种从数据中挖掘出的微观规律大部分情况下和市场的轮动逻辑是相互印证的。那么我们可以基于此来构建行业轮动策略。

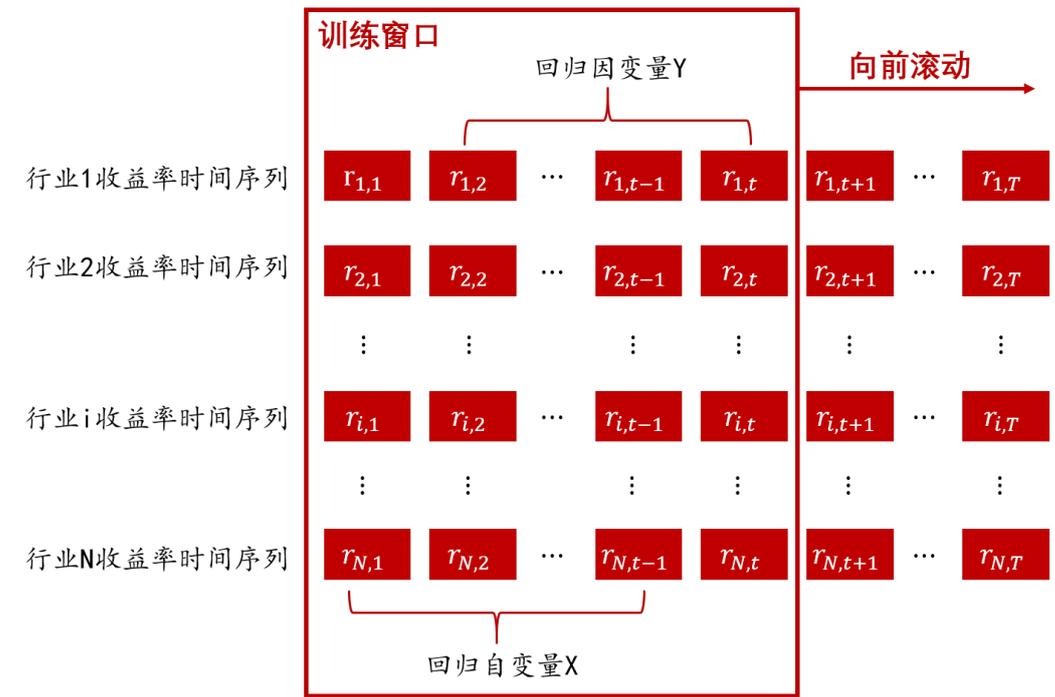
如图表 20 所示，在截面 t 处，选取 $2 \sim t$ 时刻的板块收益率作为因变量 Y ， $1 \sim t-1$ 时刻的板块收益率作为自变量 X ，训练如下回归模型：

$$r_{i,panel+1} = \alpha_i + \sum_{j=1}^N \beta_{i,j} r_{j,panel} + \varepsilon_{i,panel+1} \quad \text{for panel} = 1, 2 \dots t-1$$

也即利用六大板块当期收益率序列对各板块下期收益率序列做回归，构建定价方程，在得到回归系数估计值后，把截面 t 处最新的收益率代入定价方程，就可以得到下一期各板块的收益率预测值，进而指导行业配置。这样不断向前滚动训练窗口，每一期把预测收益率靠前的板块纳入仓位，靠后的板块则剔除仓位，就能构建一个完整的板块轮动策略了。

可以看到，该板块轮动策略的核心假设是：训练窗口内挖掘到的板块间交叉预测关系（也即回归系数的值）在下一期仍然能够保持。而随着训练窗口不断向前滚动，窗口内的轮动规律可能是时变的，回归模型会自适应的提取最显著的规律进行预测，相对比较灵活，不像动量/反转模型，模式是固定的。

图表20： 行业轮动模型示意图



资料来源：华泰证券研究所

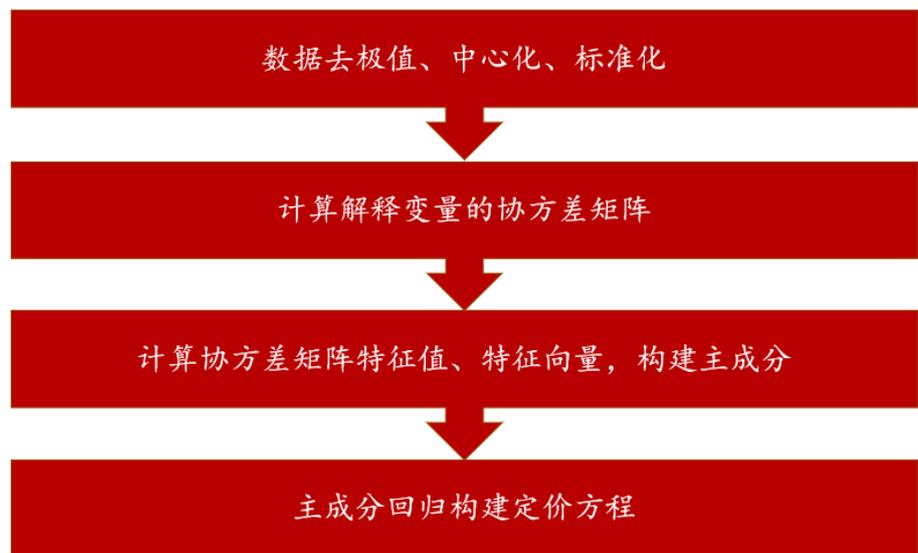
这里需要说明的一点是，我们在前文针对典型区间进行回归建模时，采用的是逐步回归法，这样可以保留对每个板块解释能力最显著的几个因子来构建板块动力图，但在构建板块轮动策略时，逐步回归法就不适用了，因为每个板块的有效解释因子可能是不同的，以图表 9 为例，对周期上游板块，解释能力最强的因子是周期下游和消费板块，而对大金融板块，解释能力最强的是消费板块和成长板块，两者的定价方程参照的基准是不一样的，这样在外延预测时，没法直接比较。所以我们在构建板块轮动策略时，采用的是主成分回归的思路，既能克服共线性的影响，又能保证定价方程参照的基准是一致的。

主成分分析是一种统计方法。它通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量，转换后的这组变量叫主成分。主成分分析在数据降维方面非常有用，利用较少的几个主成分就能保留原始变量大部分的信息，从而在筛选变量的同时又达到信息损失较小的目的。

在行业轮动策略中，主成分回归模型的构建流程如图表 21 所示：

1. 对六大板块原始收益率数据进行去极值、中心化、标准化等预处理操作，记标准化后的收益率矩阵为 X 。
2. 计算收益率矩阵 X 的协方差矩阵 Σ 。
3. 对协方差矩阵 Σ 进行特征值分解，将特征值从大到小排列，记为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_6$ ， e_1, e_2, \dots, e_6 为相应的单位正交特征向量，则新构建的主成分依次为 $Z_1 = Xe_1, Z_2 = Xe_2, \dots, Z_6 = Xe_6$ 。
4. 利用板块收益率对各主成分向量收益率回归，构建定价方程，指导行业配置。

图表21：主成分回归流程



资料来源：华泰证券研究所

在实际问题中，主成分个数 m 的选择比较灵活。常用的一个指标是主成分累计贡献率。前 m 个主成分的累积贡献定义为

$$v_m = \frac{\lambda_1 + \dots + \lambda_m}{p}$$

一般取累计贡献率 v_m 超过 70% 时对应的 m 。

我们以六大板块全样本空间（2005 年 1 月~2017 年 10 月）作为训练空间，得到收益率协方差矩阵 Σ 的特征值如下：

图表22：六大板块收益率协方差矩阵的特征值

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
特征值	5.22	0.45	0.16	0.09	0.05	0.04

资料来源：Wind，华泰证券研究所

最大特征值占比 86.82%，前两大特征值占比 94.3%，从累计贡献率的角度，前两个主成分就已经包含了原始变量绝大部分信息。这里需要强调，大部分主成分应用场景中，需求都是挑选前 m 个主成分，要求 m 尽可能小，保留信息尽可能多。但在我们行业轮动建模场景中，解释因子数日本就不多（6 大板块），计算开销并不大，如果加入更多主成分对解释性能有提升的话，是值得尝试的，后文中我们将进行详细的实证分析。

图表 23 展示了六大板块主成分分析的结果，其中，第 i 行代表第 i 个主成分在六大板块上的暴露，对应特征值 λ_i 的特征向量，也即合成相应主成分向量时的权重。

