

## 专题报告

## 金融工程

## 程序化交易

## 程序化交易专题报告系列之二

2010年04月06日

本报告的独到之处

- 利用高频数据挖掘投资机会
- 应用高频数据并接合模式识别方法能够获得相对市场更为稳健超额收益。收益波动相对较低

相关研究报告:

程序化交易专题报告系列之一: 程序化交易在股指期货中的应用

分析师 黄志文  
电话: 0755-82133928  
E-mail: huangzw@guosen.com.cn

分析师 葛新元  
电话: 0755-82130833-1870  
E-mail: gexy@guosen.com.cn

联系人 戴 军  
电话: 021-68864585  
E-mail: daijun@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。

## 专题报告

## 在高频数据中挖掘交易机会

通常来说,证券市场上发生的信息必将影响证券价格的变化,而高频数据记录了实时信息对证券价格最终影响的所有信息。相对低频数据而言能够更加准确捕捉到证券市场中发生的每个细微变化过程。故利用高频数据进行证券投资分析较用低频数据更具有优势。

已有大量的研究成果发现高频数据存在显著聚类特性,因此我们考虑尝试利用高频数据的聚类性特征结合模式识别技术对证券的短期运行趋势进行预测。

模式识别在证券投资中的运用 具体而言,是将时间序列数据挖掘应用于证券行情的高频数据中,其目标就是寻找一种交易优势,一种能够获得超越市场基准收益率一种交易优势。也就是识别或发现能够预测证券价格上涨或下跌超过一定情况的模式聚类。

**主要逻辑:** 证券的走势其实就是买卖双方对于其价格达成一致的表现,通过对其历史走势的研究可以找到相似的形态从而以此判断未来价格运动的方向。模式识别的主要逻辑就是基于此,即历史会重现,通过对证券过去走势的分析预测其未来走向。

**结论以及应用:** 由于大盘蓝筹股公司运作相对规范、受市场关注度相对较高,受市场因素影响相对较少,股价运行趋势也相对平稳。因此使得利用高频数据并接合模式识别技术对大盘股的短期预测有效性较高,运用模式识别技术针对大盘股的高频数据进行分析对机构者选股可以起到更为积极的作用。

我们通过样本外检验可以看出:应用高频数据并接合模式识别方法能够获得相对市场更为稳健超额收益。收益波动相对较低。

未来应用设想 经典图库法:我们可以利用高频数据与历史市场表现建立经典图库,然后运用图库中各种经典走势去识别当前标的证券市场走势,以此方法来捕捉投资机会,来获得超额收益。

## 目 录

高频数据与投资机会 .....	3
运用模式识别高频数据中潜藏的机会 .....	3
具体步骤 .....	3
具体分析框架 .....	4
模式识别选股实证分析 .....	4
主要逻辑 .....	4
模式识别的具体方法 .....	4
模式识别的样本选择 .....	5
实证分析 .....	6
结论以及应用 .....	9

## 图表目录

图 1、高频行情模式识别分析框架图 .....	4
表 1: 样品股明细 .....	5
图 2: 浦发银行: 模拟收益对比图 .....	6
表 2 浦发银行模拟收益—样本内检验 .....	6
图 3: 深发展-模拟收益对比图 .....	7
表 3: 深发展---模拟收益（样本内检验） .....	7
图 5: 民生银行: 模拟收益对比图 .....	8
图 6: 北京银行: 模拟收益对比图 .....	8

## 高频数据与投资机会

通常来说,证券市场上发生的信息必将影响证券价格的变化,而高频数据记录了实时信息对证券价格最终影响的所有信息。相对低频数据而言能够更加准确捕捉到证券市场中发生的每个细微变化过程。故利用高频数据进行证券投资分析较用低频数据更具有优势。

已有大量的研究成果发现高频数据存在显著聚类特性,因此我们考虑尝试利用高频数据的聚类性特征结合模式识别技术对证券的短期运行趋势进行预测。

高频时间序列数据聚类性是指某一类统计特征的持续性。在高频数据的波动性研究中,聚类性体现在大的波动性后面往往还会出现大的波动性。在超高频数据的时间间隔序列中,聚类性体现在长的时间间隔后面往往跟随着长的时间间隔。

因此我们可以利用模式识别方法预测标的证券未来一段时点可能出现运行趋势。以获得超越市场基准收益率。

## 运用模式识别高频数据中潜藏的机会

模式识别是指对表征事物或现象的各种形式的信息进行处理和分析,以对事物或现象进行描述、辨认、分类和解释的过程,是信息科学和人工智能的重要组成部分。模式识别又常称作模式分类

