

## 金工研究/深度研究

2020年03月03日

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

**李聪** 执业证书编号：S0570519080001  
研究员 01056793938  
licong@htsc.com

**刘志成** 执业证书编号：S0570518080005  
研究员 010-56793940  
liuzhicheng@htsc.com

**王佳星** 010-56793942  
联系人 wangjiaxing@htsc.com

### 相关研究

- 1 《金工：风险预算模型的预算分配方式研究》2020.02
- 2 《金工：基于量价的人工智能选股体系概览》2020.02
- 3 《金工：美元大周期与新兴市场货币危机》2020.02

# 确立研究对象：行业拆分与聚类

## 华泰中观基本面轮动系列之一

### 本文主要开展行业拆分与聚类实证，为后续系列报告确立统一的研究对象

本文作为系列研究首篇，主要目的是为后续系列报告确立统一的研究对象，核心工作是开展行业拆分和聚类实证：1、我们以中信一级行业为基础，分析了每个行业内个股收益和基本面的分化度，并综合考虑各行业的生命周期发展阶段，以及市值占比因素后，决定将食品饮料拆分成酒类、饮料、食品三个子行业；将非银行金融拆分成证券、保险、多元金融三个子行业；2、基于拆分后的行业划分体系，采用蒙特卡洛模拟和最大生成树算法构建行业间关联网络，最终将所有行业聚类成周期、消费、金融、成长、稳定五大风格，以及上游资源、中游材料、中游制造等八大主题板块。

### 行业拆分标准：综合考虑个股分化度以及生命周期阶段、市值占比等因素

好的行业划分体系应该具备“高内聚、低耦合”的特征，也即行业间的分化应尽可能大，而行业内的分化应尽可能小。然而投资实践中，我们发现部分一级行业内个股收益和基本面的分化度较高，需要进行适当拆分。其中，收益分化采用多空累计收益、回归拟合优度、平均相关系数三种方法进行刻画；基本面分化从估值（PE、PB）、盈利（ROE、ROA、ROIC、净利率、毛利率）、营运（资产负债率、总资产周转率、存货周转率）三个维度进行刻画。结合收益和基本面分化度的定量统计结果，以及各生命周期所处阶段、市值占比等因素后，综合判定目标行业是否需要拆分。

### 行业拆分结果：食品饮料和非银行金融内部分化度较大，需要拆分

综合考虑收益和基本面分化度统计结果，以及各生命周期所处阶段、市值占比等因素后，我们决定将食品饮料拆分成酒类、饮料、食品三个子行业；将非银行金融拆分成证券、保险、多元金融三个子行业。进一步，我们基于定量方法评估一个行业划分体系的好坏：1、收益同质性，可以通过检验行业内个股相关系数是否显著高于行业间个股相关系数来判定该分类方案是否把收益趋同性高的股票归到同一个行业；2、基本面同质性，不同行业划分体系的优劣可以通过联合方差检验进行判定。实证结果表明，拆分后的行业划分体系相比于原一级行业体系有显著改善。

### 行业聚类方法：结合蒙特卡洛模拟和最大生成树算法构建行业关联网络

基于自上而下的研究视角，投资者通常会从经济结构、产业链出发将各个细分行业进行归并，便于研究宏观环境对大类板块的影响，以及产业链上下游之间的传导规律。鉴于此，我们结合蒙特卡洛模拟和最大生成树算法构建行业关联网络，进行聚类研究：1、对所有行业的历史收益率序列进行K-means聚类，对于任意两个行业，如果他们被归为一类，则记录为1，否则记为0；2、重复执行多次，计算每两个行业被归为一类的概率，概率越高，说明行业间相似度越高；3、基于最大生成树算法剪枝，把距离较远的边去掉，生成全连接的行业关联网络，也即产业链全景图谱。

### 行业聚类结果：五大风格、八大主题板块

基于行业关联网络可将所有行业聚类成五大风格、八大板块：1、周期风格，下辖上游资源（石油石化、煤炭、有色金属），中游材料（钢铁、基础化工、建材），中游制造（机械、电力设备及新能源、国防军工）三大板块；2、消费风格，下辖可选消费（汽车、家电、酒类），必须消费（食品、饮料、医药、纺织服装、农林牧渔、商贸零售、轻工制造、消费者服务）两大板块；3、金融风格，下辖大金融板块（银行、证券、保险、多元金融、综合金融、房地产）；4、成长风格，下辖TMT板块（计算机、电子、传媒、通信）；5、稳定风格，下辖公共产业板块（交通运输、电力公用事业、建筑）。

风险提示：模型根据历史规律总结，历史规律可能失效。长期来看，行业拆分与聚类结果反映了国民经济体系中的产业链全景图谱，随着经济结构转型升级，未来有变动可能。

## 正文目录

本文研究导读 .....	4
行业拆分实证 .....	5
个股收益分化度刻画 .....	5
基本面分化度刻画 .....	8
行业拆分方案 .....	10
行业拆分方案评估 .....	17
个股收益同质性检验 .....	17
基本面同质性检验 .....	18
行业聚类实证 .....	19
聚类分析方法 .....	19
聚类分析结果 .....	20
总结 .....	23
风险提示 .....	23

## 图表目录

图表 1: 本文研究框架.....	4
图表 2: 收益分化度刻画 .....	5
图表 3: 各行业个股收益分化度排名均值排序（排名越靠前，分化度越大） .....	6
图表 4: 蒙特卡洛模拟各行业个股收益分化度排名均值.....	6
图表 5: 各主题基金数量统计（2020-02-25） .....	7
图表 6: 基本面分化度刻画 .....	8
图表 7: 各行业基本面指标分化度排名 .....	9
图表 8: 行业拆分方案与标准.....	10
图表 9: 初筛行业概览（截至 020-02-25） .....	11
图表 10: 医药子行业价格走势.....	12
图表 11: 医药子行业相关系数.....	12
图表 12: 医药子行业的估值走势 .....	12
图表 13: 医药子行业的净利率走势 .....	12
图表 14: 交通运输子行业价格走势 .....	13
图表 15: 交通运输子行业相关系数 .....	13
图表 16: 交通运输子行业估值（PB_LF）走势 .....	13
图表 17: 交通运输子行业 ROA 走势.....	13
图表 18: 有色金属子行业价格走势 .....	14
图表 19: 有色金属子行业相关系数 .....	14
图表 20: 有色金属子行业毛利率走势.....	14
图表 21: 有色金属子行业 ROE 走势.....	14
图表 22: 食品饮料子行业价格走势 .....	15
图表 23: 食品饮料子行业相关系数 .....	15
图表 24: 食品饮料子行业 ROE 走势.....	15
图表 25: 食品饮料子行业毛利率走势.....	15
图表 26: 非银行金融子行业价格走势.....	16
图表 27: 非银行金融子行业相关系数.....	16
图表 28: 非银行金融子行业 ROE 走势 .....	16
图表 29: 非银行金融子行业净利率走势 .....	16
图表 30: 拆分前后行业内外平均相关收益系数（近 10 年） .....	17
图表 31: 拆分前后行业内外平均相关收益系数（近 5 年） .....	17
图表 32: 拆分前后行业内外平均相关收益系数（近 3 年） .....	17
图表 33: 联合方差检验结果 .....	18
图表 34: 聚类分析方法.....	20
图表 35: 基于最大权值边构建的行业关联网 .....	20
图表 36: 基于最大生成树构建的行业关联网 .....	21
图表 37: 板块聚类结果.....	22
图表 38: 聚类结果图示.....	22

## 本文研究导读

基本面轮动系列报告主要聚焦于中观层面行业轮动研究，本文作为系列研究开篇报告，主要目的是确立研究对象，核心工作是开展行业拆分与聚类实证。

通常而言，投资者习惯采用中信一级行业或申万一级行业作为行业轮动研究对象，无论哪种划分方式，其核心理念都是将上市公司按照某种特性进行归类：从产业研究的角度，同一行业内的公司经营结构、盈利来源类似，其基本面走势应该趋同；从投资实务的角度，同一行业内公司具有比价效应，其股票价格涨跌规律也应该相近。然而在投资实践中，我们发现部分一级行业内个股基本面和价格走势的分化度较大，行业划分不够纯粹，需要进行适当的拆分。

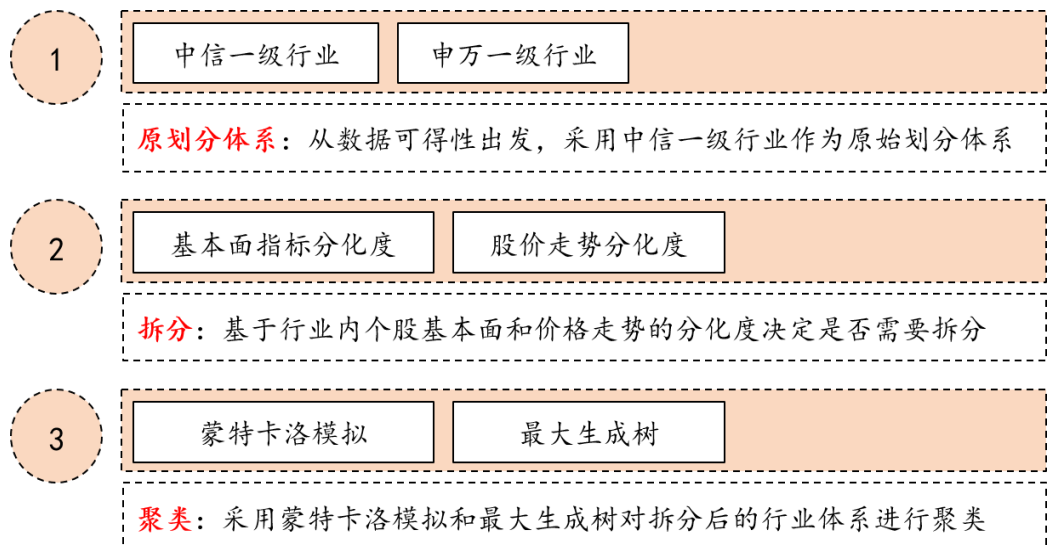
另一方面，基于自上而下的研究视角，投资者通常会从经济结构、产业链出发将各个细分行业进行归并，便于研究宏观环境对大类板块的影响，以及产业链上下游之间的传导规律。鉴于此，我们还需要对拆分后的行业划分体系进行聚类研究。