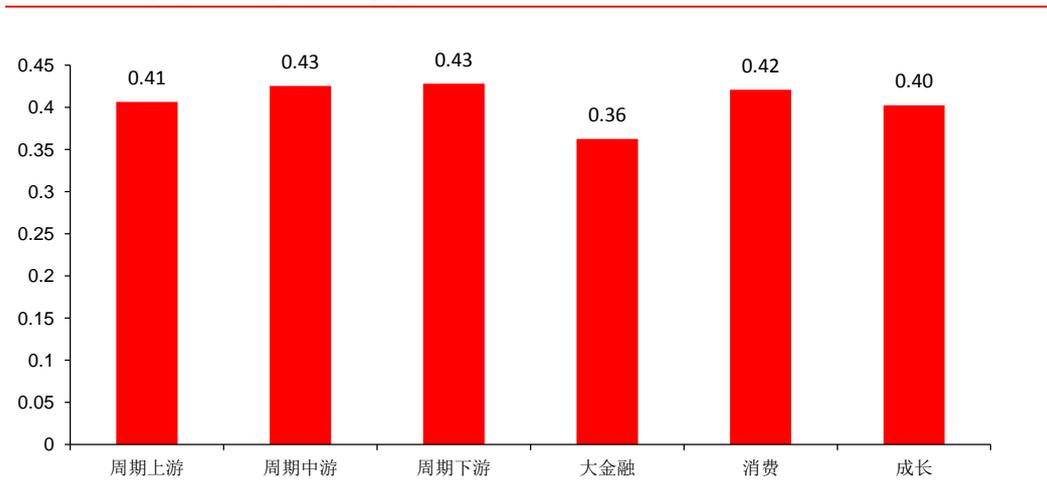
图表 23： 六大板块收益率主成分分析

	周期上游	周期中游	周期下游	大金融	消费	成长
第一主成分	0.41	0.43	0.43	0.36	0.42	0.40
第二主成分	-0.24	0.04	0.10	-0.78	0.30	0.48
第三主成分	-0.78	-0.15	0.02	0.49	0.17	0.31
第四主成分	-0.39	0.58	0.36	-0.13	0.09	-0.59
第五主成分	0.09	-0.64	0.39	-0.02	0.54	-0.36
第六主成分	0.04	0.20	-0.72	0.05	0.63	-0.18

资料来源：Wind，华泰证券研究所

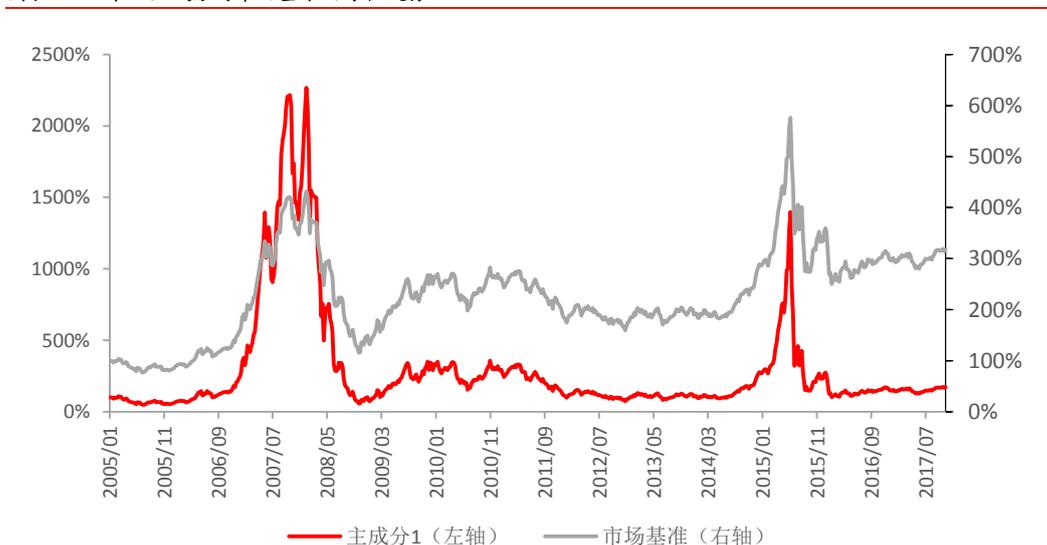
图表 24 显示，第一主成分（也即解释能力最强的主成分）在六大板块上的暴露都是正的，而且数量级基本一致，说明第一主成分其实刻画的就是市场回报率。图表 25 进一步展示了该主成分累计净值曲线和市场基准的比较，其中前者以该主成分在各板块上的暴露为权重合成加权指数，后者直接等权平均合成市场基准，可以看到两者走势基本是一致的。这是合乎逻辑的，因为市场因子是各板块收益率最显著的解释变量。

图表 24： 第一主成分在六大板块上的暴露



资料来源：Wind，华泰证券研究所

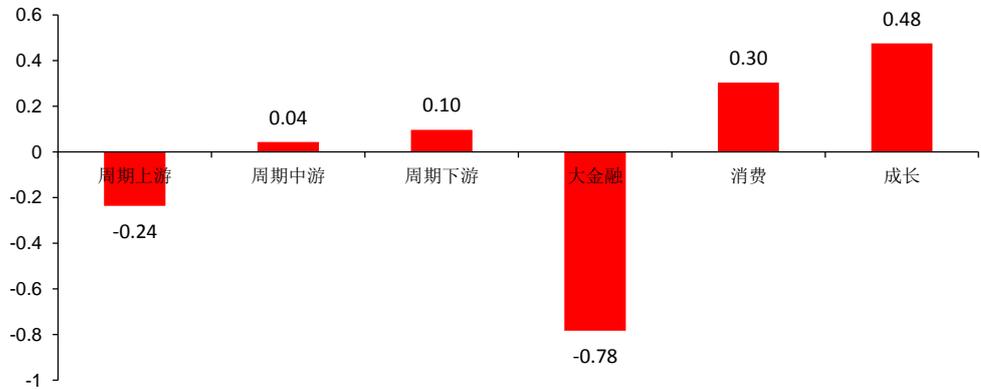
图表 25： 第一主成分与市场基准的净值走势



资料来源：Wind，华泰证券研究所

图表 26 显示，第二主成分在各板块上的暴露有所分化，正向暴露较多的是消费和成长板块，负向暴露较多的是周期上游和大金融板块，其他两个板块上暴露较少，基本可以忽略。在华泰金工周期模型中，一般把周期上、中、下游以及大金融板块归类为周期板块，把消费、成长板块归类为非周期板块。那么可以看到，第二主成分其实是对周期、非周期风格的一个区分。

图表26： 第二主成分在六大板块上的暴露



资料来源：Wind，华泰证券研究所

为了进一步验证主成分 2 的含义，我们将周期上、中、下游以及大金融板块等权平均合成周期指数，将消费、成长板块等权平均合成非周期指数，利用后者减前者构建周期/非周期风格指数。该风格指数与主成分 2 净值曲线对比如图表 27 所示，可以看到，两者走势基本一致，说明主成分 2 就是对周期、非周期风格的区分。另外注意到，在 07 年以前，市场是在五朵金花的推动下不断上涨的，所以周期风格一直占据主导地位，而从 08 年到 15 年，成长板块崛起，非周期风格更占优，进入 2016 年以后，风格再度切换，周期板块再次占据主导地位。

