

深度报告

金融工程

数量化投资系列之十

行业配置专题报告

2009年6月29日

专题报告

行业轮动与资产配置

本报告的独到之处

■全面介绍了聚类分析中行业研究中的具体应用，并将聚类分析和 Grange 相结合对不同市场阶段各个行业的轮动次序做了划分；

■首次聚类分析研究了中国股市在不同市场阶段行业分类的差异；

■使用 Granger 因果检验考察了上涨阶段三类行业之间的轮动关系，给出了具体的配置建议。

相关研究报告：

《金融工程研究-数量化投资系列之三-基于景气指数的行业选择与行业配置》

分析师 戴 军

电话：021-68864585

E-mail: daijun@guosen.com.cn

分析师 董艺婷

电话：021-68866946

E-mail: dongyt@guosen.com.cn

分析师 林晓明

电话：021-68866946

E-mail: linxming@guosen.com.cn

分析师 葛新元

电话：0755-82130833-1870

E-mail: gexy@guosen.com.cn

独立性声明：

作者保证报告所采用的数据均来自合规渠道，分析逻辑基于本人的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

伴随着经济全球化和信息技术的发展是以前所未有速度激增的金融数据。如何发现并提取隐匿浩瀚数据间的规律，获取能为投资者利用的知识并以此预测金融市场未来变化的趋势成为研究者日益关心的问题。数据挖掘为此提供有力和有效的工具，它能够从大量数据中发现有效的、新奇的、潜在有用的和最终可理解的模式。其中，聚类分析作为一种重要的数据挖掘方法在金融研究领域被广泛运用，已经被国内外的研究者应用于证券投资分析，行业间收益率关系，基金投资风格划分，股市板块羊群效应等多个研究领域。

本报告采用分阶段的谱系聚类法对计算得到的各阶段行业收益的差异进行聚类。在本文中把行业类别作为样品，把时间作为变量，每阶段的行业收益率作为指标。本文谱系聚类共计两次。第一次，采用 R 型聚类的方法对研究的时间变量进行聚类，以便确定哪些阶段，各个行业收益率相似。把聚类得到的结果作为对研究样本的时间范围分阶段的依据第二次根据划分的时间段对样品进行 Q 型聚类。聚类完成后，对所得的三类行业使用 Granger 因果性检验方法来研究彼此之间的轮动关系，提供具体的投资建议。

经实证研究表明：

- 1、从2000年至今，一、二、三类的行业分类比较稳定，一、二类行业获取超额收益（下跌阶段为超额负收益），三类行业获取市场平均收益；
- 2、在市场处于重要拐点的时候，三大类行业的收益率差异相对不明显，这也是以后使用Q型聚类分析市场是否处于拐点的一个重要方法；
- 3、获取超额收益的一二类行业比较稳定，主要有：采掘业、有色金属、黑色金属、金融服务、房地产、交运设备、信息服务和信息设备；
- 4、对于现阶段来说，时间变量的R型聚类将其归为一类（R型聚类），而一类时间变量（表2第二列）大部分是市场拐点，但是Q型聚类三类行业的收益率差异比较明显，目前还不满足拐点条件；
- 5、在上涨阶段，三类行业最先启动，二类行业的滞后一类行业在三阶左右（三个月），一类行业之后二类行业中二阶左右（两个月）。三大类行业轮动的次序为：三类行业——>二类行业（3M）——>一类行业（2M）；
- 6、目前建议超配一类行业，即使采掘业、有色金属、黑色金属、金融服务、房地产、交运设备、信息服务和信息设备。但需密切关注三类行业收益率的差异化，当差异化逐渐降低的时候，应转向第三类行业。

内容目录

研究方法	4
聚类分析	4
Granger 因果检验	5
数据说明	5
对时间变量的 R 型聚类分析	5
固定时间长度聚类	5
指定时间段聚类	7
对行业类别的 Q 型聚类分析	8
聚类后的 Granger 因果检验	9
结论与建议	10