模式识别方法之一就是决策理论方法,是发展较早也比较成熟的一种方法。被识别对象首先数字化,变换为适于计算机处理的数字信息。一个模式常常要用很大的信息量来表示。许多模式识别系统在数字化环节之后还进行预处理,用于除去混入的干扰信息并减少某些变形和失真。随后是进行特征抽取,从数字化后或预处理后的输入模式中抽取一组特征。其实特征是选定的一种度量。特征抽取后可进行分类,即从特征空间再映射到决策空间。为此而引入鉴别函数,由特征矢量计算出相应于各类别的鉴别函数值,通过鉴别函数值的比较实行分类。

**模式识别在证券投资中的运用** 具体而言,是将时间序列数据挖掘方法应于证券行情的高频数据中,是利用证券过去走势识别现在的走势来预测未来走势。其目标就是寻找一种交易优势,一种能够获得超越市场基准收益率一种交易优势。也就是识别或发现能够预测证券价格上涨或下跌超过一定情况的模式聚类。

### 具体步骤

- 在给定的观察历史高频行情时间序列 $X=\{X(t), t=1,2,3,\dots, N\}$ ,并过滤掉一些异常数据。
- 对预处理的历史高频行情序列聚成为4类( $x_1$ --上涨、 $x_2$ --平衡、 $x_3$ --下跌、 $x_4$ --

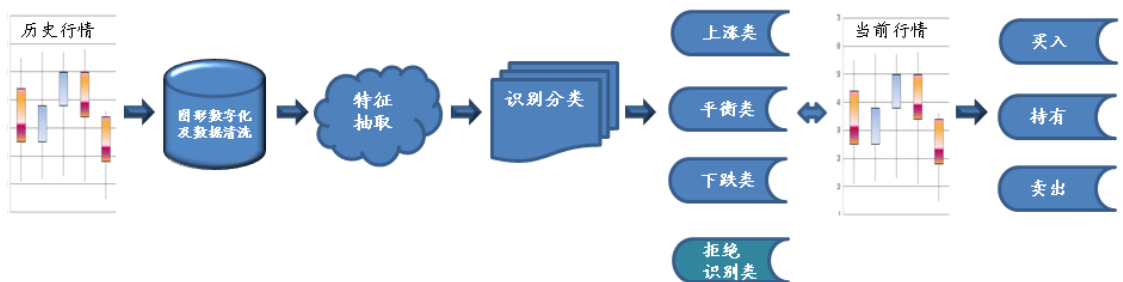
拒收识别类)。

$$\begin{aligned} X1 &= \{X1(t), t=1, 2, 3, \dots, N\}, & X2 &= \{X2(t), t=1, 2, 3, \dots, N\}, \\ X3 &= \{X3(t), t=1, 2, 3, \dots, N\}, & x4 &= \{x4(t), t=1, 2, 3, \dots, N\} \end{aligned}$$

- 训练识别样品高频行情数据，并评估训练阶段的结果，反复进行训练优化系统参数。
- 对于给定检验高频行情时间序列， $Y = \{Y(t), t=R, \dots, W\}$   $N < R < W$ ，运用在X中发现的模式来预测Y中的事件（上涨、平衡、下跌），以此辅助投资者做出正确的投资决策。

## 具体分析框架

图 1、高频行情模式识别分析框架图



## 模式识别选股实证分析

### 主要逻辑

证券的走势其实就是买卖双方对于其价格达成一致的表现，通过对其历史走势的研究可以找到相似的形态从而以此判断未来价格运动的方向。模式识别的主要逻辑就是基于此，即历史会重现，通过对证券过去走势的分析预测其未来走向。

### 模式识别的具体方法

我们通过对样本证券高频数据抽取频率及每天的时间周期模拟结果表明，时间周期以每一天数据抽取频率为1分钟识别效果较为显著，因此我们运用行情数据频率为1分钟线，选取每天行情（一共240个数据点）。根据标的证券每天行情走势图与此标的证券历史行情走势图进行对比，找出最为相似的走势图，然后根据历史上最为相似的那天的N日后涨跌幅来决定买卖，再加上止损策略。

具体的操作方法如下：

- 1、分类历史高频行情数据。
- 2、识别当前行情与历史行情走势图的相似程度是否达到C。
- 3、当满足条件1时且N天后涨幅超过R1时则作买入策略。
- 4、当满足条件1时且N天后跌幅超过R2则作入卖出策略。
- 5、当以上三个条件不满足时，根据目前持有状况作出：