图表27： 第二主成分与周期/非周期风格指数的净值走势



资料来源：Wind，华泰证券研究所

分层回测

分层回测是展示模型预测能力最直观的手段，构建方法如下：

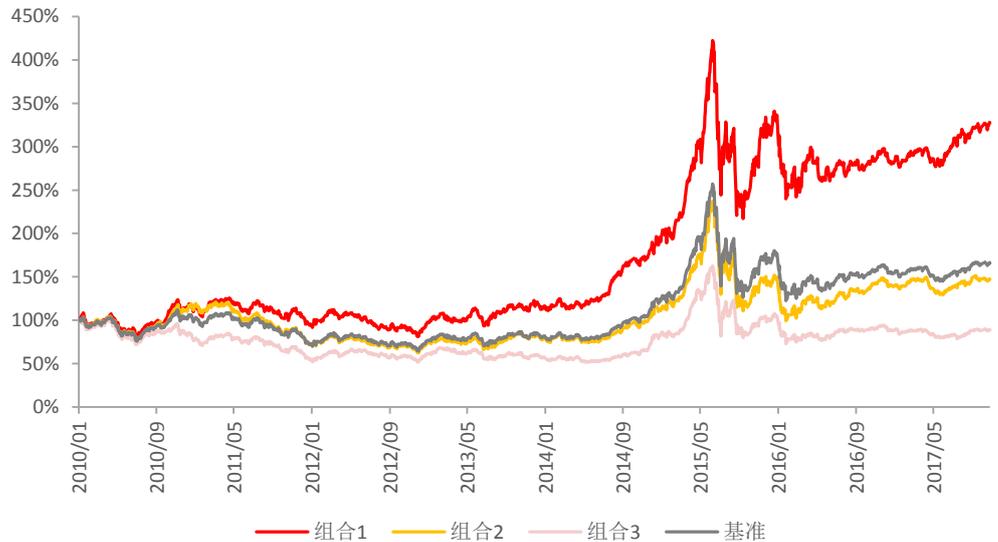
- 可选标的：周期上、中、下游，大金融，消费，成长六大板块。
- 回溯区间：2010年1月4日至2017年10月27日。
- 参数设置：训练窗口长度为242周，考虑全部主成分向量构建回归模型。
- 组合构建：在每周最后一个交易日，基于过去242周六大板块周频对数收益率数据训练主成分回归模型，构建定价方程，并外延预测下一周收益率，按照预测的收益率排名构建分层组合，在下周首个交易日以收盘价换仓。
- 基准组合：等权配置六大板块。
- 评价方法：回测年化收益率、夏普比率、信息比率、最大回撤、胜率等

首先分3层进行回测，组合1持有每期预测排名前两位的板块，组合2、组合3依次类推，得到各分层组合的净值曲线和业绩表现如图表28~图表31所示，可以看到：

- 1、组合1无论从净值曲线还是业绩指标来看都是最优的，年化收益率16.9%，夏普比率0.61，信息比率1.28，胜率接近60%，显著强于基准。
- 2、组合3表现最弱，年化收益率、夏普比率、信息比率均小于0，胜率不到45%，显著弱于基准。
- 3、组合2的表现介于组合1和组合3之间，稍弱于基准。

整体而言，组合1到组合3的表现单调下降，说明该行业轮动模型具备良好的预测能力，能够把握过去一段时间内的板块轮动规律，并获取超额收益。

图表28： 分层回测净值曲线（分3层）



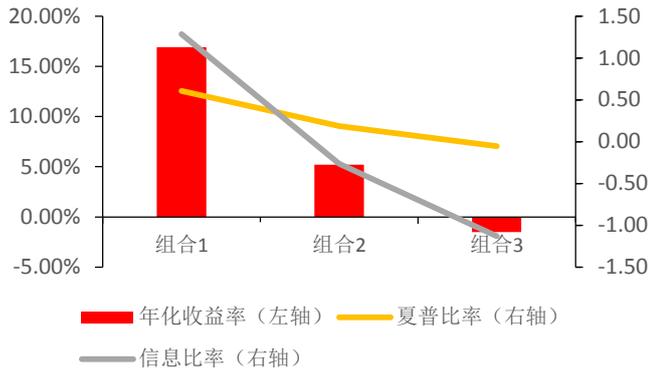
资料来源：Wind，华泰证券研究所

图表29： 分层回测业绩指标（分3层）

	年化收益率	年化波动率	夏普比率	最大回撤	年化超额收益率	超额收益 年化波动率	信息比率	相对基准 月胜率	超额收益 最大回撤
组合1	16.90%	27.91%	0.61	-48.62%	9.25%	7.20%	1.28	59.25%	-9.86%
组合2	5.22%	28.14%	0.19	-58.07%	-1.51%	5.71%	-0.26	46.00%	-31.49%
组合3	-1.49%	28.22%	-0.05	-55.26%	-7.84%	6.92%	-1.13	44.25%	-48.70%
基准	6.90%	27.30%	0.25	-52.42%					

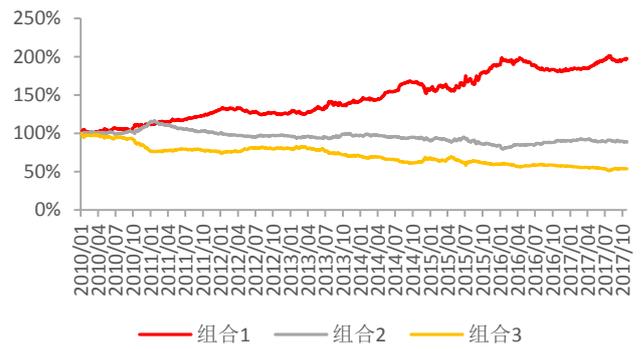
资料来源：Wind，华泰证券研究所

图表30: 分层回测业绩指标对比图示(分3层)



资料来源: Wind, 华泰证券研究所

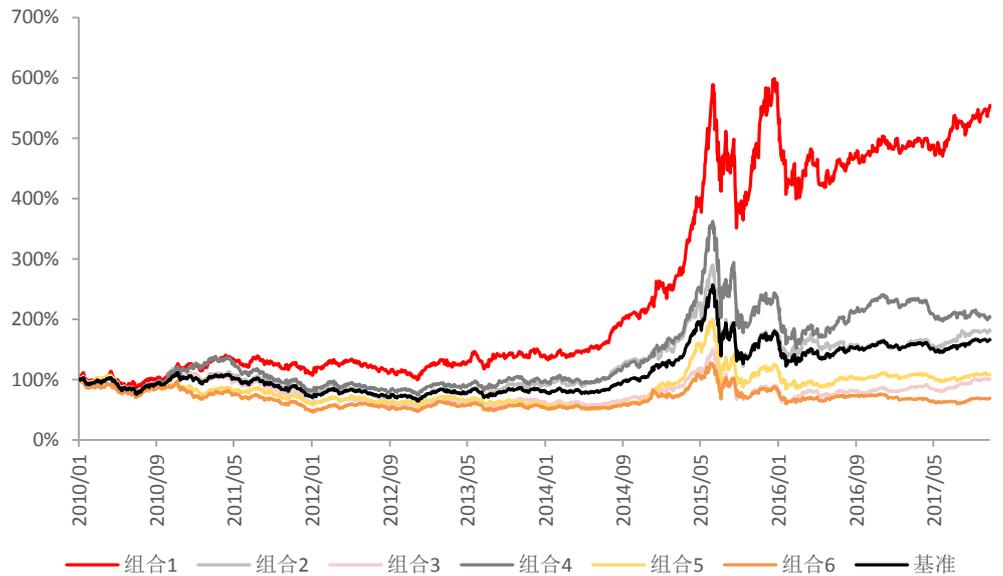
图表31: 分层组合净值除以基准组合净值示意图(分3层)



资料来源: Wind, 华泰证券研究所

进一步, 我们分6层进行回测, 也即组合1~组合6分别持有预测收益排名1~6的板块, 得到其净值曲线和业绩表现如图表32~图表35所示。各组合表现综合排名为: 组合1 > 组合4 > 组合2 > 组合5 > 组合3 > 组合6, 组合1仍是表现最强的, 且显著强于基准, 组合6仍是表现最差的, 且显著弱于基准, 但组合2~组合5的表现并非单调下降。我们猜测原因是在某些截面处, 针对六大板块构建的回归定价方程不够显著, 或者说过去的训练窗口期内并不存在显著的交叉预测关系, 导致在外延预测下一期收益率时, 类似于在随机猜测, 因而各组合表现并不严格单调。

图表32: 分层回测净值曲线(分6层)



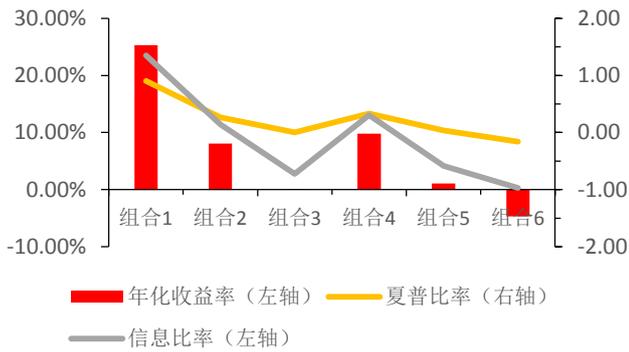
资料来源: Wind, 华泰证券研究所

图表33: 分层回测业绩指标(分6层)