图表目录

图 1: 固定时间长度进行 R 型聚类的结果	6
图 2: 2000.1-2009.6 上证综指走势图及各阶段编号	7
图 3: 指定时间段进行 R 型聚类的结果	8
表 1: R 型聚类说明	6
表 2: 对行业类别 Q 型聚类结果说明	8
表 3: 三类行业 ADF 单位根检验	9
表 4: 三类行业 Granger Causality 检验结果	9

概述

市场在不同的阶段运行，即便是处于一个长期上涨的景气阶段也不例外，都会经历相同的周期：上涨，顶部，下跌，底部，而每个阶段都会对不同的行业产生不同的影响。行业配置是获取超额收益的重要来源，据国外的实证研究统计，共同基金大部分超额收益都可以用行业配置来解释，也就是说在成熟市场，行业配置在基金的投资组合管理中占据了首要的地位，其作用强于个股的选择。在我国基金规模不断扩大到前景下，个股的选择在基金中的位置将有所下降，由于行业内股票同质化倾向严重，使得行业配置的作用显得尤为重要。

行业轮动效应也就是说市场总是在变换着追捧某些行业，不存在长期受追捧的某一行业。在经济周期的不同阶段，不同行业的受益情况不同，某些行业价格指数会独立与其他行业指数的走势，因此，进行动态的行业配置或者行业轮动策略会创造出客观的超额收益。事实上，从资产配置的视角看，我们也可以把行业轮动策略看成是基于行业景气周期的资产配置策略。

伴随着经济全球化和信息技术的发展是以前所未有速度激增的金融数据。如何发现并提取隐匿浩瀚数据间的规律，获取能为投资者利用的知识并以此预测金融市场未来变化的趋势成为研究者日益关心的问题。数据挖掘为此提供有力和有效的工具，它能够从大量数据中发现有效的、新奇的、潜在有用的和最终可理解的模式。其中，聚类分析作为一种重要的数据挖掘方法在金融研究领域被广泛应运，它无需依赖预先定义的类和带类标号的训练实例，通过采用一定的距离准则和相似性准则对物理或抽象的对象集合合并或分割，使得合并或分割后类内距离小、类间距离大。聚类方法已经被国内外的研究者应用于证券投资分析，行业间收益率关系，基金投资风格划分，股市板块羊群效应等多个研究领域。

研究方法

聚类分析

本文旨在利用聚类方法研究不同时间段行业收益率的相互关系并分析聚类结果产生的原因。由于各行业收益率的差异同样不是一成不变的，因此本文采用动态聚类方法进一步研究不同行业间收益率的相互关系。所谓的分阶段的动态聚类其本质是某一种聚类方法，只不过假设聚类样品之间的聚类关系会随着时间的推移而发生变化。所以需要把研究的时间范围划分成几个阶段，在各个阶段中采用该种聚类方法对样品聚类。谱系聚类法是一个应用较为广泛的聚类方法。根据谱系聚类法，首先各个样品(或变量)自成一类，然后把最相似(指标间距离最小)的样品(或变量)聚为小类，再将已聚合的小类按其相似性(用类间距离度量)进行进一步聚合，如此逐步扩大，直到所有的样品(或变量)都聚集完毕，形成一个表示亲疏关系的谱系图。最后，根据谱系图按照某种标准对样品(或变量)进行分类。其中对样品的谱系聚类称为 Q 型聚类，对变量的聚类称为 R 型聚类。

本文采用分阶段的谱系聚类法对计算得到的各年行业收益的差异进行聚类。在本文中把行业类别作为样品，把时间作为变量，每阶段的行业收益率指标。本文谱系聚类共计两次。第一次，采用 R 型聚类的方法对研究的年份变量进行聚类，以便确定哪些年份，各个行业收益率相似。把聚类得到的结果作为对研究样本的时间范围分阶段的依据第二次根据划分的时间段对样品进行 R 型聚类。