综上，全文内容主要分为两大块，第一部分聚焦于行业拆分研究，主要解答如下问题：

1. 如何评估行业内个股基本面和个股走势的分化度，进而确定该行业是否需要拆分；
2. 如何在保持灵活性和可跟踪性的基础上，对分化度大的行业进行拆分；
3. 如何定量评估拆分后的行业划分体系和原方案孰优孰劣；

第二部分聚焦于行业聚类研究，我们结合蒙特卡洛模拟和最大生成树算法构建了稳健的聚类方案，最终将拆分后的行业体系聚类成了周期、消费、金融、成长、稳定五大风格，以及上游资源、中游材料、中游制造、可选消费、必须消费、大金融、TMT、公共产业八大主题板块。

图表1： 本文研究框架



资料来源：华泰证券研究所

## 行业拆分实证

无论是中信一级行业还是申万一级行业，其核心理念都是将上市公司按照主营业务收入或利润来源进行归类：从产业研究的角度，同一行业内的公司经营结构、盈利来源类似，其基本面走势应该趋同；从投资实务的角度，同一行业内公司具有比价效应，其股票价格涨跌规律也应该相近。然而在投资实践中，我们发现部分一级行业内个股基本面和价格走势的分化度较大，行业划分不够纯粹，需要进行适当的拆分。

本节内容将以中信一级行业作为研究对象，从价格和基本面两个角度刻画行业内个股的分化度，在兼顾灵活度和可跟踪性的前提下，对分化度大的行业进行适当拆分，而拆分的核心目的是期望各行业内部股票价格和基本面走势尽可能趋同，行业划分更加纯粹。

### 个股收益分化度刻画

本文采用如下三种方式刻画行业内的个股收益分化度

- 多空累计收益**：将行业内可投资个股（剔除 ST 股、停牌股等）按当日收益率从高到低排序分为 5 层，第 1 层为多头组合，第 5 层为空头组合，层内取个股收益率中位数作为组合收益率，计算日内多空收益，并汇总考察区间内的多空累计收益，多空累计收益越高，说明行业内股票价格走势的分化越大。
- 回归拟合优度**：建立个股收益率对行业指数收益率的线性回归模型，计算行业内个股拟合优度  $R^2$  均值， $R^2$  均值越大，说明个股走势和行业指数走势越一致，分化度越小。为了保证统计口径上的一致性，以  $1 - R^2$  均值作为分化度，也即值越大，分化度越大。
- 平均相关系数**：建立行业内个股收益率相关系数矩阵，计算相关系数平均值  $\bar{\rho}$ ， $\bar{\rho}$  越大，说明行业内个股涨跌越一致，分化度越低。同样，为了保证统计口径上的一致性，以  $1 - \bar{\rho}$  作为分化度，值越大，分化度越高。

在实证过程中，为了避免统计结果受到特定考察区间的影响，针对每个指标，我们都进行了 1000 次模拟。每次模拟都会随机设置起止日期（为了避免短期噪音影响，每次考察区间的长度都超过 750 天），然后根据分化度指标计算结果在行业间进行排序。进一步，对 1000 次排序结果取平均，得到最终的行业分化度排序。

图表2：收益分化度刻画

多空 累计 收益	<ul style="list-style-type: none"> <li>将收益率从高到低分5层，第1层为多头，第5层为空头，计算多空收益</li> <li>多空收益越高，说明行业内个股分化度越大</li> </ul>
回归 拟合 优度	<ul style="list-style-type: none"> <li>将行业内个股对行业指数回归，计算回归拟合优度的平均值 <math>R^2</math></li> <li><math>1 - R^2</math> 越大，说明个股与行业走势的一致性越弱，分化度越大</li> </ul>
平均 相关 系数	<ul style="list-style-type: none"> <li>建立行业内个股相关系数矩阵，计算平均相关系数 <math>\bar{\rho}</math></li> <li><math>1 - \bar{\rho}</math> 越大，说明个股间的趋同性越弱，分化度越大</li> </ul>
不同行业之间按照分化度计算结果从高到低排序	
蒙特 卡洛 模拟	<ul style="list-style-type: none"> <li>为了保证统计结果的稳健性，进行蒙特卡洛模拟，随机挑选考察区间</li> <li>每次随机设置起止日期，将1000次模拟的排序结果取平均得到最终排序</li> </ul>

资料来源：华泰证券研究所



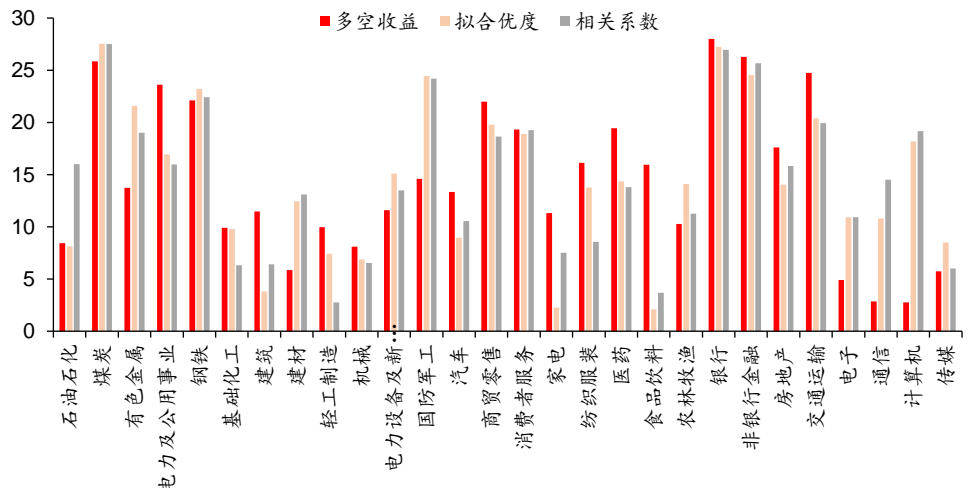
实证结果表明，三种分化度刻画算法在高趋同型行业（分化度排名靠后）上的结论是相似的，主要集中在银行、非银行金融、煤炭、钢铁、交通运输等行业，其中银行、非银行金融和煤炭在三种算法下都是趋同度最高的三个行业，属于典型的金融和强周期风格。而对于分化度较高的行业，三类算法的结果有所区别，其中，多空累计收益视角下分化度排名靠前的主要是计算机、通信、电子、传媒等成长型行业；回归拟合优度和平均相关系数视角下分化度排名靠前的行业主要是食品饮料、轻工制造、家电等消费型行业。

图表3：各行业个股收益分化度排名均值排序（排名越靠前，分化度越大）

排名	多空累计收益分化度平均排名	拟合优度分化度平均排名	相关系数分化度平均排名
1	计算机 2.8	食品饮料 2.1	轻工制造 2.8
2	通信 2.9	家电 2.3	食品饮料 3.7
3	电子 4.9	建筑 3.8	传媒 6.0
4	传媒 5.7	机械 6.9	基础化工 6.3
5	建材 5.9	轻工制造 7.4	建筑 6.4
6	机械 8.1	石油石化 8.1	机械 6.5
7	石油石化 8.4	传媒 8.5	家电 7.5
8	基础化工 9.9	汽车 9.0	纺织服装 8.6
9	轻工制造 10.0	基础化工 9.8	汽车 10.5
10	农林牧渔 10.3	通信 10.8	电子 10.9
11	家电 11.3	电子 10.9	农林牧渔 11.3
12	建筑 11.5	建材 12.4	建材 13.1
13	电力设备及新能源 11.6	纺织服装 13.8	电力设备及新能源 13.5
14	汽车 13.3	房地产 14.0	医药 13.8
15	有色金属 13.7	农林牧渔 14.1	通信 14.5
16	国防军工 14.6	医药 14.3	房地产 15.8
17	食品饮料 15.9	电力设备及新能源 15.1	电力及公用事业 16.0
18	纺织服装 16.1	电力及公用事业 16.9	石油石化 16.0
19	房地产 17.6	计算机 18.2	商贸零售 18.7
20	消费者服务 19.3	消费者服务 18.9	有色金属 19.0
21	医药 19.4	商贸零售 19.8	计算机 19.2
22	商贸零售 22.0	交通运输 20.4	消费者服务 19.2
23	钢铁 22.1	有色金属 21.6	交通运输 19.9
24	电力及公用事业 23.6	钢铁 23.2	钢铁 22.4
25	交通运输 24.8	国防军工 24.4	国防军工 24.2
26	煤炭 25.9	非银行金融 24.5	非银行金融 25.7
27	非银行金融 26.3	银行 27.2	银行 27.0
28	银行 28.0	煤炭 27.5	煤炭 27.5

资料来源：Wind，华泰证券研究所

图表4：蒙特卡洛模拟各行业个股收益分化度排名均值



资料来源：Wind，华泰证券研究所

各行业个股收益的分化可以从主题基金的分布中得到印证。我们按照 Wind 开放式基金分类，统计了不同类别下消费、医药、TMT、周期、金融地产五类主题基金的个数，其中，主题划分主要根据基金名称中的关键字匹配确定，具体而言：

1. 消费类关键词：消费、食品、白酒、饮料。
2. 医药类关键词：医药、医疗、健康、卫生。
3. TMT 类关键词：TMT、电子、传媒、计算机、通信、互联网、信息、大数据。
4. 周期类关键词：周期、资源、原材料、钢铁、煤炭、有色、石油、化工、建材。
5. 金融地产类关键词：金融、地产、银行、非银、证券、保险。

统计结果显示，主题类基金在所有基金中的占比偏低，绝大多数基金仍然以沪深 300 等宽基指数作为业绩比较基准，选股范围并非限定在某一特定的行业板块。但主题基金数量结构上的分化仍然反映出一些有意思的结论：

1. 在普通股票型、增强指数型、偏股混合型等偏向于主动管理的基金中，消费、医药、TMT 类基金的数量要显著多于周期和金融地产类，这一方面是因为前三者符合经济结构转型的需要，更加受到投资者青睐；另一方面是因为这些行业内个股收益分化度较大，可以充分发挥基金经理的  $\alpha$  获取能力。
2. 在被动指数型基金中，周期和金融地产类主题基金的数量大幅提升，这两类行业内个股收益趋同度较高，通过主动管理获取  $\alpha$  较为困难，所以主要作为  $\beta$  收益获取渠道（被动指数型基金具有成本优势）。

综上，主题类基金的数量分布和前文收益分化度的统计结果是可以相互印证的，这也为 FOF 投资带来一个具有实操价值的结论：对于消费、医药、TMT 类收益分化度比较大的行业，应该侧重于基金优选来获取更高的相对收益，适合长期持有；对于周期、金融地产类收益趋同度较高的行业，应该侧重于大势研判获取绝对收益，适合波段操作。