（1）当持有时：加上止损策略（每次买入后持有期亏损达到X作出卖出，否则持有；）

（2）空仓时：不作任何操作，继续空仓。

### 模式识别的样本选择

通常来说，市场投资者与各研究机构对大公司跟踪密切程度相对小公司来说要高得多。研究得也更加全面透彻。同时大公司交易量也相对较大，信息披露也更加充分及时，因此大公司的信息不对称程度较低，受市场操控可能性较小，公司要低得多。因此投资者对大公司内在价值认识的不确定性较低，交易能够反映其内在价值的信息，从而减少投资者对其内在价值认识的不确定性。

同时大公司流动性也要相对强得多，不同投资者拥有的信息就越有可能融入到股票价格中，从而降低了信息不对称性和投资者对资产内在价值认识的不确定性。

考虑蓝筹公司透明度与市场关注高的原因，以及较其他一些股票波动性相对较低、突变性特征较小，因此我们首先选择沪深交易所的银行股为样本进行识别结果分析。

**表 1：样本股明细**

证券代码	证券简称	首发上市日期
000001.SZ	深发展 A	1991-04-03
002142.SZ	宁波银行	2007-07-19
600000.SH	浦发银行	1999-11-10
600015.SH	华夏银行	2003-09-12
600016.SH	民生银行	2000-12-19
600036.SH	招商银行	2002-04-09
601009.SH	南京银行	2007-07-19
601166.SH	兴业银行	2007-02-05
601169.SH	北京银行	2007-09-19
601328.SH	交通银行	2007-05-15
601398.SH	工商银行	2006-10-27
601939.SH	建设银行	2007-09-25
601988.SH	中国银行	2006-07-05
601998.SH	中信银行	2007-04-27

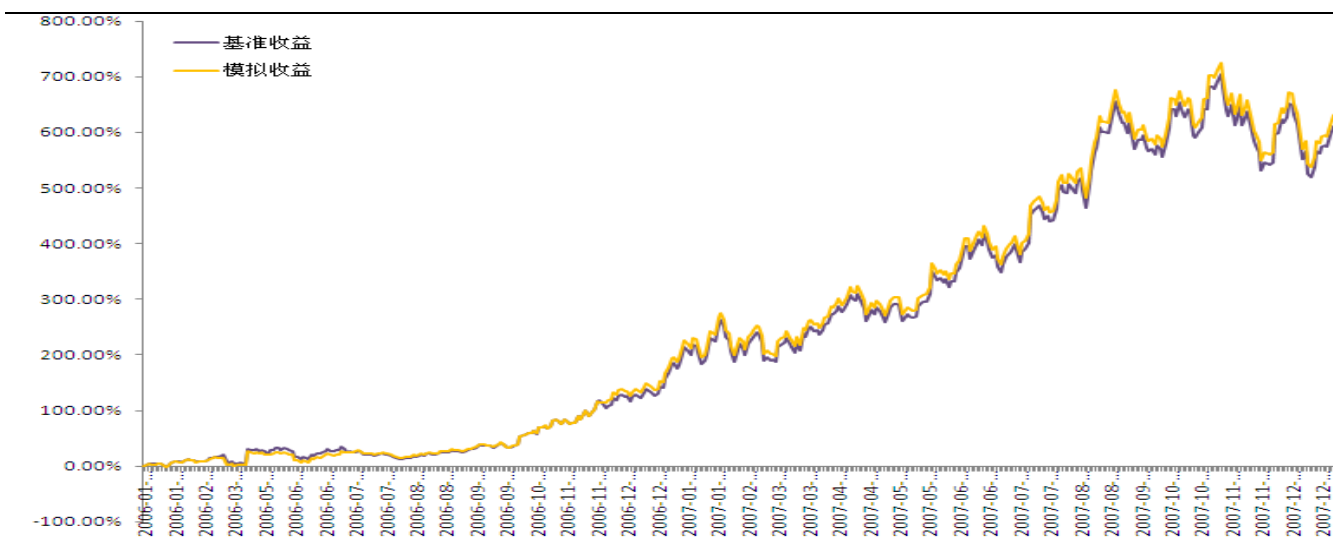
数据来源：国信证券经济研究所

## 实证分析

### 样本内检验及参数优化

考虑银行板块中浦发银行与深发展上市时间较长，样本数据较多一些，因此我们选择浦发银行作参数数据优化。深发展的样本中内检验。选取时间为：2006年至2007年高频行情（2005年1月1日至2005年12月31日为历史样品，模拟从2006年1月1日到2007年12月31日）。交易成本为单边千分之三。