Granger 因果检验

Granger 因果检验是检验经济变量之间因果关系的一种常用方法。尽管使用聚类分析能得出不同市场阶段各个行业大致的分类，但这是基于市场现象得出的一种相关关系，互为相关不一定互为因果，因此对不同时间段内三类行业因果关系的检验是把握行业轮动的基本手段。透过充满随机性和不确定性的市场现象，揭示三类行业之间轮动的因果关系，是在行业轮动效应下对资产进行配置的前提。本报告中对行业进行 Q 型聚类分析的基础上，对不同市场阶段三大类行业进行 Granger 因果检验，从而把握不同市场阶段行业轮动所带来的超额收益。

数据说明

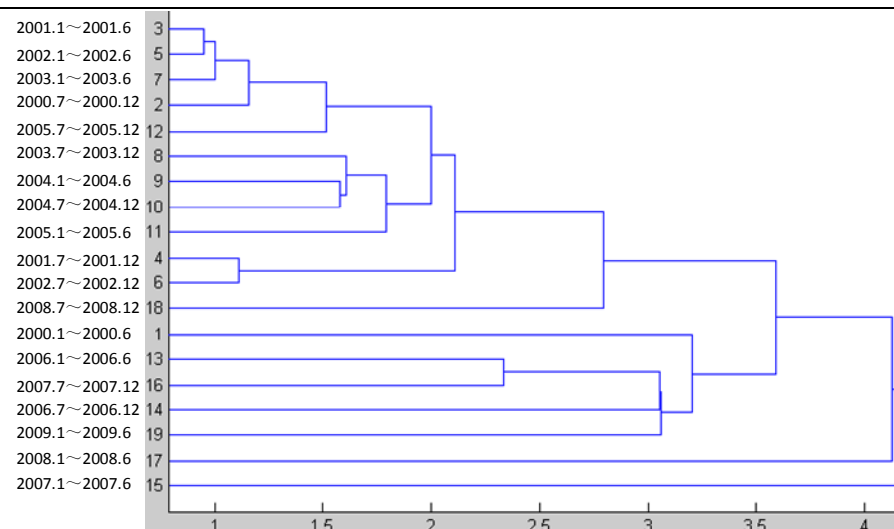
本报告使用申万一级行业指数的月收益率数据作为样本，时间段为2001年1月~2009年6月，市场基准指数使用Wind全A指数。

对时间变量的 R 型聚类分析

固定时间长度聚类

本文首先对年度变量就行R型聚类，以确定哪些阶段的各行业的收益率相类似。以此作为第二步对各个行业收益率进行Q型聚类划分时间段的依据。首先采用固定时间段（6个月）进行R型聚类分析，结果如图1、表1所示。

图 1: 固定时间长度进行 R 型聚类的结果



数据来源: 国信证券经济研究所

从上图可以看出, 2000年1月至2009年5月, 不同的市场阶段可以分为6大类, 分别为: 震荡上涨、震荡下跌、上涨、下跌以及07年的大涨和08年的大跌。明细参阅表1。图2给出了行情走势图与各个分类之间的对应关系。

表 1: R 型聚类说明

类别	时间编号	市场走势	备注
第一类	2,3,5,7,12	震荡上涨	
第二类	8,9,10,11	震荡下跌	
第三类	4,6	下跌	
第四类	1,13,14,16,19	上涨	
第五类	15	07 年上涨	说明 07 年的单边上涨行情比较特殊
第六类	17	08 年下跌	说明 08 年的单边下跌行情比较特殊
第七类	18	08 年下跌	说明 08 年下跌前半段和后半段存在差异