	年化收益率	年化波动率	夏普比率	最大回撤	年化超额收益率	超额收益年化波动率	信息比率	相对基准月胜率	超额收益最大回撤
组合1	25.28%	28.02%	0.90	-40.38%	16.55%	12.27%	1.35	58.75%	-16.54%
组合2	8.08%	30.55%	0.26	-58.00%	1.49%	10.57%	0.14	51.00%	-26.07%
组合3	0.12%	28.57%	0.00	-60.26%	-6.37%	8.79%	-0.72	44.00%	-57.01%
组合4	9.82%	29.52%	0.33	-58.07%	2.91%	9.55%	0.31	49.50%	-19.86%
组合5	1.09%	29.36%	0.04	-57.87%	-5.27%	9.00%	-0.59	43.75%	-42.28%
组合6	-4.68%	29.06%	-0.16	-53.93%	-10.97%	11.31%	-0.97	41.50%	-61.67%
基准	6.90%	27.30%	0.25	-52.42%					

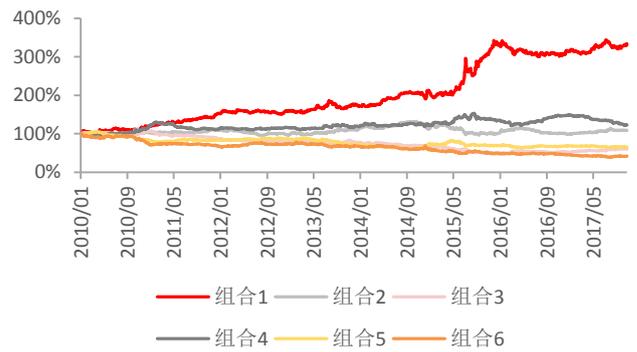
资料来源: Wind, 华泰证券研究所

图表34： 分层回测业绩指标对比图示（分6层）



资料来源：Wind，华泰证券研究所

图表35： 分层组合净值除以基准组合净值示意图（分6层）



资料来源：Wind，华泰证券研究所

对比两种分层场景下的多头组合，分6层场景下组合1的年化收益率、夏普比率、信息比率都更优，但对应的超额收益波动率、超额收益最大回撤也更大，这符合风险收益对等的原则，获取高收益的同时，也需要承担更高的风险。

图表36： 不同分层场景多头策略对比

	年化收益率	年化波动率	夏普比率	最大回撤	年化超额收益率	超额收益年化波动率	信息比率	相对基准月胜率	超额收益最大回撤
组合1(分3层)	16.90%	27.91%	0.61	-48.62%	9.25%	7.20%	1.28	59.25%	-9.86%
组合1(分6层)	25.28%	28.02%	0.90	-40.38%	16.55%	12.27%	1.35	58.75%	-16.54%
基准	6.90%	27.30%	0.25	-52.42%					

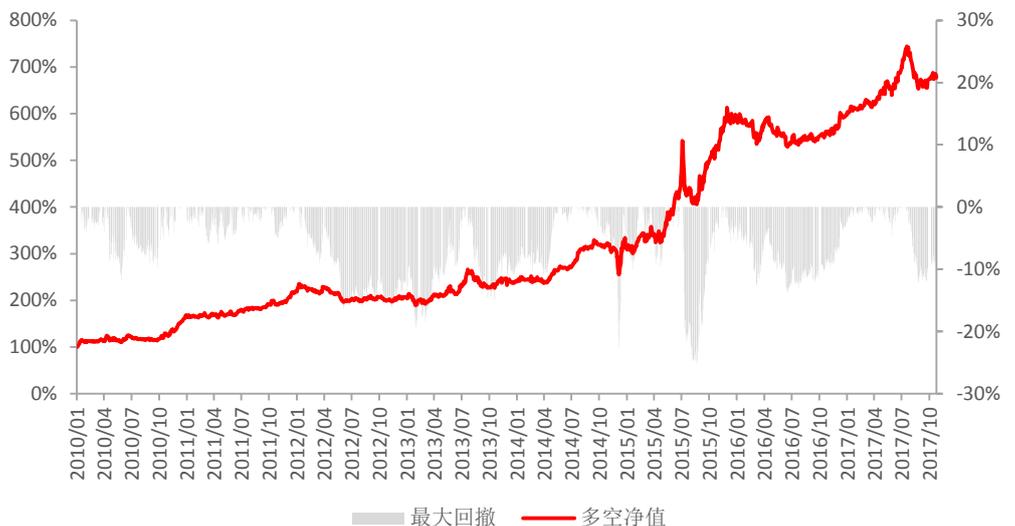
资料来源：Wind，华泰证券研究所

多空回测

假设所有板块指数都可以卖空，则可以构建多空策略组合，也即在每个截面处，买入预测收益率最高的板块，卖空预测收益率最低的板块，以对冲市场风险，获取 alpha 收益。

多空策略回测参数设置参照分层回测场景，其净值走势和业绩表现分别如图表 37，图表 38 所示，年化收益率达到 28.68%，夏普比率 1.51，最大回撤 25.13%，比纯多头策略的表现更优。

图表37： 多空策略净值曲线



资料来源：Wind，华泰证券研究所

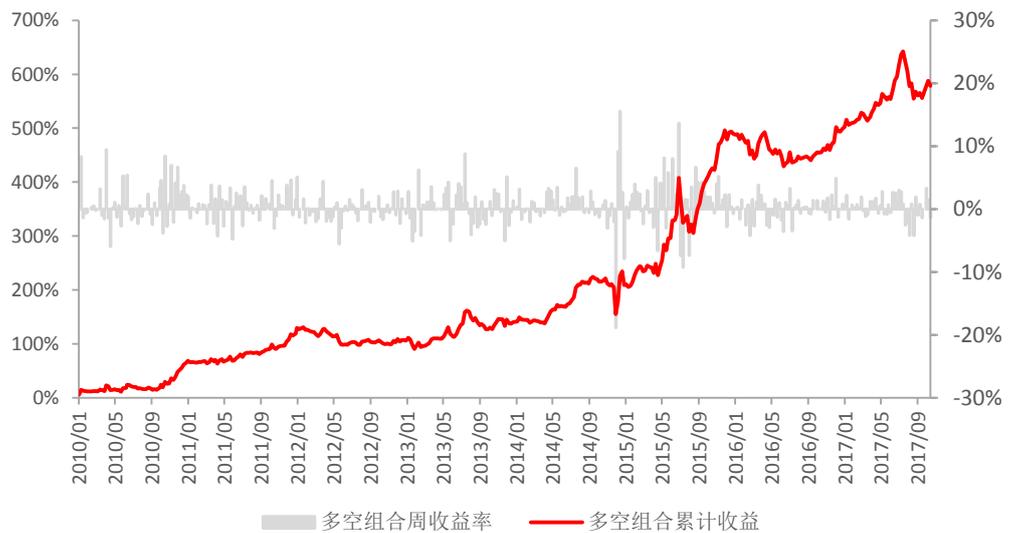
图表38：多空策略业绩表现

	年化收益率	年化波动率	夏普比率	最大回撤
多空策略	28.68%	19.04%	1.51	-25.13%
纯多头策略	25.28%	28.02%	0.90	-40.38%
基准	6.90%	27.30%	0.25	-52.42%

资料来源：Wind，华泰证券研究所

进一步，我们统计每一周多空策略的收益表现如图表 39，图表 40 所示。多空策略周收益率均值约 0.5%，胜率为 59.15%。当预测正确时，策略平均周收益率为 2.15%，预测错误情况下，策略平均周收益率为-1.86%，也即猜准的情况下收益率绝对值更大，策略整体盈亏比为 1.67。其中盈亏比的计算公式如下：

$$\text{盈亏比} = (\text{成功率} \times \text{平均获利幅度}) / (\text{败率} \times \text{平均亏损幅度})$$

图表39：多空策略净周收益率及累计收益率

资料来源：Wind，华泰证券研究所

图表40：多空策略业绩表现

	平均值	最小值	最大值	>0 占比	>0 部分均值	<0 部分均值	盈亏比
周收益率统计	0.51%	-18.85%	15.50%	59.15%	2.15%	-1.86%	1.67

资料来源：Wind，华泰证券研究所

总体而言，基于主成分回归的行业轮动策略是一个追求盈亏比的策略，其胜率优势并不十分显著，毕竟可选资产只有六大板块。但即便如此，该策略也能获得接近 30% 的 alpha 收益，说明板块背后的交叉预测关系是显著且有效的。