图表5：各主题基金数量统计（2020-02-25）

	总数	消费类	医药类	TMT 类	周期类	金融地产
普通股票型基金	422	30	33	25	9	5
增强指数型基金	169	1	2	0	0	0
被动指数型基金	884	23	37	58	30	86
偏股混合型基金	1066	42	35	13	13	5
灵活配置型基金	1917	27	35	28	9	15
汇总	4458	123	142	124	61	111

资料来源：Wind，华泰证券研究所

## 基本面分化度刻画

本文主要从财务指标维度衡量个股的基本面分化：

1. 估值维度，包含市盈率  $PE\_TTM$  和市净率  $PB\_LF$ 。估值水平是股票定价的基础，同一行业内不同公司的经营结构和盈利来源类似，估值应该具有可比性，否则套利资金的存在会抹平估值溢价。
2. 盈利维度，包含净资产收益率  $ROE$ 、总资产收益率  $ROA$ 、资本回报率  $ROIC$ 、净利率、毛利率。同一行业内不同公司的景气驱动因素类似，理论上应该是一荣俱荣、一损俱损，分化过大意味着盈利结构差别较大。
3. 营运维度，包含资产负债率、总资产周转率、存货周转率，主要考察同一行业内不同公司的资本结构、偿债能力等营运状况的差异。

上述指标计算过程中，为了剔除季节性影响，凡是流量类数据（也即利润表、现金流量表中的数据）均采用 TTM 方式构建。此外，资本回报率（ $ROIC$ ）定义如下：

$$ROIC = \text{息前税后利润} / \text{期初期末全部投入资本均值}$$

该指标通常可用于直观评估某公司的价值创造能力，其中息前税后利润为：

$$\text{息前税后利润} = \text{息税前利润} - \text{税收}$$

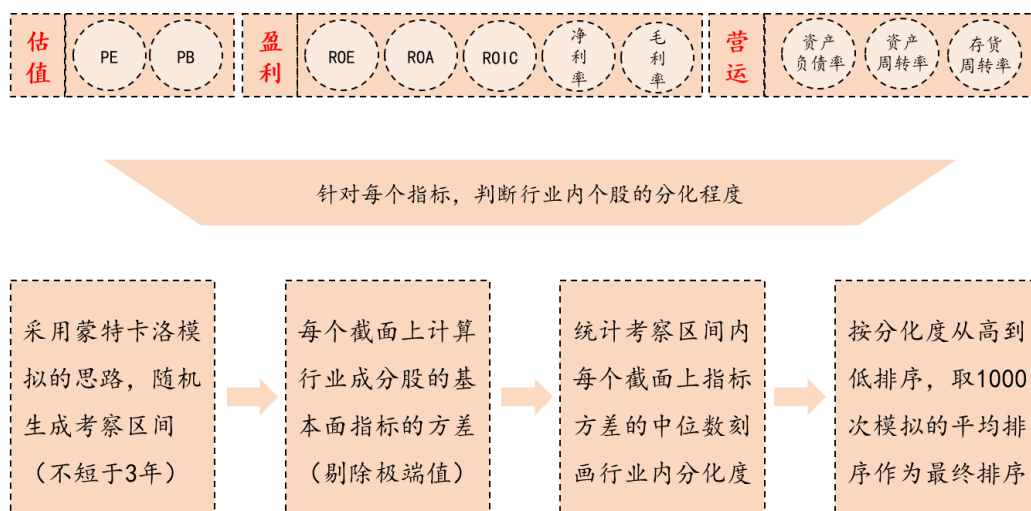
全部投入资本按如下方式计算：

$$\begin{aligned} \text{全部投入资本} = & \text{股东权益} + \text{短期借款} + \text{应付票据} + \text{应付利息} \\ & + \text{一年内到期的非流动负债} + \text{长期借款} + \text{应付债券} \end{aligned}$$

在实证过程中，针对每个指标，进行如下处理：

1. 蒙特卡洛模拟生成随机考察区间，为了避免结论受到短期噪声冲击的影响，考察区间的长度至少为 3 年，也即 12 个季度。
2. 针对考察区间内每个财报季，计算个股的财务指标，统计行业内成分股的方差来刻画行业分化度，这里为了剔除极端值的影响，只保留成分股中财务指标值处于[5%,95%]分位数之间的样本。
3. 对考察区间内每个截面的方差求中位数，作为该行业在考察区间内的分化度衡量，不同行业间按照分化度从大到小排序。
4. 以上步骤重复 1000 次，取 1000 次排序结果的平均值作为最终的行业分化度排序。

图表6： 基本面分化度刻画



资料来源：华泰证券研究所



实证结果表明，不同基本面指标给出的分化度结果排序并不一致：

1. PB 维度分化度最大的三个行业是医药、有色金属、传媒；
2. PE 维度分化度最大的三个行业是国防军工、有色金属、农林牧渔；
3. ROA 维度分化度最大的三个行业是食品饮料、医药、基础化工；
4. ROE 维度分化度最大的三个行业是食品饮料、钢铁、房地产；
5. ROIC 维度分化度最大的三个行业是非银行金融、食品饮料、医药；
6. 净利率维度分化度最大的三个行业是非银行金融、交通运输、房地产；
7. 毛利率维度分化度最大的三个行业是医药、交通运输、食品饮料；
8. 资产负债率维度分化度最大的三个行业是石油石化、基础化工、有色金属；
9. 总资产周转率维度分化度最大的三个行业是石油石化、商贸零售、有色金属；
10. 存货周转率维度分化度最大的三个行业是交通运输、传媒、消费者服务；

如果以至少三项指标排名前三作为高分化度标准，那么基本面分化度最为显著的行业包括有色金属、食品饮料、医药、交通运输。对于非银行金融行业，由于其财务数据中一般没有营业成本项，所以无法计算毛利率和存货周转率指标，而剩余八项指标中，ROIC 和净利率分化度均排名第一，因此将非银行金融行业也列入基本面分化度较高的行业中。

图表7：各行业基本面指标分化度排名

	PB	PE	ROA	ROE	ROIC	净利率	毛利率	资产负债率	总资产周转率	存货周转率
石油石化	20	5	19	16	22	21	16	1	1	7
煤炭	27	20	8	4	10	14	10	26	13	8
有色金属	2	2	9	6	16	15	14	3	3	12
电力及公用事业	19	24	22	22	28	6	12	19	25	4
钢铁	25	18	16	2	21	23	25	9	4	16
基础化工	18	17	3	8	12	17	18	2	18	13
建筑	17	25	26	24	20	26	26	20	16	10
建材	12	6	17	11	19	16	20	4	22	18
轻工制造	9	8	5	10	17	19	21	10	23	20
机械	15	9	18	13	15	12	17	12	24	25
电力设备及新能源	13	16	20	23	25	18	19	13	20	23
国防军工	11	1	25	27	27	25	15	14	21	24
汽车	6	10	13	9	6	22	24	11	9	14
商贸零售	23	15	23	17	18	28	22	23	2	5
消费者服务	22	11	10	20	8	7	6	22	8	3
家电	21	23	11	5	4	24	23	7	7	21
纺织服装	16	4	4	14	7	20	8	15	15	22
医药	1	22	2	12	3	5	1	5	10	19
食品饮料	8	14	1	1	2	9	3	25	6	11
农林牧渔	14	3	6	7	5	10	11	16	5	15
银行	28	28	28	28	23	27	--	28	28	--
非银行金融	24	27	27	26	1	1	--	27	27	--
房地产	4	7	24	3	24	3	9	18	26	26
交通运输	26	26	21	25	26	2	2	6	11	1
电子	10	12	12	15	9	13	13	17	14	17
通信	5	13	14	18	14	11	7	8	17	9
计算机	7	19	7	21	13	8	5	24	19	6
传媒	3	21	15	19	11	4	4	21	12	2

注：银行和银行金融行业的财务数据中一般没有营业成本一项，所以毛利率和存货周转率无法计算

资料来源：Wind，华泰证券研究所

整体来看，在收益分化度中比较突出的成长类行业（计算机、电子、传媒、通信），其基本面分化度并不突出（除了传媒行业在个别指标上分化度较高）；与之对应的，在收益分化度中并不突出的医药行业，基本面分化度却非常明显（PB、ROA、ROIC、毛利四项指标中分化度均排名前三）；而食品饮料行业则是在收益分化度和基本面分化度中均排名靠前，是重点考察对象。

## 行业拆分方案

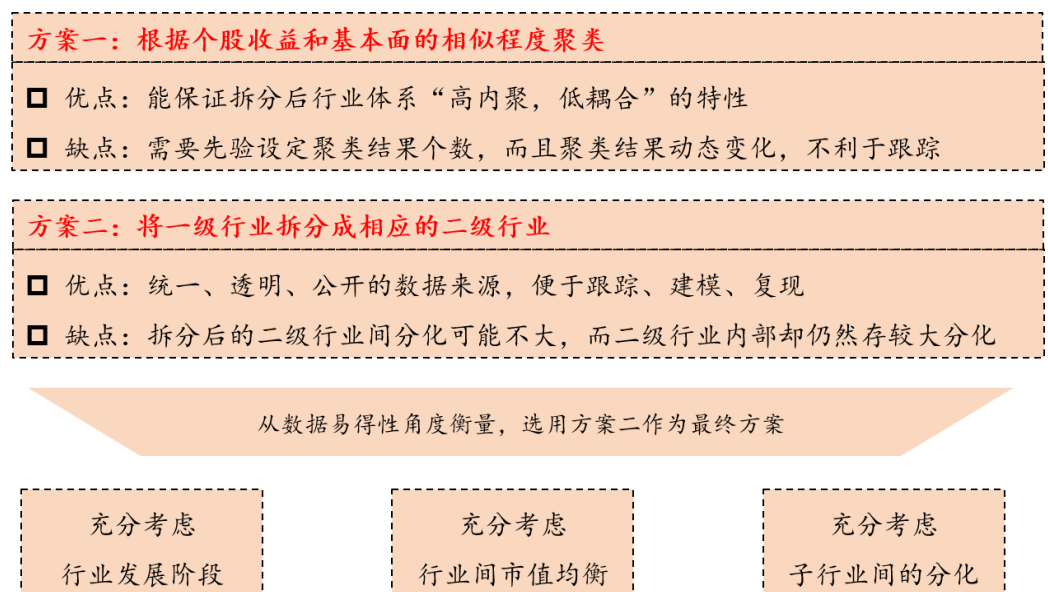
总结前文，我们从收益和基本面出发，对各行业内个股的分化情况进行了定量梳理。从收益分化角度，初步筛选**计算机、通信、电子、食品饮料、家电、建筑、轻工制造**作为研究对象；从基本面分化角度，初步筛选**有色金属、食品饮料、医药、非银行金融、交通运输、传媒**作为研究对象。本节内容主要讨论如何在保持灵活性和可跟踪性的基础上，对部分行业进行拆分。拆分的核心目的是期望新的行业划分体系具有“高内聚、低耦合”的特征，也即行业之间的分化度应该尽可能大，而行业内的分化度应该尽可能小。