数据来源: 国信证券经济研究所

我们认为, 07年的上涨之所以和其他阶段的上涨无法归类在一起, 除了涨幅巨大这一特征外, 还因为本轮上涨的推动因素与以往不同, 开始与06年的股权分置改革, 是中国股市的一个历史性事件, 它对未来中国股市的发展影响深远, 这是导致本轮大牛市不同于以往最本质的原因。开始与07年底的大幅调整也是不同寻常的, 首先是在此之前的一轮大牛市所积聚的巨大的系统性风险需要释放, 其次是开始与07年上半年的次贷危机愈演愈烈。同时, 2008年9月14日美国第四大投行雷曼兄弟宣布破产, 同一天美国第三大投行美林证券被收购, 以此为标志的金融风暴爆发, 国内股市更是雪上加霜, 这也是08年熊市内在成因前后两段相异的重要原因。

图 2: 2000.1-2009.6 上证综指走势图及各阶段编号



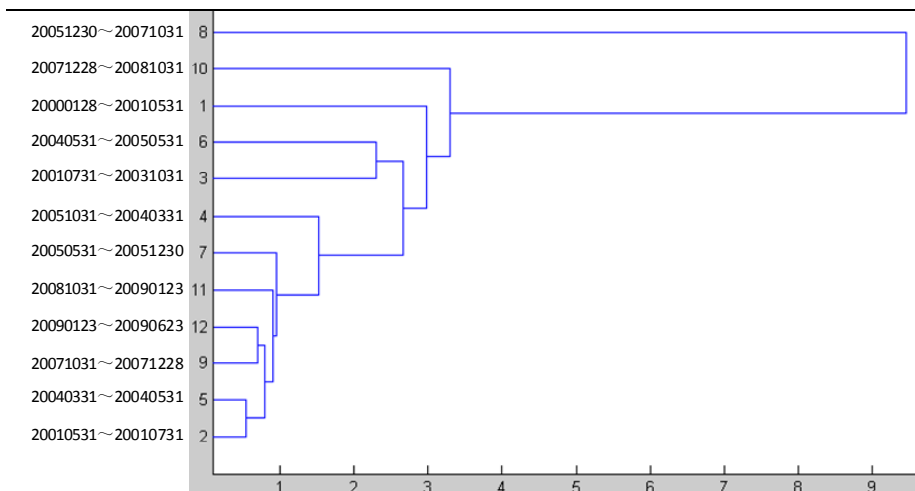
资料来源: 国信证券经济研究所, 同一种颜色为同一分类, 对数坐标

指定时间段聚类

上文固定时间段, 从2000年开始, 对每半年的中国股市时间变量进行R型聚类, 结果与直观认识一致, 并且能发掘一些细微的差异。说明使用聚类分析方法对中国股市进行分析是可行的。本部分按照中国股市每次的上涨、下跌、顶部、底部的分类, 指定对应的时间段对其进行聚类分析, 具体时间段如下:

- 1、20000128~20010531: 上涨, 5个月
- 2、20010531~20010731: 顶部, 3个月
- 3、20010731~20031031: 下跌, 15个月
- 4、20031031~20040331: 反弹, 6个月
- 5、20040331~20040531: 顶部, 3个月
- 6、20040531~20050531: 下跌, 12个月
- 7、20050531~20051230: 底部, 6个月
- 8、20051230~20071031: 上涨, 22个月
- 9、20071031~20071228: 顶部, 3个月
- 10、20071228~20081031: 下跌, 10个月
- 11、20081031~20090123: 底部, 3个月
- 12、20090123~20090623: 上涨, 6个月-

图 3: 指定时间段进行 R 型聚类结果



数据来源: 国信证券经济研究所

从上图可以看出, 对于指定的时间段, R型聚类能有效的区分不同的市场阶段。这对使用聚类方法来判断当前的市场状况有很大的指导意义。一般来说, 市场处于顶部和底部的时间在三个月左右, 这样, 我们就可以使用最近三个月的数据, 加入到指定时间段的时间序列中去进行聚类分析, 以判断当前时点市场所处的位置。当然, 这里面就涉及不同行业中不同市场状况下的表现, 这是本文下一部分所要研究的内容。对于指定时间段详细的分析结果, 可以参阅表2 (含行业类别Q型聚类结果分析)。