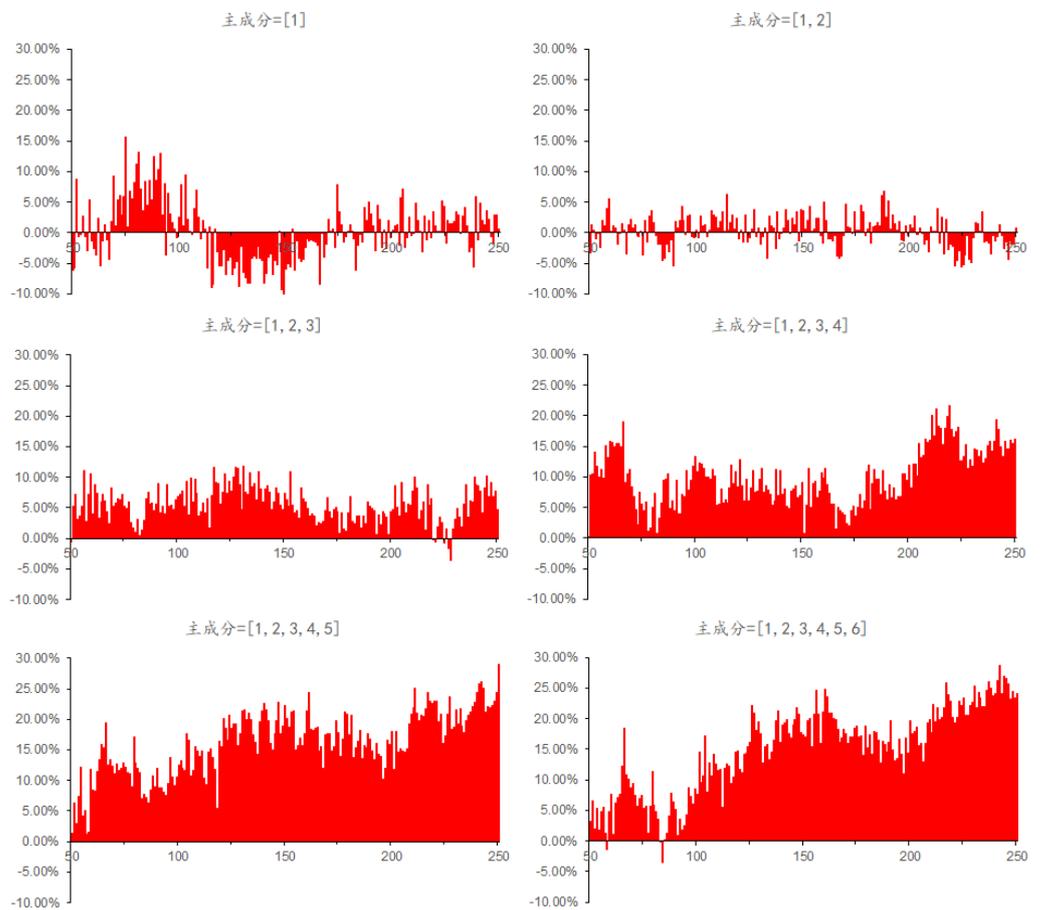
参数敏感性分析

在基于主成分回归模型的行业轮动策略构建中，最核心的两个参数是回归训练窗口长度，以及参与回归的主成分个数。在上文分层回测以及多空回测场景中，我们统一设置了训练窗口长度为242周，且所有6个主成分均参与回归。在本节中，我们将遍历参数的各种可能取值，考察多空策略对参数的敏感程度。具体而言：

- 参与回归的主成分个数从1个到6个，也即全集遍历。
- 训练窗口长度从50周到250周，这里的考虑是：窗口长度若取得太小，学习到的都是短期噪声规律，可能是特定市场环境下的炒作结果，难以外延；窗口长度若取得太大，则测试样本太少，策略没法经过足够长市场轮回的考验。

图表41展示了不同参数设置下的多空策略年化收益率，其中六个子图分别表示纳入1~6个主成分进行回归的结果，每个子图里面，纵坐标对应多空年化收益率，横坐标对应训练窗口长度，为了方便比较，我们将坐标轴的刻度调整到了一致。图表42则展示了每个子图里面的收益率统计值。

图表41：参数遍历场景下的多空收益率图示



资料来源：Wind，华泰证券研究所

图表42：参数遍历场景下的多空收益率统计结果

	最小值	最大值	平均值	中位数	大于0比例
主成分[1]	-10.17%	15.64%	0.21%	0.07%	51.24%
主成分[1,2]	-5.52%	6.76%	0.27%	0.41%	55.72%
主成分[1,2,3]	-3.63%	11.79%	5.50%	5.40%	97.01%
主成分[1,2,3,4]	0.01%	21.59%	9.92%	9.59%	100.00%
主成分[1,2,3,4,5]	1.17%	29.02%	16.07%	16.36%	100.00%
主成分[1,2,3,4,5,6]	-3.57%	28.68%	14.97%	16.20%	98.51%

资料来源：Wind，华泰证券研究所

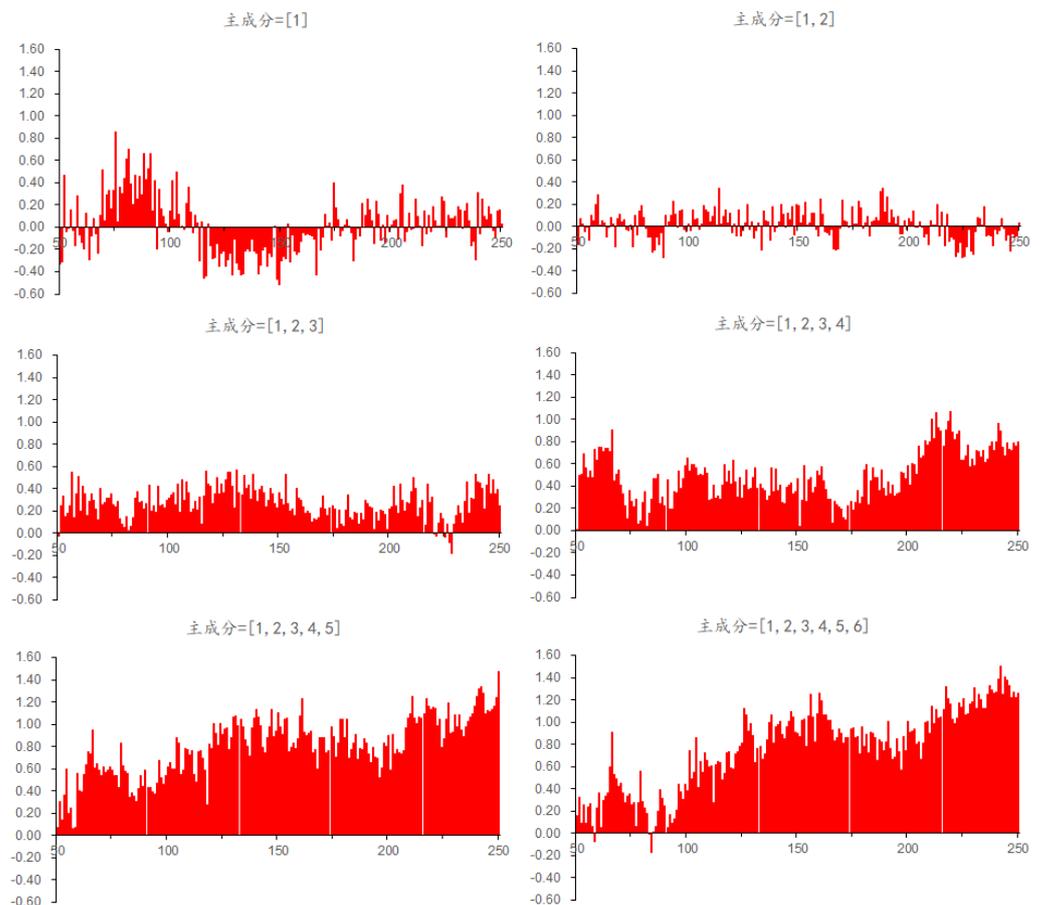
可以看到：

- 1、随着纳入的主成分变多，策略表现也趋向于变好。如果只考虑 1 个或 2 个主成分，不同训练长度下的收益率分布没有太多规律，且只有不到 60% 的场景获得了正收益。而当主成分个数超过 3 个时，几乎所有训练窗口长度下，策略都获得了正收益，而且随着主成分增多，平均收益率也不断抬升。这是合乎逻辑的，因为主成分回归中，不同解释变量之间是相互正交的，所以引入新的解释变量时，至少不会恶化整个回归方程的效果。综合来看，采用前 5 个主成分进行回归的效果是最优的，引入第 6 个主成分后模型表现并没有提升，这说明第 6 个主成分对回归方程边际解释能力的提升有限，甚至带来了一定的噪声影响。
- 2、当主成分回归数目超过 3 个时，模型表现比较稳定，一方面，不同训练窗口长度下，策略的收益率基本都是正的，而且相邻参数之间的结果差距并不大，说明该模型对参数的敏感性在容忍限度内；另一方面，随着训练窗口长度变大，收益率也有变大的趋势，且分布也趋于稳定，说明采用更长的训练窗口，训练出的规律更稳定，外延预测时更可靠。
- 3、年化收益率最高的一组参数是采用前 5 个主成分进行回归，训练窗口长度为 250，年化收益率达到 29.02%。