行业拆分一般有两种思路：1、直接提取一级行业成分股，按照个股收益和基本面的相似程度进行聚类，该方法的好处是能保证拆分后行业的纯粹性，缺点是聚类结果动态变化，不利于跟踪，而且需要先验指定聚类个数，主观性较强；2、直接将一级行业拆分成二级行业，该方法的好处是数据来源透明，易于跟踪，缺点是难以保证拆分后行业体系的“高内聚、低耦合”特性，也即拆分出来的二级行业之间可能分化并不大，但二级行业内部却仍然存在较大的分化。

从定量模型构建的角度，数据来源的公开、透明、一致性应该是第一考量要素，因此本文采用了方案二的思路，也即将分化度较大的一级行业直接拆分成相应的二级行业。展开具体讨论之前，还需要明确一下拆分标准：

- 充分考虑行业当前发展阶段：**根据产业生命周期理论，行业的生命发展周期包括幼稚期、成长期、成熟期、衰退期。不同阶段下，行业内部的竞争格局、增长潜力相差较大，分析时不能一概而论。一般而言，处于成长期的行业，市场需求快速增长，竞争格局尚不稳定，内部多元性和不确定性较强，根据当前状态做的行业划分很难适应未来一段时间的变化；对应地，处于成熟期、衰退期的行业，经过市场的优胜劣汰之后，内部竞争格局已经相对稳定，未来很难出现大的变化，拆分时应该重点考虑这些行业。
- 充分考虑行业间的市值均衡：**从量化研究来看，行业划分体系是市场中性策略中控制偏离的标准，因此各行业之间的市值不应差距过大；从投资实务的角度，如果拆分后的行业总市值过低、股票数目过少、流动性不足，也没有跟踪的必要。
- 充分考虑子行业间的分化程度：**一般而言，行业内个股高分化是子行业高分化的必要条件，但非充分条件。因此，为了保证拆分结果的“高内聚、低耦合”特性，还需要检查子行业间的分化程度。

图表8：行业拆分方案与标准



资料来源：华泰证券研究所

图表9：初筛行业概览（截至 020-02-25）

一级行业	总市值(亿)	自由流通市值(亿)	成份股个数	二级行业	总市值(亿)	自由流通市值(亿)	成份股个数
有色金属	14454	6936	103	贵金属	3337	1651	12
				工业金属	5195	2429	52
				稀有金属	5921	2855	39
建筑	13791	5734	126	建筑施工	12076	4937	83
				建筑装修Ⅱ	1154	490	28
				建筑设计及服务Ⅱ	561	306	15
轻工制造	8057	3268	116	造纸Ⅱ	1572	770	22
				包装印刷	2773	1158	38
				家居	2459	894	36
				文娱轻工Ⅱ	974	345	10
				其他轻工Ⅱ	279	102	10
家电	14487	7950	73	白色家电Ⅱ	8921	5922	10
				黑色家电Ⅱ	1282	561	9
				小家电Ⅱ	1729	439	18
				照明电工及其他	1891	731	30
				厨房电器Ⅱ	663	297	6
医药	51093	22753	310	化学制药	15492	7533	97
				中药生产	8268	3790	73
				生物医药Ⅱ	8371	3993	42
				其他医药医疗	18962	7437	98
食品饮料	38103	15455	102	酒类	26754	10507	38
				饮料	2925	2052	16
				食品	8425	2895	48
非银行金融	60679	23744	65	证券Ⅱ	27940	12982	45
				保险Ⅱ	29541	9885	7
				多元金融	3198	877	13
交通运输	25104	6777	123	公路铁路	7973	1486	38
				物流	5452	1784	37
				航运港口	6062	1534	35
				航空机场	5617	1974	13
电子	55132	23950	241	半导体	16025	6164	48
				元器件	8557	3767	39
				光学光电	14628	6341	54
				消费电子	13901	6769	72
				其他电子零部件Ⅱ	2021	909	28
通信	21185	8274	114	电信运营Ⅱ	1918	552	3
				通信设备制造	17133	6508	85
				增值服务Ⅱ	1366	793	11
				通讯工程服务	769	421	15
计算机	36869	17779	235	计算机设备	9492	4708	69
				计算机软件	18523	9580	142
				云服务	4303	1869	18
				产业互联网	4551	1622	6
传媒	16378	7293	143	媒体	3751	1693	39
				广告营销	2881	1587	31
				文化娱乐	7344	3164	58
				互联网媒体	2401	849	15

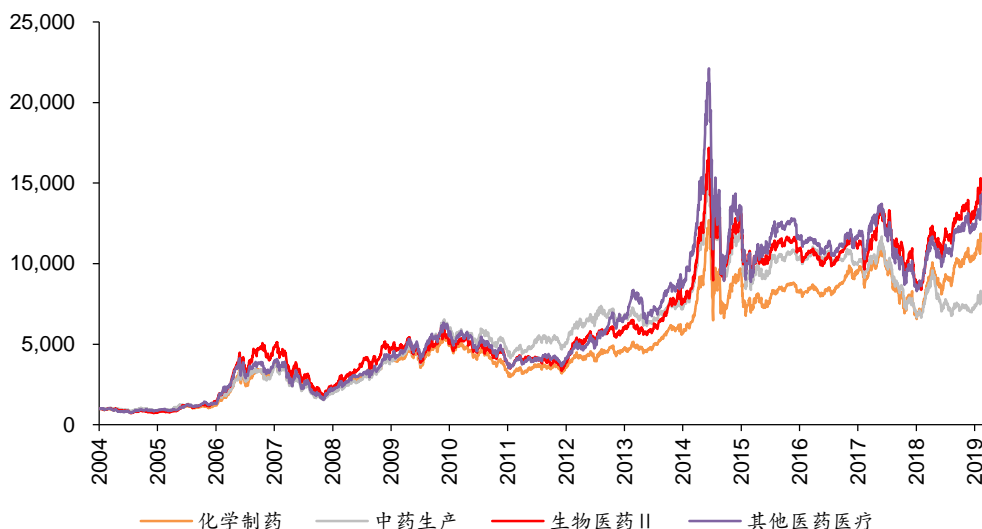
资料来源：Wind，华泰证券研究所

按照前述标准，初筛行业中，建筑、轻工制造、家电三个行业内子行业间的市值分布极不均匀，基本上呈现出一家独大的格局，因此本文不予拆分。而电子、通信、计算机、传媒四个行业均属于典型的成长风格，目前还处于高速增长期，内部竞争格局尚未稳定，个股间收益和基本面的分化呈现出短期性、暂时性、阶段性的特点，在经济结构转型和产业升级的大背景下，未来不确定性较大，为了保证划分体系的灵活性，本文暂不予拆分。综上，后文中我们主要考察有色金属、医药、食品饮料、非银行金融、交通运输五个行业。

### 医药：子行业间收益趋同性高，不予拆分

医药行业下辖化学制药、中药生产、生物医药 II、其他医药医疗 4 个子行业：1、从走势来看，四个子行业间趋同度较高，相关系数均在 0.9 左右；2、从基本面来看，生物制药子行业由于高投入，高风险，高回报的特点，近年来估值溢价更高，而且净利率水平也要显著高于其他三个子行业；而其他医药医疗子行业以医疗服务和医疗器械为主，属于重资产行业，折旧损耗较大，所以净利率水平比其他三个子行业低。整体而言，医药子行业间的基本面分化较为明显，但是收益分化度较低，换言之，该行业可以通过优选个股获得较高的超额回报，但是子行业间轮动效应不明显，因此本文不予拆分。

图表10：医药子行业价格走势



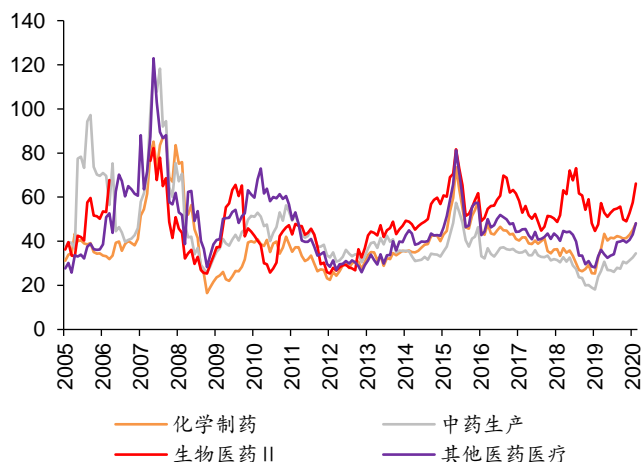
资料来源：Wind，华泰证券研究所

图表11：医药子行业相关系数

	化学制药	中药生产	生物医药 II	其他医药医疗
化学制药	1			
中药生产	0.93	1		
生物医药 II	0.91	0.89	1	
其他医药医疗	0.92	0.90	0.90	1