对行业类别的 Q 型聚类分析

采用指定时间段作为时间变量, 对同一时间段内的行业进行Q型聚类分析, 以便分析不同时间段各个行业是如何表现的, 它们的分类有没有稳定的关系。Q型聚类分析的结果如表2所示:

表 2: 对行业类别 Q 型聚类结果说明

事后描述	R 型聚类结果	起始时间	截止时间	一类行业	一类收益率	三类行业	三类收益率	二类行业	二类收益率	基准指数	一类-三类	二类-三类
上涨	五类	20000128	20010531	V	58.68%	ACDEFGHIJKLM NOPQRSTU	51.36%	B	79.42%	51.63%	7.33%	28.06%
顶部	一类	20010531	20010731	ACEFGHIJKL MNOSTUV W	-10.08%	DPQR	-7.88%	B	-6.25%	-9.19%	-2.21%	1.63%
下跌	二类	20010731	20031031	JV	-66.76%	ABCDEFGHIJKLM NOPQRSTUW	-54.79%	S	-26.55%	-56.07%	-11.97%	28.24%
上涨	一类	20031031	20040331	I	9.94%	ABCEFGHIJKLMN OPQRSTUWV	18.38%	D	37.00%	19.54%	-8.44%	18.62%
顶部	一类	20040331	20040531	B	-1.80%	ACDEFGHIJKLM NOPQRSTUWV	-8.60%	I	-22.37%	-8.77%	6.80%	-13.77%
下跌	二类	20040531	20050531	D	-21.12%	ACEFGHIJKLMN OPQRSTUWV	-42.18%	B	-20.23%	-42.53%	21.06%	21.95%
底部	一类	20050531	20051230	B	-1.05%	ACDEFGHIJKLM NOPRTUW	-0.12%	SV	10.80%	1.44%	-0.93%	10.91%
上涨	三类	20051230	20071031	E	269.82%	ACDFGHIJKLMN OPQRSTUWV	178.56%	B	229.71%	194.15%	91.25%	51.14%
顶部	一类	20071031	20071228	B	-25.49%	ACDEFGHIJKLM NOPQRSTUW	2.07%	RS	-4.54%	-1.12%	-27.56%	-6.60%

下跌	四类	20071228	20081031	B	-99.03%	ACDEFGHIJKLM NOPQRSTUW	-80.58%	S	-94.73%	-85.64%	-18.44%	-14.14%
底部	一类	20081031	20090123	KLOPRTV	4.41%	ABCDEFGHIJMN QSUW	4.05%	F	18.31%	1.50%	0.36%	14.26%
上涨 OR 拐 点?	一类	20090123	20090623	E	87.06%	ACDFGHIJKLMN OPQRSTUWV	53.32%	B	75.06%	56.59%	33.75%	21.74%

数据来源：国信证券经济研究所

行业代码说明：

A 农林牧渔；B 采掘；C 化工；D 黑色金属；E 有色金属；F 建筑建材；
G 机械设备；H 电子元器件；I 交运设备；J 信息设备；K 家用电器；
L 食品饮料；M 纺织服装；N 轻工制造；O 医药生物；P 公用事业；
Q 交通运输；R 房地产；S 金融服务；T 商业贸易；U 餐饮旅游；
V 信息服务；W 综合

仔细分析上表可以发现：

- 1、从2000年至今，一、二、三类的行业分类比较稳定，一、二类行业获取超额收益（下跌阶段为超额负收益），三类行业获取市场平均收益¹；
- 2、在市场处于重要拐点的时候，三类行业的收益率差异相对不明显，这也是以后使用Q型聚类分析市场是否处于拐点的一个重要方法；
- 3、获取超额收益的一二类行业比较稳定，主要有：采掘业、有色金属、黑色金属、金融服务、房地产、交运设备、信息服务和信息设备；
- 4、对于现阶段来说，时间变量的R型聚类将其归为一类（R型聚类），而一类时间变量（表2第二列）大部分是市场拐点，但是Q型聚类三类行业的收益率差异比较明显，目前还不满足拐点条件。