图表 43、图表 44 是从夏普比率的角度，展示了不同参数设置下的多空策略表现，其结论和前面基本一致：

- 1、参与回归的主成分越多，模型表现越好；训练窗口长度越长，模型表现越稳定。
- 2、夏普比率最高的一组参数是采用所有主成分进行回归，训练窗口长度为 242，夏普比率达到 1.51。这也是前文中进行分层回测和多空回测的参数设置。

图表43： 参数遍历场景下的多空夏普比率图示



资料来源：Wind，华泰证券研究所

图表44: 参数遍历场景下的多空夏普比率统计结果

	最小值	最大值	平均值	中位数	大于0比例
主成分[1]	-0.52	0.85	0.01	0.00	51.24%
主成分[1,2]	-0.28	0.35	0.01	0.02	55.72%
主成分[1,2,3]	-0.18	0.57	0.27	0.26	97.01%
主成分[1,2,3,4]	0.00	1.07	0.49	0.47	100.00%
主成分[1,2,3,4,5]	0.06	1.47	0.80	0.81	100.00%
主成分[1,2,3,4,5,6]	-0.17	1.51	0.76	0.83	98.51%

资料来源: Wind, 华泰证券研究所

基于主成分回归的行业轮动策略业绩归因

在前文中，我们详细阐述了基于主成分回归模型的行业轮动策略原理，并通过分层回测、多空回测等手段实证了其表现，取得了不错的效果。本节我们将构建一个业绩归因模型来解释该策略的收益来源，弄清楚该策略在哪些风险上有暴露并获得了相应补偿。

业绩归因模型

著名的 Fama-French (1993) 三因子模型认为，一个投资组合的超额回报率可由它在三因子上的暴露度来解释，这三个因子分别是：市场资产组合 ($r_m - r_f$)、市值因子 (SMB)、账面市值比因子 (HML)。Carhart (1997) 在 Fama-French (1993) 三因子模型的基础上增加了动量因子并提出了适用性更高的四因子模型，得到了广泛的关注。David E. Rapach (2015) 在 Carhart (1997) 四因子模型的基础上，加上流动性因子，解释能力进一步增强。本节我们将基于该 5 因子归因模型对行业轮动策略的收益率进行回归，测试其在各个因子上的风险暴露，以此来解释策略收益的来源。

五因子业绩归因模型如下（注意由于 A 股实证反转风格更明显，所以我们将原始回归模型中的动量因子替换成了反转因子）：

$$r_{p,t} = \alpha + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{REV}REV_t + \beta_{LIQ}LIQ_t + e_{p,t}$$

其中， $r_{p,t}$ 为策略在截面 t 的收益率， MKT_t 、 SMB_t 、 HML_t 、 REV_t 、 LIQ_t 分别代表市场因子、市值因子、账面市值比因子、反转因子、流动性因子在截面 t 的因子收益率。 α 、 β_{MKT} 、 β_{SMB} 、 β_{HML} 、 β_{REV} 、 β_{LIQ} 均为回归待拟合系数， $e_{p,t}$ 为残差项。此五因子归因模型中的因子收益率，实际上是沿着某一个或几个因子维度，将股票分层，计算下个月每层内部股票的平均收益率，之后用不同层股票的平均收益率做差得到。层间的差异实际上衡量了该种因子的收益效应。通过这个过程，我们完成了因子收益率的计算，且使得各个因子的量纲得到统一，不同因子收益率之间也具有可比性。各因子构建方法如下：

1. 市场因子

- **含义：**市场风险是指大盘走势变化所引起的不确定性。简单来说，就是大盘波动导致个股随之波动的风险。比如表现较好的公司，其股票价格却跟随大盘下降；或者表现不好的公司，股价却跟随牛市上涨。根据风险收益对等的原则，对于市场风险，应该有个对个股收益的超额补偿。
- **构建方法：**市场因子收益率等于本月市场所有股票平均收益率。

2. 市值因子

- **含义：**市值因子的引入源于“小公司效应”这一金融市场异象，即一般而言金融市场上小公司将获取相对于大公司的超额收益。这是因为公司规模对该公司股票的风险有直接影响：资产规模小，风险就会相对增加，反之，风险就会相对减少，所以小公司相对大公司具有更高的风险溢价。
- **构建方法：**市值因子收益率等于小市值股票组合（市值最小的 30% 组合）减小市值股票组合（市值最大的 30% 组合）的月平均收益率。

3. 账面市值比因子

- **含义：**账面市值比是账面的所有者权益除以市值，也即市净率的倒数 BP。账面市值比较高的股票一般为价值股，账面市值比较低的股票一般为成长股。一般而言，价值股将获得比成长股获取更高的收益，其原因如下：由于投资者对企业历史绩效的过度反应，成长股因其历史表现良好而股价被定得过高，而价值股因其历史表现差而股价被定得过低，而过度反应所导致的错误定价最终会被纠正，从而导致价值股具有相对高的收益，而成长股具有相对低的收益。
- **构建方法：**账面市值比因子收益率等于高 BP 股票组合（BP 最大的 30% 组合）减低 BP 股票组合（BP 最小的 30% 组合）的月平均收益率。

4. 反转因子

- **含义：**反转效应是指在一定时期内表现较差的股票在接下来的一段时期内有均值回归的需要，所以表现会较好。
- **构建方法：**反转因子收益率等于最近 1 个月收益率 30% 分位以下组合减去 70% 分位以上组合的月平均收益率。

5. 流动性因子

- **含义：**流动性是衡量证券变现能力的指标，流动性强，说明证券能够在折价较小的情况下，顺利卖出；相反，流动性弱，则表明交易证券需要在成交价格上做出较大让步；因此，流动性差也是风险之一。按照风险与收益相匹配的特征，理论上市场会赋予流动性差的证券一定的收益补偿。
- **构建方法：**参照宋光辉（2017）的处理，我们将成交量和价格的变动联系起来，反映流动性。即对每个月内每日收益率和成交金额的比值求平均。

$$ILLIQ_{iy} = \frac{1}{D_{iy}} \sum_{d=1}^{D_{iy}} \frac{|R_{iyd}|}{VOLD_{iyd}}$$