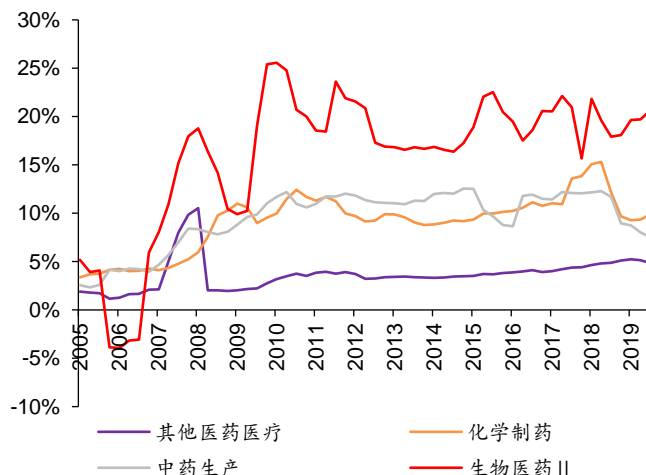
资料来源：Wind，华泰证券研究所

图表12：医药子行业的估值走势



资料来源：Wind，华泰证券研究所

图表13：医药子行业的净利率走势

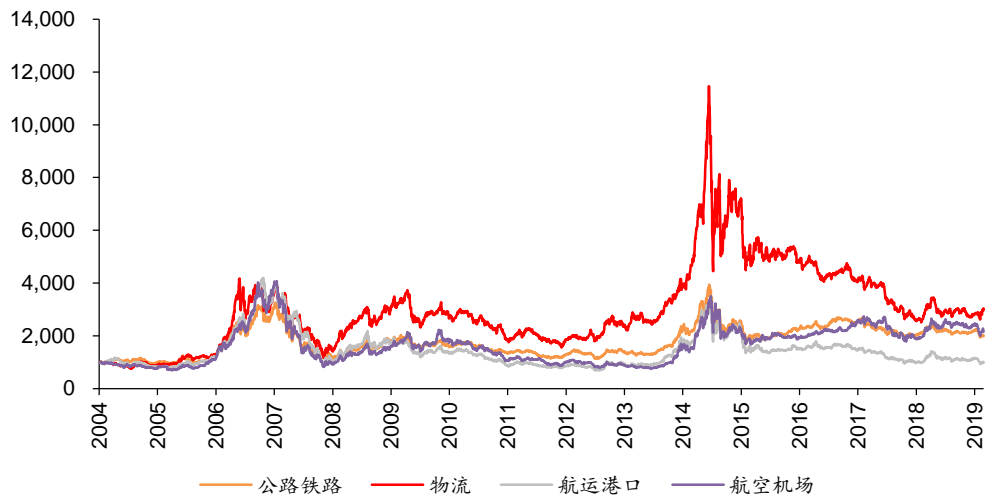


资料来源：Wind，华泰证券研究所

### 交通运输：拆分能提升“高内聚”，但会破坏“低耦合”，综合考虑后不予拆分

交通运输行业下辖公路铁路、物流、航运港口、航空机场 4 个子行业：1、从走势来看，四个子行业间物流行业弹性较大，其他三个子行业的趋同度较高；2、从基本面来看，物流行业受益于近年来电子商务的发展，未来前景可期，兼具成长属性，估值溢价更高；航运港口、航空机场子行业的业绩波动较大，受经济周期的影响更显著；公路铁路子行业的业绩稳定性较高，是典型的防御型行业。整体而言，交通运输子行业间的基本面分化较为明显，而收益分化度适中。拆分的好处是能够提升“高内聚”，也即各子行业内部的核心驱动因素更为纯粹，坏处是破坏了“低耦合”，因为拆分后新得到的航运港口和航空机场行业之间，收益和基本面的分化度都偏低。权衡利弊后，本文暂不拆分交通运输行业。

图表14：交通运输子行业价格走势



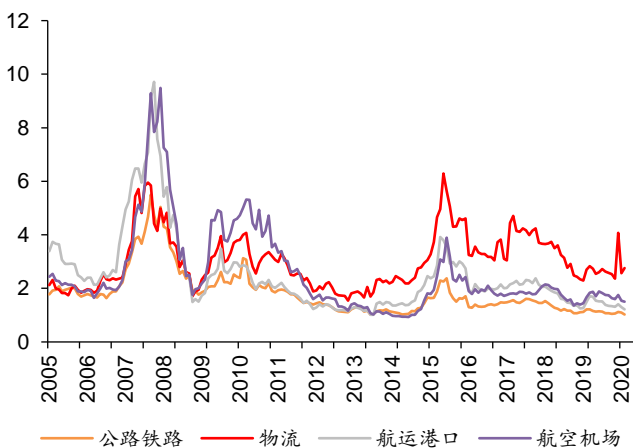
资料来源：Wind，华泰证券研究所

图表15：交通运输子行业相关系数

	公路铁路	物流	航运港口	航空机场
公路铁路	1			
物流	0.79	1		
航运港口	0.82	0.82	1	
航空机场	0.77	0.72	0.76	1

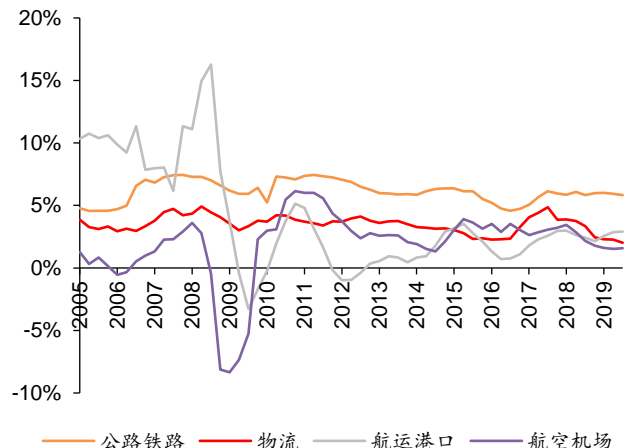
资料来源：Wind，华泰证券研究所

图表16：交通运输子行业估值（PB\_LF）走势



资料来源：Wind，华泰证券研究所

图表17：交通运输子行业 ROA 走势



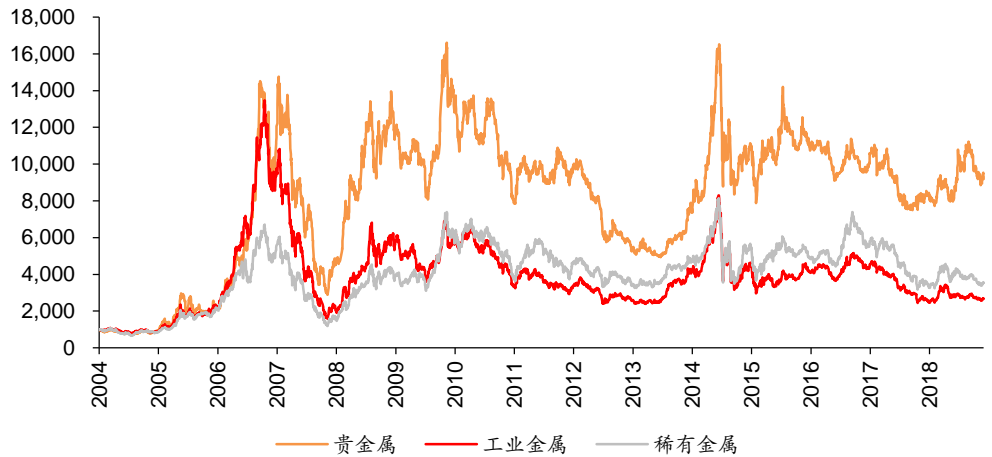
资料来源：Wind，华泰证券研究所



### 有色金属：拆分能提升“高内聚”，但会破坏“低耦合”，综合考虑后不予拆分

有色金属行业下辖贵金属、工业金属、稀有金属 3 个子行业：1、从走势来看，三个子行业间工业金属和稀有金属的趋同度较高，相关系数高达 0.91，贵金属的走势则相对独立；2、从基本面来看，工业金属的业绩波动较大，受经济周期的影响显著，是典型的周期类行业；而稀有金属的业绩波动虽大，但作为新能源汽车产业链的上游，近年来景气度向好，毛利率显著高于其他两个行业，有一定的成长属性；而贵金属行业中，以黄金为代表的品种兼具货币属性、商品属性、金融属性，定价机制较为复杂，基本面走势显著易于其他两个行业。整体而言，无论是走势还是基本面，工业金属和稀有金属的趋同度较高，而贵金属则相对独立，拆与不拆同样是在“高内聚”和“低耦合”之间取舍，为了保持处理口径上的一致性，本文暂不拆分有色金属行业。