聚类后的 Granger 因果检验

Granger 因果检验的一个前提是要求涉及的时间序列为平稳序列，因此首先对上述四个时间序列进行单位根检验。为了保证检验的有效性，采用ADF单位根检验方法检验三类行业收益率的平稳性。

表 3：三类行业 ADF 单位根检验

序列名称	样本数	T 检验临界值 (5%显著性)	T 统计量	原假设	平稳性	是否差分
一类行业	24	-2.0687	-10.2373	接受	平稳	一阶差分
二类行业	24	-2.0687	-2.9176	接受	平稳	一阶差分
三类行业	24	-2.0687	-3.2786	接受	平稳	一阶差分

数据来源：国信证券经济研究所，20051230-20071031

由上表可知，三类行业的收益率一阶差分序列满足平稳性要求，可以做 Granger Causality 检验。检验结果如表

表 4：三类行业 Granger Causality 检验结果

P 统计值	滞后阶数				
	1	2	3	5	10
一类不影响二类	0.0579	0.0225	0.0109	0.0076	0.0008

¹ 使用 Wind 全 A 指数作为基准，三类行业的收益率为平均值，未加权处理。

二类不影响一类	1.09E-10	7.93E-20	1.45E-37	2.15E-43	1.44E-29
一类不影响三类	0.9268	0.6592	0.2402	0.01069	0.02521
三类不影响一类	2.40E-09	5.40E-60	6.40E-58	2.10E-54	4.90E-53
二类不影响三类	0.8737	0.7314	0.2011	0.00976	0.00764
三类不影响二类	3.40E-10	2.47E-30	4.53E-48	6.71E-44	4.50E-50

数据来源：国信证券经济研究所

检验结果表明，在牛市中三类行业收益率是一类和二类行业收益率的Granger原因，同时二类行业是一类行业的Granger原因，但没有三类行业对一类行业影响显著。三类行业对一类、二类行业的影响从1阶到5阶都比较显著，但从三阶开始，显著性有所降低。二类行业对一类行业的影响2阶以后显著性降低。也就是说，在上涨阶段，三类行业最先启动，一类、二类行业的滞后阶在三阶左右（三个月）。三类行业——>二类行业（3M）——>一类行业（2M）。

结论与建议

- 1、从2000年至今，一、二、三类的行业分类比较稳定，一、二类行业获取超额收益（下跌阶段为超额负收益），三类行业获取市场平均收益；
- 2、在市场处于重要拐点的时候，三大类行业的收益率差异相对不明显，这也是以后使用Q型聚类分析市场是否处于拐点的一个重要方法；
- 3、获取超额收益的一二类行业比较稳定，主要有：采掘业、有色金属、黑色金属、金融服务、房地产、交运设备、信息服务和信息设备；
- 4、对于现阶段来说，时间变量的R型聚类将其归为一类（R型聚类），而一类时间变量（表2第二列）大部分是市场拐点，但是Q型聚类三类行业的收益率差异比较明显，目前还不满足拐点条件；
- 5、在上涨阶段，三类行业最先启动，二类行业的滞后一类行业在三阶左右（三个月），一类行业之后二类行业中二阶左右（两个月）。三大类行业轮动的次序为：三类行业——>二类行业（3M）——>一类行业（2M）；
- 6、目前建议超配一类行业，即使采掘业、有色金属、黑色金属、金融服务、房地产、交运设备、信息服务和信息设备。但需密切关注三类行业收益率的差异化，当差异化逐渐降低的时候，应转向三大类行业。

国信证券投资评级

类别	级别	定义
股票 投资评级	推荐	预计 6 个月内，股价表现优于市场指数 20%以上
	谨慎推荐	预计 6 个月内，股价表现优于市场指数 10%-20%之间
	中性	预计 6 个月内，股价表现介于市场指数 $\pm 10\%$ 之间
	回避	预计 6 个月内，股价表现弱于市场指数 10%以上
行业 投资评级	推荐	预计 6 个月内，行业指数表现优于市场指数 10%以上
	谨慎推荐	预计 6 个月内，行业指数表现优于市场指数 5%-10% 之间
	中性	预计 6 个月内，行业指数表现介于市场指数 $\pm 5\%$ 之间
	回避	预计 6 个月内，行业指数表现弱于市场指数 5%以上