其中， $ILLIQ_{iy}$ 为第*i*只股票为在第*y*期的流动性指标， D_{iy} 为第*i*只股票在第*y*期的交易天数； R_{iyd} 为第*d*天的股票收益率； $VOLD_{iyd}$ 为*d*天的成交金额(百万元)。 $ILLIQ$ 反映了在单位成交金额下，证券价格波动的大小。 $ILLIQ$ 越小，表明在成交金额相同的情况下，证券价格受到的冲击越小，流动性越好。

$$LIQ_y = High\ ILLIQ_{iy} - Low\ ILLIQ_{iy}$$

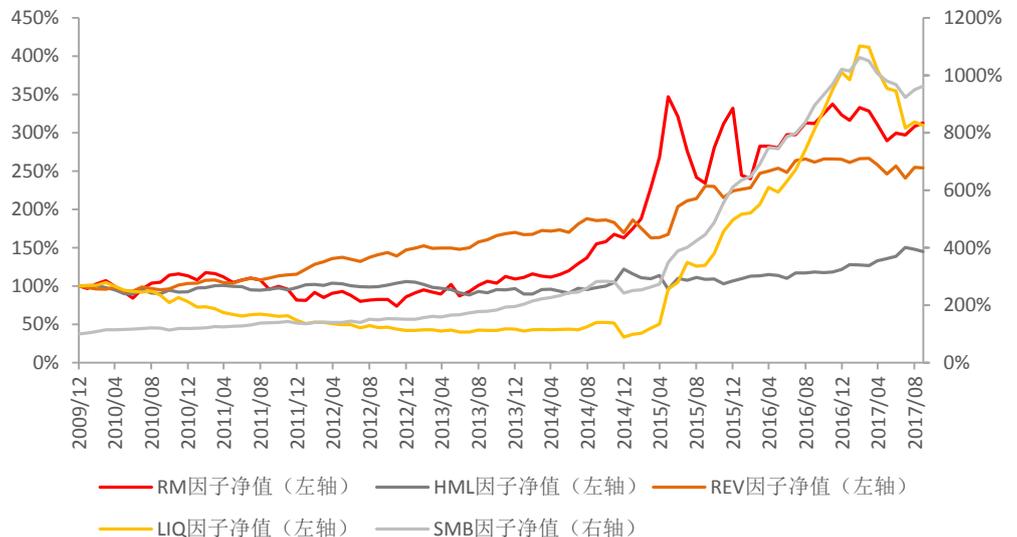
LIQ_y 为第*y*期的流动性因子收益率， $High\ ILLIQ_{iy}$ 为第*y*期的流动性指标位于90%分位以上组合的平均值， $Low\ ILLIQ_{iy}$ 为第*y*期的流动性指标位于10%分位以下组合的平均值。

业绩归因实证分析

前文我们阐述了5因子月度收益率的构建方法，其统计信息如图表46所示，累计净值曲线如图表45所示，可以看到：

- 市场因子（RM）走势基本和上证综指一致，但表现更强，这是因为我们基于所有股票月度收益率求平均来构建市场因子，所以在小盘风格上暴露较多。
- 市值因子（SMB）是月度平均收益率最显著的因子，说明过去绝大部分时间内，小盘股相对大盘股都有稳定的超额收益。从净值曲线看也是一直平稳上升，特别是14年后进入加速上涨阶段，但今年遭遇较大回撤。
- 账面市值比因子（HML）是月度收益率最小的因子，从净值曲线看，16年之前一直处于震荡状态，趋势不明显，16年以后开始逐渐抬头，说明低估值公司相比于高估值公司逐渐获得更多青睐，市场表现为价值投资主导。
- 反转因子（REV）是月度波动率最小的因子，相对来说是很稳定的收益来源，净值曲线也一直处于平稳攀升的走势，在14年底和今年初遭遇了较大的回测。
- 流动性因子（LIQ）月度平均收益率仅次于市值因子，但也可能看到其净值走势和市值因子非常相似，有共线性的可能，后文会做进一步分析。

图表45：5因子累计净值曲线



资料来源：Wind，华泰证券研究所

图表46： 5因子月度收益率统计

	RM	SMB	HML	REV	LIQ
因子月度收益率均值	1.59%	2.56%	0.49%	1.08%	1.81%
因子月度收益率标准差	8.51%	4.60%	4.18%	3.93%	12.06%

资料来源：Wind，华泰证券研究所

从上面的净值走势可以看出，流动性因子（LIQ）和市值因子（SMB）的走势非常相似，可能存在共线性。我们将5个因子做互相关，结果如下表，可以看到，流动性因子与市值因子强正相关，与账面市值比因子有一定负相关性。这个现象也好理解，因为流动性因子衡量了单位成交金额下，证券价格波动的大小，从逻辑上讲，小市值股票的资金容量有限，同等资金冲击下容易造成更大的价格波动，所以小盘股在流动性风险上暴露较多，导致市值因子和流动性因子之间有很强的正相关。

图表47： 5因子之间相关性分析

	RM	SMB	HML	REV	LIQ
RM	1.00	0.39	-0.21	0.07	0.47
SMB	0.39	1.00	-0.45	0.26	0.87
HML	-0.21	-0.45	1.00	0.19	-0.51
REV	0.07	0.26	0.19	1.00	0.18
LIQ	0.47	0.87	-0.51	0.18	1.00

资料来源：Wind，华泰证券研究所

多重共线性的存在会影响回归方程的效果，必须对这些因子进行去共线性处理。具体做法是用流动性因子对其他四个因子做回归，取回归后的残差作为新的流动性因子：

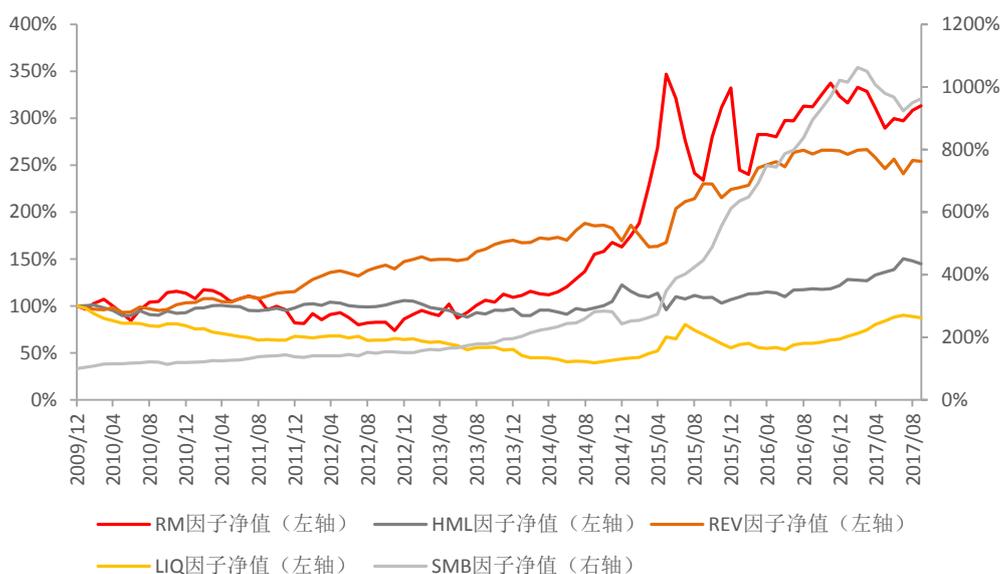
$$LIQ = LIQ - \widehat{LIQ}$$

$$\widehat{LIQ} = \beta_0 + \beta_1 RM + \beta_2 SMB + \beta_3 HML + \beta_4 REV$$

最终构建的业绩归因模型为：

$$r_{p,t} = \alpha + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \beta_{REV} REV_t + \beta_{LIQ} LIQ_t + e_{p,t}$$

去除共线性后，各因子累计净值的表现如下，流动性因子和市值因子间的共线性现象得到明显抑制。

图表48： 去共线性后5因子累计净值曲线


资料来源：Wind，华泰证券研究所

利用该业绩归因模型对行业轮动策略进行分析，结论如图表 49 所示，其中每个单元格展示了相应因子的回归系数和 t 统计量 (*, **, ***分别代表双边 1%, 5%, 10%置信度)；多空对对冲组合和纯多头组合的参数设置与前文一样，训练窗口长度 242 周，全部 6 个主成分都参与回归。

图表49：基于主成分回归模型的行业轮动策略业绩归因结果

投资组合	Annualized alpha	alpha	MKT	SMB	HML	REV	LIQ	R ²
多空对冲组合	17.32%	1.34%**	-0.05	0.29*	-0.02	0.05	0.00	6.45%
		[2.04]	[-0.69]	[1.91]	[-0.13]	[0.33]	[0.00]	
纯多头组合	9.38%	0.75%*	0.88***	-0.19*	0.34***	-0.03	0.04	80.66%
		[1.65]	[18.48]	[-1.79]	[3.20]	[-0.31]	[0.57]	

资料来源：Wind，华泰证券研究所

业绩归因结论如下：

- 1、多空对冲组合在 5%置信度水平下，获得了 17.32%的年化超额收益，且仅在市值因子上有一定的风险暴露，这说明：
 - a) 该策略的收益来源主要是 alpha，而不是通过在特定风险因子上持续暴露敞口来获得收益补偿。或者换个说法，该多空策略可能在不同市场环境下把握了占主导地位的风格因子，而不是持续押注于某一类特定的风格。
 - b) 由于整个归因窗口内，小市值绝大部分时间下都是主导风格，其净值曲线基本处于单调向上的走势，与多空策略的走势非常像，所以在统计意义上，策略在市值因子上有一定暴露。
- 2、纯多头组合在 10%置信度水平下，获得了 9.38%的年化超额收益。从风格上看，该策略在市场因子上暴露最为显著，市场的好坏大概率决定了组合的表现，另一方面，该多头策略在账面市值比上也有一定暴露，说明低估值的板块更容易被选中。整体而言，纯多头策略的表现不如多空策略稳健。