图表18：有色金属子行业价格走势



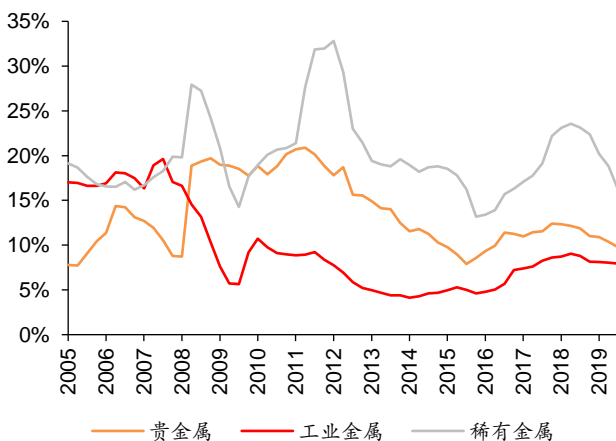
资料来源：Wind，华泰证券研究所

图表19：有色金属子行业相关系数

	贵金属	工业金属	稀有金属
贵金属	1		
工业金属	0.70	1	
稀有金属	0.66	0.91	1

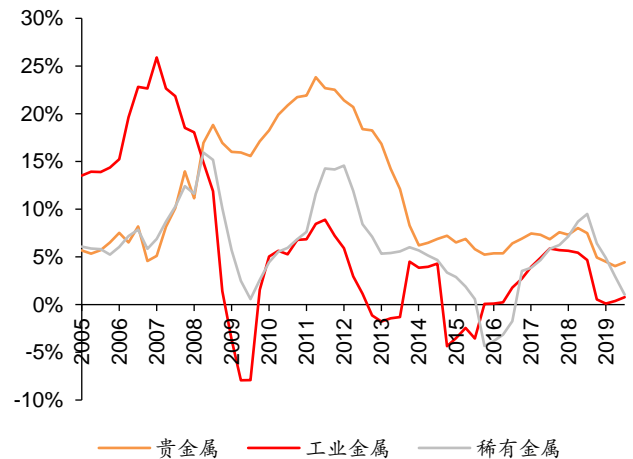
资料来源：Wind，华泰证券研究所

图表20：有色金属子行业毛利率走势



资料来源：Wind，华泰证券研究所

图表21：有色金属子行业 ROE 走势

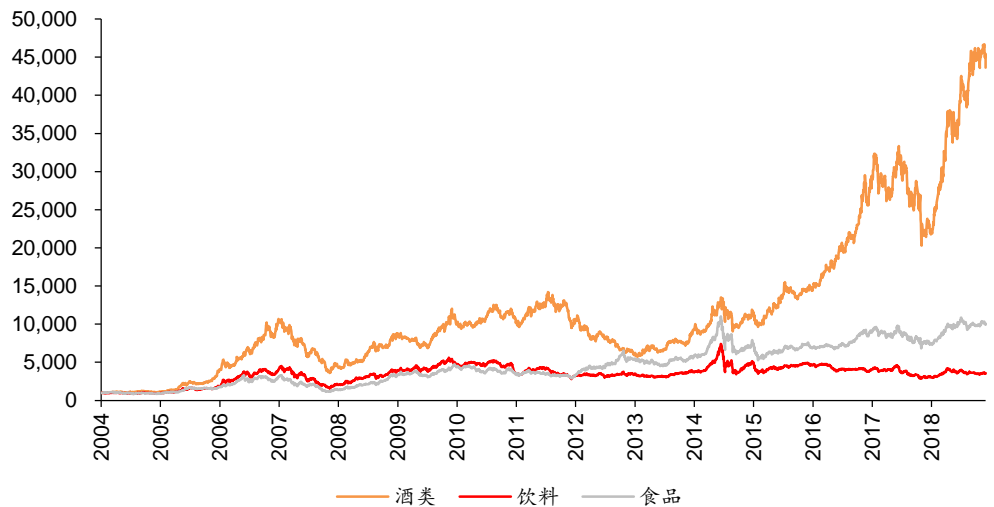


资料来源：Wind，华泰证券研究所

### 食品饮料：价格和基本面走势均分化较大，需要拆分

食品饮料行业下辖酒类、饮料、食品 3 个子行业：1、从走势来看，酒类行业的弹性显著大于饮料和食品行业，是近几年表现最好的行业之一，三个子行业间相关系数均低于 0.8，相较而言趋同度并不高；2、从 ROE 来看，酒类行业的业绩波动更大，属于可选消费风格，具备一定的周期特性；而食品行业受上游猪肉价格的影响明显，2010 年以前业绩波动较大，但近年来相对稳定；对比之下，饮料行业的业绩最为稳定；3、从毛利率来看，三个子行业间存在明显的层次分化，酒类最高，饮料次之，食品最低。整体而言，无论是走势还是基本面，酒类、饮料、食品三个子行业的分化度均较为明显，驱动逻辑各不相同，需要进行拆分。

图表22： 食品饮料子行业价格走势



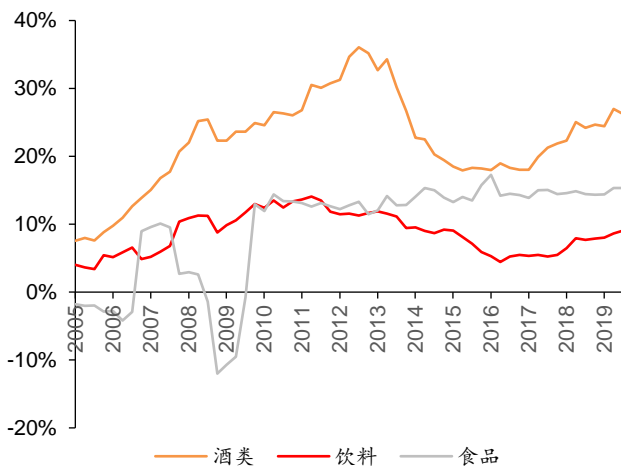
资料来源：华泰证券研究所

图表23： 食品饮料子行业相关系数

	酒类	饮料	食品
酒类	1		
饮料	0.71	1	
食品	0.66	0.78	1

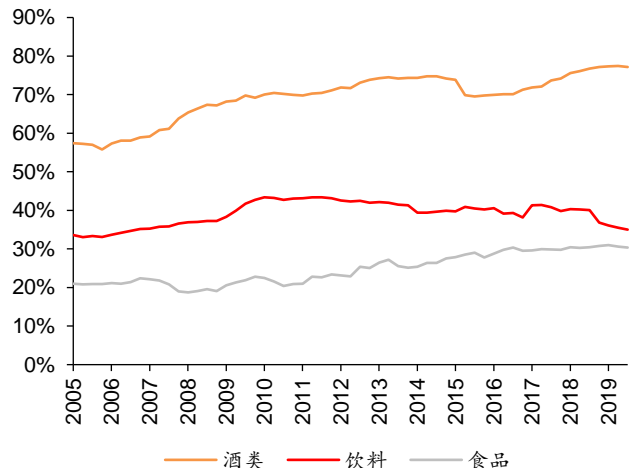
资料来源：Wind，华泰证券研究所

图表24： 食品饮料子行业 ROE 走势



资料来源：Wind，华泰证券研究所

图表25： 食品饮料子行业毛利率走势

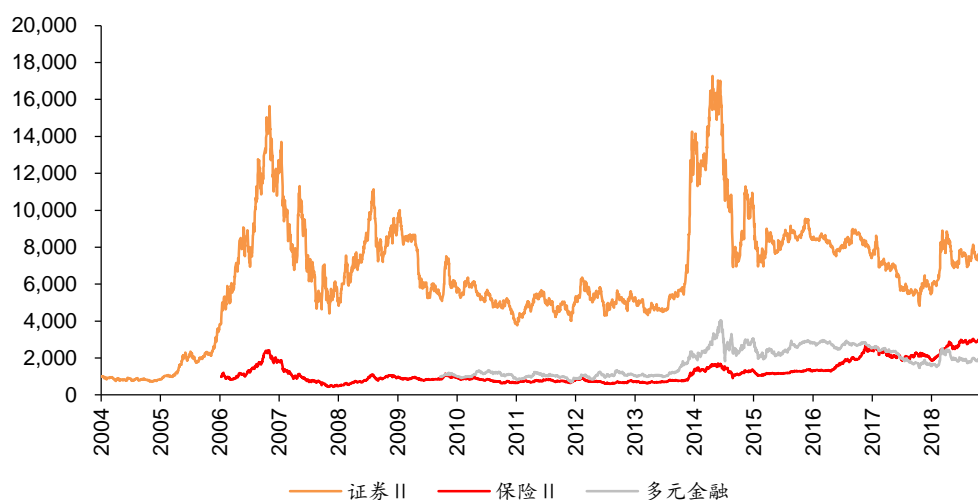


资料来源：Wind，华泰证券研究所

### 非银行金融：价格和基本面走势均分化较大，需要拆分

非银行金融行业下辖证券、保险、多元金融 3 个子行业：1、从走势来看，证券行业的弹性显著大于保险和多元金融行业，三个子行业间相关系数均低于 0.75，其中保险和多元金融行业的相关系数只有 0.5 左右，趋同度较低；2、从 ROE 来看，证券行业的业绩波动更大，是典型的周期风格；而保险行业业绩也有一定的周期波动特征，其周期性主要体现在投资收益的波动上，但同时我们注意到，2016 年以来，保险行业的盈利能力（ROE）是在逐步攀升的，有研究表明，随着人均 GDP 提升，保费支出也呈现出加速增长的态势，因而保险行业兼具一定的成长性；3、从净利率来看，保险行业的波动中枢显著低于证券和多元金融行业。整体而言，无论是走势还是基本面，证券、保险、多元金融三个子行业的分化度均较为明显，驱动逻辑各不相同，需要进行拆分。

图表26： 非银行金融子行业价格走势



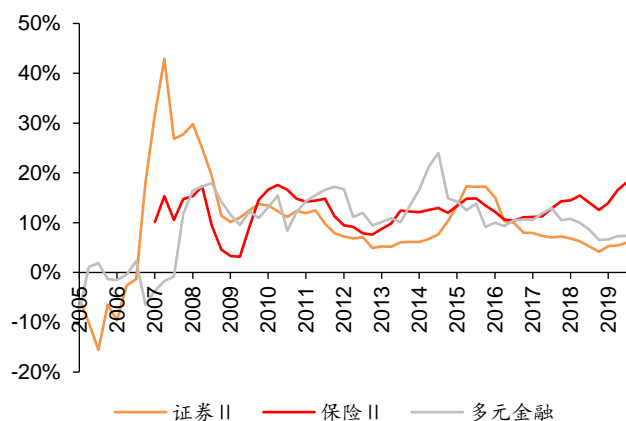
资料来源：Wind，华泰证券研究所

图表27： 非银行金融子行业相关系数

	证券 II	保险 II	多元金融
证券 II	1		
保险 II	0.71	1	
多元金融	0.70	0.54	1

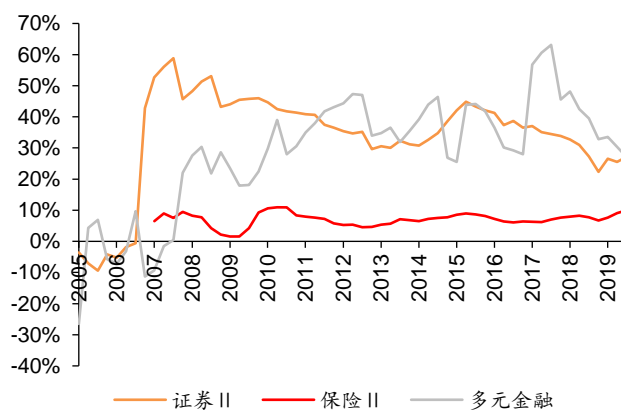
资料来源：Wind，华泰证券研究所

图表28： 非银行金融子行业 ROE 走势



资料来源：Wind，华泰证券研究所

图表29： 非银行金融子行业净利率走势



资料来源：Wind，华泰证券研究所

## 行业拆分方案评估

总结前文，我们在综合考虑个股分化和子行业分化后，决定将食品饮料拆分成酒类、饮料、食品三个子行业，将非银行金融拆分成证券、保险、多元金融三个子行业。本节将通过定量手段，证明新的行业划分体系相比于原一级行业划分体系更加纯粹，“高内聚、低耦合”特性更加显著，也即行业间的分化更大，而行业内的分化更小。

### 个股收益同质性检验

对于特定行业分类方案，可以通过检验行业内个股相关系数是否显著高于行业间个股相关系数来判定该分类方案是否把收益趋同性高的股票归到同一个行业。

具体而言，针对行业 $s$ ，假设行业内个股集合为 $I$ ，行业外个股集合为 $C$ ，行业内个股的平均相关系数记为：

$$\rho_{II} = \text{Mean}(\rho(r_i, r_j)), \quad i, j \in I$$

而行业间的平均相关系数记为：

$$\rho_{IC} = \text{Mean}(\rho(r_i, r_j)), \quad i \in I, j \in C$$

假设对于任意行业 $s$ ，均有 $\rho_{II} > \rho_{IC}$ ，则说明行业分类方案有效的将收益趋同性高的个股归为了一类。实证过程中，新方案和旧方案的对比，可以通过被拆分行行业在拆分前后行业内外平均相关系数的差异进行判断。