图 2：浦发银行：模拟收益对比图



数据来源：国信证券经济研究所，

通过样本内模拟整体总收益率超越基准 19.35%，总发生买卖次数为 9 次。由于 2006 至 2007 是市场历史从未有过的单边上涨阶段。所以总体超越基准收益较低。

表 2 浦发银行模拟收益—样本内检验

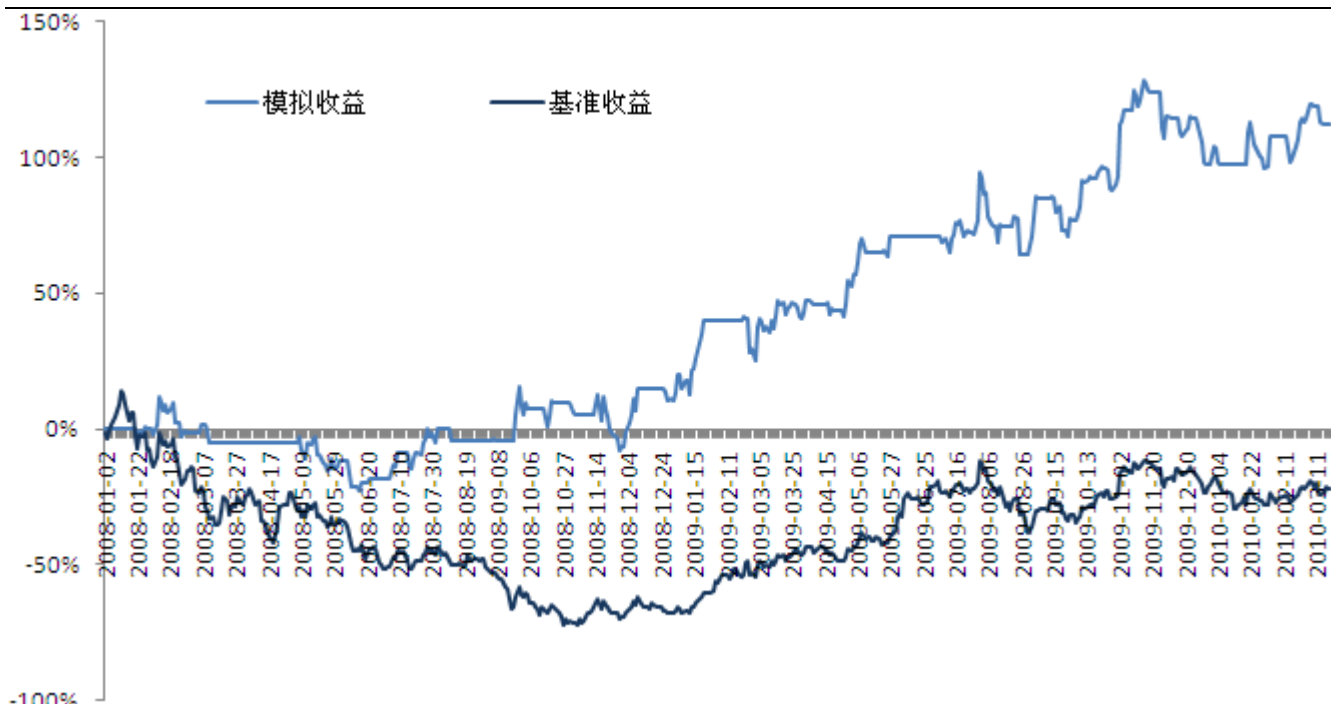
	基准收益	模拟收益	超越基准	买卖信号
2006	82.76%	93.59%	10.83%	9
2007	148.79%	148.79%	0.00%	0
全段	611.10%	630.45	19.35%	9

数据来源：国信证券经济研究所

我们通过浦发银行在 2006 至 2007 年样本模拟出模型参数，以此参数计算在一个完整的市场周期内检验深发展在 2008 年 1 月 1 日到 2010 年 3 月 20 日内

模拟收益为 112.25%。超越买入持有策略 134.39%。并且收益波动幅度也低于买入持有策略。但交易的次数较多，总体交易了 99 次。

图 3: 深发展-模拟收益对比图



数据来源：国信证券经济研究所，

表 3: 深发展---模拟收益（样本内检验）

	基准收益	模拟收益	超越基准	买卖信号
2008	-66.42%	10.50%	76.92%	43
2009	49.54%	90.10%	40.56%	49
2010	-2.26%	14.25%	16.51%	7
全段	-22.14%	112.25%	134.39%	99