免责声明

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。本报告版权归国信证券所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。

国信证券经济研究所研究团队

宏观			策略			交通运输		
杨建龙			汤小生	021-68864595		唐建华	0755-82130468	
林松立	010-82254212		黄海培	021-68864598		孙菲菲	0755-82133400	
任泽平	010-82254206		崔嵘	021-68866202		高健	0755-82130678	
周炳林	0755-82133339		廖喆	021-68866236		黄金香	010-82252922	
			赵谦	021-68866295				
银行			房地产			计算机		
邱志承	021-68864597		方焱	0755-82130648		凌晨	021-68866233	
黄飙	0755-82133476		陈林	0755-82133343		电子元器件		
谈煜	010-82254212		区瑞明	0755-82130678		王俊峰	010-82254205	
戴志锋	0755-82133343							
有色金属			汽车及零配件			钢铁		
彭波	0755-82133909		李君	021-68866235		郑东	010-82254160	
李洪翼	010-82252922		左涛	021-68866253		秦波	010-82252922	
商业贸易			基础化工			非银行金融		
胡鸿轲	021-68866206		邱伟	0755-82133263		武建刚	010-82250828	
吴美玉	010-82252911		陆震	0755-82130532		王一峰	010-82250828	
石油与石化			电力设备			传媒		
李晨	021-68866252		彭继忠	021-68866203		陈财茂	021-68866236	
严蓓娜	021-68866253		皮家银	021-68866205		廖绪发	021-68866237	
机械			电力与公用事业			医药		
余爱斌	0755-82133400		徐颖真	021-68864007		贺平鸽	0755-82133396	
李筱筠	010-82254205		谢达成	021-68866236		丁丹	0755-82130678	
通信			造纸			家电		
严平	021-68865025		李世新	0755-82130565		王念春	0755-82130407	
纺织服装			建材			旅游		
方军平	021-68866202		杨昕	021-68864596		陈财茂	021-68866236	
						廖绪发	021-68866237	
中小股票			食品饮料					
高芳敏	021-68864586		黄茂	0755-82133476				
陈爱华			谢鸿鹤	0755-82130646				
农业			煤炭			建筑		
张如	021-68866233		李然	0755-82130681		邱波	0755-82133390	
固定收益与基金			金融工程			金融工程		
皮敏	021-68864011		葛新元	0755-82133332		王军清	0755-82133297	
张旭	010-82254210		董艺婷	021-68866946		黄志文	0755-82130532	
杨涛	0755-82133339		戴军	021-68864585		林晓明		

国信证券经济研究所机构销售部

华南区			华东区			华北区		
万成水	0755-82133147 13923401205 wancs@guosen.com.cn		盛建平	021-68864592 15821778133 shengjp@guosen.com.cn		王立法	010-82252236 13910524551 wanglf@guosen.com.cn	
刘宇华	0755-82130818 13823380182 liuyh@guosen.com.cn		马小丹	021-68866025 13801832154 maxd@guosen.com.cn		王晓建	010-82252615 13701099132 wangxj@guosen.com.cn	
邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn		郑毅	021-68866205 13795229060 zhengyi@guosen.com.cn		谭春元	010-82254209 13810118116 tancy@guosen.com.cn	
庞博	0755-82130818 0755-82133343 pangb@guosen.com.cn		黄胜蓝	021-68866011 13761873797 huangsl@guosen.com.cn		焦骥	010-82254209 13601094018 jiaojian@guosen.com.cn	
祝彬	0755-82133456 15814403667 zhubin@guosen.com.cn		刘塑	021-68866236 13817906789 liusu@guosen.com.cn				