图表30： 拆分前后行业内外平均相关收益系数（近 10 年）

拆分前	行业内	行业外	差异	拆分后	行业内	行业外	差异
食品饮料	0.31	0.28	0.03	酒类	0.44	0.30	0.13
				饮料	0.32	0.29	0.04
				食品	0.27	0.27	0.00
非银行金融	0.49	0.31	0.18	证券 II	0.60	0.33	0.28
				保险 II	0.53	0.26	0.27
				多元金融	0.31	0.27	0.04

资料来源：Wind，华泰证券研究所

图表31： 拆分前后行业内外平均相关收益系数（近 5 年）

拆分前	行业内	行业外	差异	拆分后	行业内	行业外	差异
食品饮料	0.32	0.29	0.03	酒类	0.45	0.32	0.13
				饮料	0.33	0.30	0.03
				食品	0.27	0.27	0.00
非银行金融	0.52	0.33	0.19	证券 II	0.63	0.35	0.29
				保险 II	0.52	0.25	0.27
				多元金融	0.36	0.30	0.06

资料来源：华泰证券研究所

图表32： 拆分前后行业内外平均相关收益系数（近 3 年）

拆分前	行业内	行业外	差异	拆分后	行业内	行业外	差异
食品饮料	0.29	0.25	0.04	酒类	0.41	0.26	0.15
				饮料	0.31	0.26	0.04
				食品	0.25	0.24	0.01
非银行金融	0.53	0.31	0.23	证券 II	0.65	0.33	0.32
				保险 II	0.56	0.22	0.34
				多元金融	0.37	0.27	0.10

资料来源：Wind，华泰证券研究所

实证结果表明，两个被拆分行业在拆分后，行业内个股收益同质性都有一定改善。对于食品饮料行业，拆分前行业内外的收益分化度并不显著，而拆分后酒类子行业的个股收益趋同性显著提升；对于非银行金融行业，拆分前个股的收益趋同性就比较高，行业内平均相关系数要显著高于行业间平均相关系数，而拆分后，证券和保险子行业内个股收益的趋同度进一步提升。

### 基本面同质性检验

对于特定行业划分方案，其行业内个股的基本面同质性可以通过财务指标的联合方差表征，因此不同行业划分体系的优劣可以通过联合方差检验进行判定。

具体而言，对于行业分类 $a$ 和行业分类 $b$ ，在财务指标 $S$ 下，分别计算每个行业分类的联合方差：

$$S_a^2 = \frac{\sum_{i=1}^N (n_i - 1) V_i}{\sum_{i=1}^N (n_i - 1)}$$

其中， $N$ 表示方案中分成的行业个数， $n_i$ 表示方案中第 $i$ 个行业中股票数量， $V_i$ 表示方案中第 $i$ 个行业中各个成分股财务指标 $S$ 的方差。对于两个行业划分方案 $a$ 和 $b$ ，其关于财务指标 $S$ 的同质性可以通过计算检验统计量 $F$ 表征：

$$F = \frac{S_a^2}{S_b^2} \sim F(N_a - 1, N_b - 1)$$

其中 $N_a$ 及 $N_b$ 分别表示行业划分方案 $a$ 和 $b$ 中的行业个数。对于任何两种行业划分方案 $a$ 和 $b$ ，可分别计算各财务指标对应的行业联合方差及检验统计量 $F$ 。若 $F$ 显著大于1，则说明方案 $a$ 中股票关于财务指标 $S$ 的方差更大，其同质性低于方案 $b$ ，也即个股基本面趋同度更低；反之则说明其同质性高于方案 $b$ 。

在实证中，记新的行业拆分体系为 $a$ （剔除综合后一共有 33 个行业），原一级行业划分体系记为 $b$ （剔除综合后一共有 29 个行业），分别利用 2010 年至 2018 年的年报数据计算联合方差统计量。结果表明，拆分后的行业体系相比于原一级行业体系基本面同质性更高，接近 70% 的场景下  $F$  检验统计量都小于 1，尤其是在 ROIC、毛利率等盈利指标下，拆分后的行业划分体系中，个股基本面趋同度更显著。

图表33：联合方差检验结果

	2010	2011	2012	2013	2014	2015	2016	2017	2018
ROA	0.9901	0.9736	0.9721	1.0034	1.0032	0.9967	0.9913	0.9956	0.9979
ROE	0.9837	0.9762	0.9837	1.0033	1.0150	0.9982	0.9963	0.9952	0.9983
ROIC	0.7847	0.8159	0.6736	0.5916	0.6990	0.7945	0.7038	0.9833	0.9281
净利率	0.9866	0.9754	0.9874	1.0037	1.0335	1.0272	0.9700	0.9638	0.9922
毛利率	0.9791	0.9652	0.9667	0.9666	0.9730	0.9714	0.9676	0.9701	0.9515
资产周转率	1.0064	1.0018	1.0024	1.0022	0.9970	1.0030	0.9986	1.0030	1.0038
资产负债率	0.9990	0.9975	0.9993	0.9996	1.0011	1.0025	1.0015	0.9981	0.9981
存货周转率	1.0022	1.0020	1.0009	1.0008	1.0016	0.9420	0.0556	0.9944	0.1410

资料来源：Wind，华泰证券研究所



## 行业聚类实证

基于自上而下的研究视角，投资者通常会从经济结构、产业链出发将各个细分行业进行归并，便于研究宏观环境对大类板块的影响，以及产业链上下游之间的传导规律。鉴于此，我们还需要对拆分后的行业划分体系进行聚类研究。

### 聚类分析方法

聚类是一种常见的数据分析工具，其目的是把大量数据点的集合分成若干类，使得归于同一类的数据之间相似性尽可能大，而归于不同类别的数据之间差异性尽可能大。K-means 算法作为一种经典的基于距离的聚类方法，逻辑清晰，计算效率高，在实证中得到广泛应用，其基本步骤为：

1. 根据先验知识事先确定需要聚类的簇的个数  $k$ ；
2. 在样本中随机选取  $k$  个对象，每个对象初始地代表了一个簇的中心；
3. 对剩余的每个对象，根据其与各簇中心的距离，将它赋给最近的簇；
4. 计算每个簇的平均值，作为新的簇中心；
5. 判定：若簇中心不再发生变动或达到迭代次数，结束，否则返回第 2 步，不断重复。

可见，K-means 算法的核心思想是采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。然而实证过程中我们发现，该算法的聚类结果对先验设定的簇个数  $k$ ，以及随机选取的初始点非常敏感。换言之，对于不同的  $k$  值，同样的两个行业是否归为同一类结论可能不同；即便是固定  $k$  值，以及输入数据，结论也会因为初始值的随机性而不固定，尤其是在样本较大的时候，算法容易陷入局部最优解。

为了解决  $k$  值敏感性，需要结合先验知识进行合理设定，中信行业划分体系在三级行业的基础上构建了周期、金融、消费、成长、稳定五大风格板块，因而本文实证中统一设置簇个数  $k$  为 5；为了解决单次运行中的初值敏感问题，我们采用了多次模拟求概率的方法，具体而言，针对目标区间的行业收益率序列，执行如下步骤：

1. 基于收益率序列之间的欧式距离刻画行业的远近程度，并进行 K-means 聚类，对于任意两个行业，如果他们被归为一类，则记录为 1，否则记为 0；由于拆分后一共有 33 个行业（剔除综合），所以一共有 1089 个行业配对关系；
2. 上述流程重复执行 1000 次，计算每两个行业被归为一类的次数占比，占比越高，说明这两个行业的相似性越强。

得到每两个行业被归为一类的概率占比之后，可以通过**最大生成树**算法构建最终的行业关联关系，并梳理出聚类结果。简单来说，假设一共有  $N$  个节点，最大生成树就是找到  $N - 1$  条边将所有节点连接起来，并且使得这  $N - 1$  条边的权值之和达到最大，具体而言，我们可以通过每两个行业之间被归为一类的概率统计结果构建无向图  $G = (V, E)$ ，其中， $V$  代表所有的节点集合（本例中就是各个行业）， $E$  代表了所有的连接边集合，每条边的权重就是其所连接的两个行业被归为一类的概率。若存在一个无回路的子集  $T$ ， $T$  是  $E$  的子集，它连接了所有的节点，并且使得

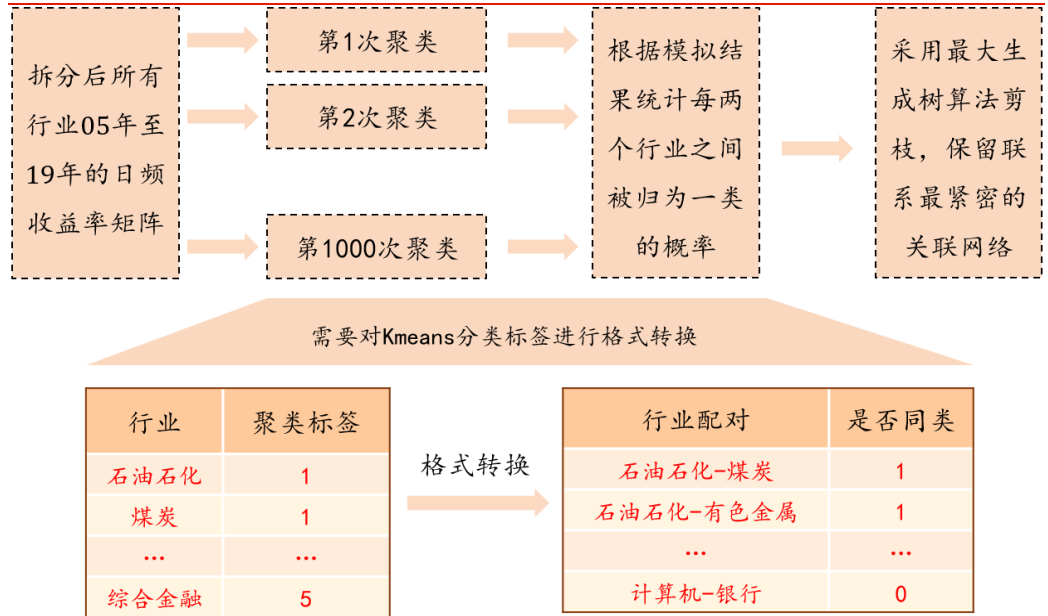
$$w(T) = \sum_{(\mu, \nu) \in T} w(\mu, \nu)$$

达到最大，则称  $T$  为  $G$  的最大生成树。在实现过程中一般采用 Kruskal 算法，原理如下：

1. 将所有的边按照权重大小（本例中就是两个节点被归为一类的概率）递减排序；
2. 选择最大权重的边，判断是否在当前的生成树中形成了一个环路，如果环路没有形成，则将该边加入最大生成树中，否则放弃；
3. 重复步骤 2，直到有  $V - 1$  条边在最大生成树中。