风险提示：通用回归模型是历史经验的总结，存在失效的可能。

附录

CT 统计量推导

考虑经典的资产组合配置问题。在时刻 t 有两个资产，一为无风险资产，收益率为 $R_{f,t+1}$ 。另一为风险资产，收益率为 R_{t+1} ，均值为 $E_t R_{t+1}$ ，方差为 σ_t^2 。注意，收益率的下标 $t+1$ 代表从时刻 t 到时刻 $t+1$ 实现的收益。均值和方差下标为 t 代表投资者当前所处的时刻。假设投资者在风险资产上配置的比例为 θ_t ，则该资产组合的收益率为 $R_{p,t+1}$ 为：

$$R_{p,t+1} = \theta_t R_{t+1} + (1 - \theta_t) R_{f,t+1} = \theta_t (R_{t+1} - R_{f,t+1}) + R_{f,t+1}$$

则组合的平均收益率为

$$E_t R_{p,t+1} = \theta_t (E_t R_{t+1} - R_{f,t+1}) + R_{f,t+1}$$

其中，组合的期望收益率的方差为 $\sigma_{pt}^2 = \theta_t^2 \sigma_t^2$ 。

现考虑线性收益-风险效用函数，即投资者最大化如下的收益和风险的线性组合：

$$\max_{\theta_t} E_t R_{p,t+1} - \frac{k}{2} \sigma_{pt}^2,$$

其中 k 为正数，代表风险偏好水平。根据前面计算的 $E_t R_{p,t+1}$ 和 σ_{pt}^2 ，可知上式等价于：

$$\max_{\theta_t} \theta_t (E_t R_{t+1} - R_{f,t+1}) - \frac{k}{2} \theta_t^2 \sigma_t^2,$$

容易得到，该二次优化问题的解为：

$$\hat{\theta}_t = \frac{E_t R_{t+1} - R_{f,t+1}}{k \sigma_t^2}.$$

也即，当组合中的风险资产的比例为 $\hat{\theta}_t$ 时，可以使组合达到给定效用函数下的最优解。此时，组合的平均超额收益率为

$$E_t R_{t+1} - R_{f,t+1} = \hat{\theta}_t (E_t R_{t+1} - R_{f,t+1}) = \frac{(E_t R_{t+1} - R_{f,t+1})^2}{k \sigma_t^2}.$$

注意到风险资产的夏普比率为：

$$S_t = \frac{E_t R_{t+1} - R_{f,t+1}}{\sigma_t}$$

代入前面的式子得到组合的平均超额收益为：

$$\frac{S_t^2}{k}$$

综上所述，在收益-风险效用函数约束下，最优组合的平均预期超额收益率为风险资产夏普比率的平方与风险偏好水平之比（Campbell & Viceira, 2001）。

接下来利用该结论解释 CT 统计量的意义。假设现有两位不同风格的投资者 A 和 B。投资者 A 在对风险资产的预期超额收益做预测时，仅仅利用历史的平均超额收益，完全忽略所有对超额收益有影响的解释变量。即投资者 A 考虑方程：

$$r_{t+1} = \alpha + \varepsilon_{t+1},$$

其中 r_{t+1} 为风险资产的超额收益， α 为该风险资产的平均超额收益， ε_{t+1} 为风险项，其均值为 0，方差为 σ_ε^2 。于是 r_{t+1} 的预期值为 α ，方差为 σ_ε^2 ，夏普比率为

$$S_1 = \frac{\alpha}{\sqrt{\sigma_\varepsilon^2}}$$

因此，当风险偏好水平为 k 时，最优组合的平均预期超额收益率为

$$\frac{S_1^2}{k} = \frac{1}{k} \left(\frac{\alpha^2}{\sigma_\varepsilon^2} \right)$$

投资者 B 能够在风险项 ε_{t+1} 中剥离出 r_{t+1} 的一部分解释变量 x_t ，相当于考虑如下方程（记为等式 1）：

$$r_{t+1} = \alpha + x_t + \delta_{t+1}$$

其中 x_t 均值为 0，方差为 σ_x^2 。方程中 δ_{t+1} 为随机残差项，与 x_t 不相关，其均值为 0，方差满足 $\sigma_\delta^2 + \sigma_x^2 = \sigma_\varepsilon^2$ 。

此时， r_{t+1} 的期望值为 $\alpha + x_t$ ，方差为 σ_δ^2 ，夏普比率为

$$S_2 = \frac{\alpha + x_t}{\sigma_\delta}$$

因此，当风险偏好水平为 k 时，最优组合的平均预期超额收益率为

$$\frac{S_2^2}{k} = \frac{1}{k} \left(\frac{(\alpha + x_t)^2}{\sigma_\delta^2} \right)$$

在上式中对解释变量 x_t 取平均得到，在平均意义下，最优组合的预期超额收益率如下所示（记为等式2）

$$\frac{S_2^2}{k} = \frac{1}{k} \left(\frac{\alpha^2 + \sigma_x^2}{\sigma_\delta^2} \right)$$

对于一元线性回归模型，回归拟合度 $R^2 = r_{X,Y}^2$ ，其中 $r_{X,Y}$ 为解释变量 X 和因变量 Y 的相关系数。因此等式1的回归拟合度为：

$$R^2 = \left(\frac{\text{cov}(r_{t+1}, x_t)}{\sqrt{\text{var}(r_{t+1})\text{var}(x_t)}} \right)^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\delta^2}$$

于是，等式2可以改写为

$$\frac{S_2^2}{k} = \frac{1}{k} \left(\frac{S_1^2 + R^2}{1 - R^2} \right)$$

综上，相对于完全忽略所有对超额收益有影响的解释变量，当投资者从风险项中剥离出一部分解释变量时，最优组合平均预期超额收益的增加幅度为：

$$\frac{\frac{S_2^2}{k} - \frac{S_1^2}{k}}{\frac{S_1^2}{k}} = \left(\frac{R^2}{1 - R^2} \right) \left(\frac{1 + S_1^2}{S_1^2} \right) = \text{CT}$$

其中 R^2 为回归方程的拟合度， S_1 为风险资产的夏普比率。

由此我们得到，CT代表的意义为：相对于完全忽略所有对超额收益有影响的解释变量，当投资者从风险项中剥离出一部分解释变量时，平均预期超额收益的增加幅度。

免责声明

本报告仅供华泰证券股份有限公司（以下简称“本公司”）客户使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司对该等信息的准确性及完整性不作任何保证。本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，本公司可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。本公司不保证本报告所含信息保持在最新状态。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。

本公司力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成所述证券的买卖出价或征价。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本公司及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，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，也可能为之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。本公司的资产管理部、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。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权力。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

本公司具有中国证监会核准的“证券投资咨询”业务资格，经营许可证编号为：Z23032000。全资子公司华泰金融控股（香港）有限公司具有香港证监会核准的“就证券提供意见”业务资格，经营许可证编号为：A0K809

©版权所有 2017 年华泰证券股份有限公司

评级说明

行业评级体系

一报告发布日后的6个月内的行业涨跌幅相对同期的沪深300指数的涨跌幅为基准；

一投资建议的评级标准

增持行业股票指数超越基准

中性行业股票指数基本与基准持平

减持行业股票指数明显弱于基准

公司评级体系

一报告发布日后的6个月内的公司涨跌幅相对同期的沪深300指数的涨跌幅为基准；

一投资建议的评级标准

买入股价超越基准20%以上

增持股价超越基准5%-20%

中性股价相对基准波动在-5%~5%之间

减持股价弱于基准5%-20%

卖出股价弱于基准20%以上

华泰证券研究

南京

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码：210019

电话：86 25 83389999/传真：86 25 83387521

电子邮件：ht-rd@htsc.com

深圳

深圳市福田区深南大道4011号香港中旅大厦24层/邮政编码：518048

电话：86 755 82493932/传真：86 755 82492062

电子邮件：ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层

邮政编码：100032

电话：86 10 63211166/传真：86 10 63211275

电子邮件：ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码：200120

电话：86 21 28972098/传真：86 21 28972068

电子邮件：ht-rd@htsc.com