可见，最大生成树是一种剪枝算法，它把那些距离较远，权值较低的边去掉，而剩下的边就是我们要重点关注的行业关联关系。

图表34： 聚类分析方法

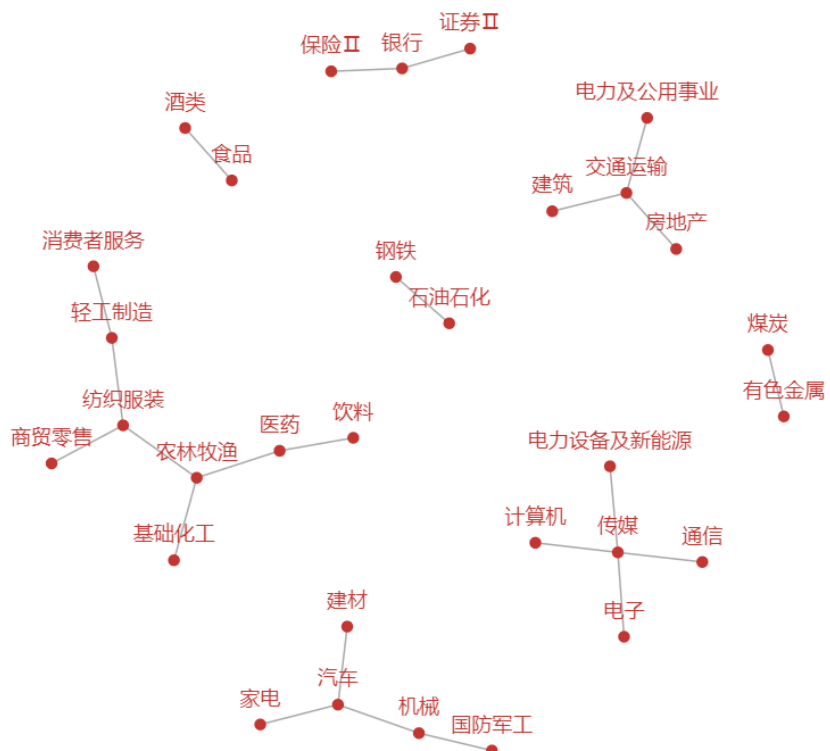


资料来源：华泰证券研究所

## 聚类分析结果

首先考察基于最大权值边构建的简单网络，也即针对每个行业，只保留它与外界连接的边中权值最大的那一条。结果表明，所有行业被拆分成相对独立的多个类别，比如以银行、证券、保险为核心的金融风格；以计算机、传媒、电子、通信为核心的成长风格；以家电、汽车为核心的可选消费品风格；以医药、纺织服装为核心的必须消费品风格；以有色、煤炭、钢铁、石油石化为核心的强周期风格。整体来看，各类别的划分结果与定性认知基本一致，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35： 基于最大权值边构建的行业关联网络

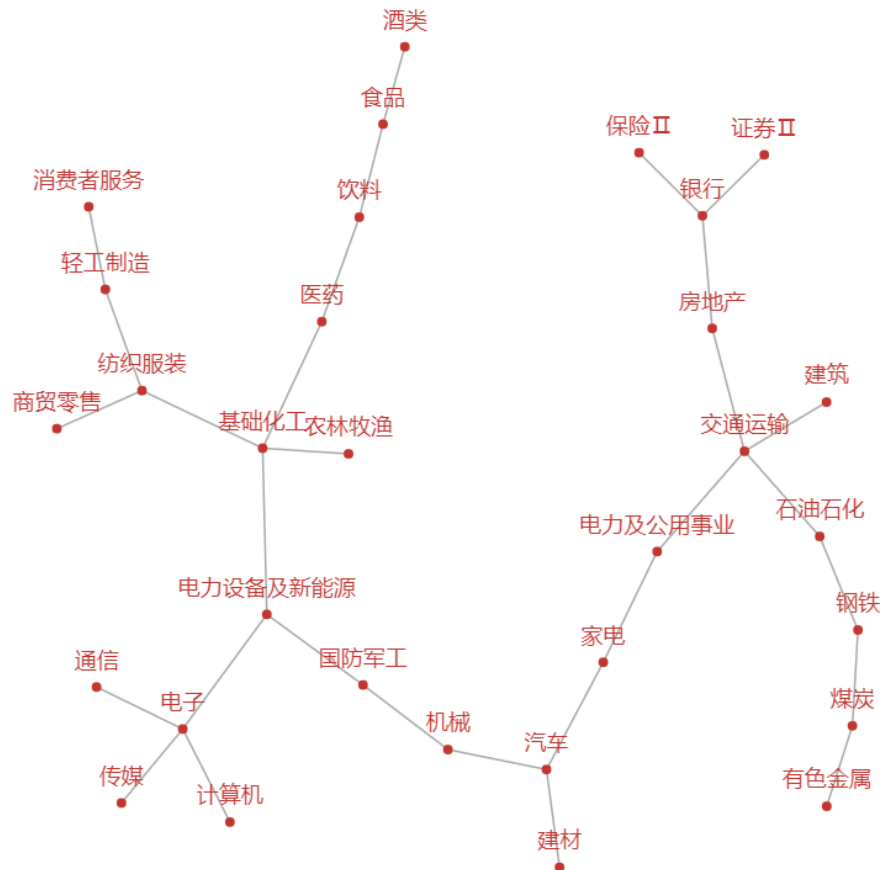


资料来源：Wind，华泰证券研究所

进一步，考察基于最大生成树构建的全连接行业关联网络，图中节点与节点间的距离反映了行业之间的相似程度，结果表明：

1. 关联网络图左半部分更贴近下游消费端，以基础化工、纺织服装为核心的轻工产业链串联起医药、食品、饮料、农林牧渔等必须消费品行业，而以计算机、电子、通信、传媒为核心的成长板块则通过电力设备及新能源行业与必须消费品链条衔接起来。
2. 关联网络图右半部分更贴近上游生产端，以有色、煤炭、钢铁、石油石化为核心的资源类强周期行业，和以银行、保险、证券、房地产为代表的金融类板块通过交通运输行业衔接起来，而交通运输、电力及公用事业都是典型的基础设施类行业，现金流相对稳定，属于弱周期风格。
3. 左半部分的下游消费端行业集群，和右半部分的上游生产端行业集群，通过机械、汽车、国防军工等中游制造业行业联系起来，构建出完整的产业链图谱。
4. 从节点距离来看，以计算机、电子、通信、传媒为代表的成长类行业，和以银行、证券、保险为代表的金融类行业，相互间距离最远，分化最大；而以有色金属、钢铁、煤炭、石油石化为代表的上游资源类行业，和以食品、饮料、纺织服装为代表的必须消费类行业距离最远，分化最大。

图表36： 基于最大生成树构建的行业关联网络



资料来源：Wind，华泰证券研究所

综上，我们按照先验设定的周期、消费、金融、成长、稳定五大风格，得到如下聚类结果：

1. 周期风格，下辖上游资源、中游材料、中游制造三个主题板块。其中，上游资源包含石油石化、煤炭、有色金属三个行业；中游材料包含钢铁、建材、基础化工三个行业；中游制造包含机械、电力设备及新能源、国防军工三个行业；
2. 消费风格，下辖可选消费和必须消费两个主题板块，其中，可选消费包含汽车、家电、酒类三个行业，它们都具备一定的周期属性，对消费需求变动的弹性较大；必须消费包含食品、饮料、纺织服装、医药、农林牧渔、消费者服务、商贸零售、轻工制造八个行业。整体而言，必须消费偏衣食类需求，而可选消费偏住行类需求；
3. 金融风格，下辖大金融主题板块，包含银行、证券、保险、多元金融、综合金融、房地产六个行业；
4. 成长风格，下辖 TMT 主题板块，包含计算机、电子、传媒、通信四个行业；
5. 稳定风格，下辖公共产业主题板块，包含电力及公用事业、交通运输、建筑三个行业。

图表37： 板块聚类结果

大类风格	主题板块	细分行业
周期	上游资源	石油石化、煤炭、有色金属
	中游材料	钢铁、建材、基础化工
	中游制造	机械、电力设备及新能源、国防军工
消费	可选消费	汽车、家电、酒类
	必须消费	食品、饮料、纺织服装、医药、农林牧渔、消费者服务、商贸零售、轻工制造
金融	大金融	银行、证券Ⅱ、保险Ⅱ、多元金融、综合金融、房地产
成长	TMT	计算机、电子、传媒、通信
稳定	公共产业	电力及公用事业、交通运输、建筑

资料来源：华泰证券研究所

图表38： 聚类结果图示



资料来源：华泰证券研究所

## 总结

本文作为基本面轮动系列报告第一篇，主要目的是为后续系列报告确立统一的研究对象，核心工作是开展行业拆分与聚类实证。

一个理想的行业划分体系应该具备“高内聚，低耦合”的特征，也即行业间的分化度应尽可能大，而行业内的分化度应尽可能小。基于此原则，我们以中信一级行业为基础，分析了每个行业内个股收益和基本面的分化度，并综合考虑各个行业的生命周期发展阶段，以及市值占比因素后，最终决定将食品饮料拆分成酒类、饮料、食品三个子行业；将非银行金融拆分成证券、保险、多元金融三个子行业；剔除掉综合行业后，拆分后的行业划分体系一共包含 33 个细分行业，分别是：石油石化、煤炭、有色金属、钢铁、基础化工、建材、机械、电力设备及新能源、国防军工、汽车、家电、酒类、食品、饮料、医药、纺织服装、农林牧渔、商贸零售、轻工制造、消费者服务、银行、证券、保险、多元金融、综合金融、房地产、计算机、电子、传媒、通信、交通运输、电力及公用事业、建筑。

进一步，基于拆分后的行业划分体系，我们采用蒙特卡洛模拟和最大生成树算法构建行业间关联网络，进行聚类研究。结果表明，33 个细分行业可以聚类成周期、消费、金融、成长、稳定五大风格，以及上游资源、中游材料、中游制造、可选消费、必须消费、大金融、TMT、公共产业八大主题板块。整体而言，聚类结果与基于市场经验的定性认知基本一致，充分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。

## 风险提示

模型根据历史规律总结，历史规律可能失效。长期来看，行业拆分与聚类结果反映了国民经济体系中的产业链全景图谱，随着经济结构转型升级，未来有变动可能。



## 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门、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：91320000704041011J。

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 2020 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

深圳市福田区益田路 5999 号基金大厦 10 楼/邮政编码：518017

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