数据来源：国信证券经济研究所

## 样本外分析

以选取最优参数模拟银行股板块中所有个股，模拟时间区间为：2008 年 1 月 1 日至 2010 年 3 月 20 日。交易费用单边为千分之三。

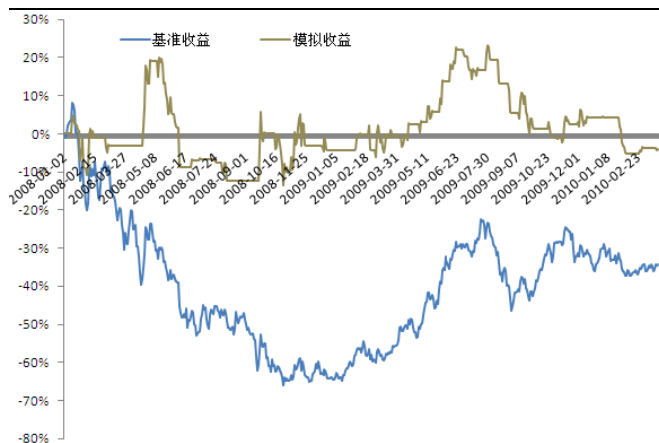
在一个较为完整市场周期中(20080101至20100320)模拟出银行股板块平均收益率超越买入持有策略25.29%。波动幅度远低于买入持有策略。持有策略收益标准差平均值为0.1617，而模拟收益率标准差的平均值为0.113。由此可以看出采用这种策略可获得更为稳健的相对收益。



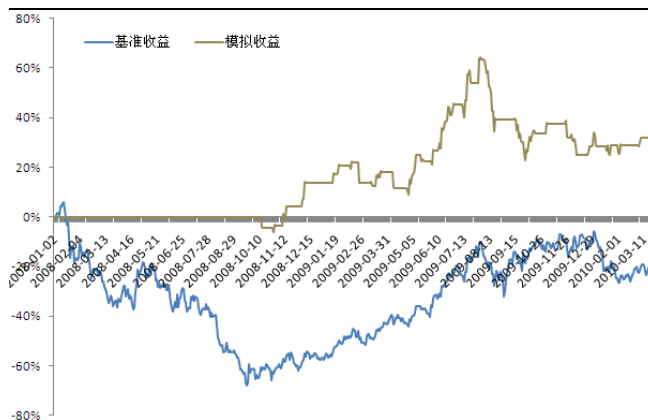
**表 4: 深沪银行股模拟收益率 (20080101 至 20100320)**

代码	简称	基准收益率	模拟收益率	超额收益	基准收益标准差	模拟收益标准差
600000	浦发银行	-23.95%	-34.23%	-10.28%	0.1849	0.1501
600015	华夏银行	-37.26%	-39.71%	-2.45%	0.1533	0.0790
600016	民生银行	-33.56%	-4.04%	29.52%	0.1600	0.0839
600036	招商银行	-45.41%	-24.79%	20.62%	0.1745	0.0914
601009	南京银行	-6.49%	-0.78%	5.71%	0.1969	0.0741
601166	兴业银行	-29.50%	-24.97%	4.53%	0.1973	0.1438
601169	北京银行	-17.17%	31.69%	48.86%	0.1735	0.1711
601328	交通银行	-44.75%	-10.50%	34.25%	0.1487	0.1204
601398	工商银行	-36.51%	-31.28%	5.23%	0.1187	0.0904
601939	建设银行	-40.51%	-32.26%	8.25%	0.1337	0.1093
601988	中国银行	-33.93%	-19.09%	14.84%	0.1172	0.0847
601998	中信银行	-28.18%	-22.82%	5.36%	0.1458	0.0731
000001	深发展 A	-22.14%	112.25%	134.39%	0.1895	0.4488
002142	宁波银行	-25.37%	29.89%	55.26%	0.1704	0.1419
<b>平均</b>		<b>-30.34%</b>	<b>-5.05%</b>	<b>25.29%</b>	<b>0.1617</b>	<b>0.1330</b>

数据来源: 国信证券经济研究所

**图 5: 民生银行: 模拟收益对比图**


数据来源: 国信证券经济研究所,

**图 6: 北京银行: 模拟收益对比图**


数据来源: 国信证券经济研究所,



## 结论以及应用

由于大盘蓝筹股公司运作相对规范、受市场关注度相对较高,受市场因素影响相对较少,股价运行趋势也相对平稳。因此使得利用高频数据并接合模式识别技术对大盘股的短期预测有效性较高,运用模式识别技术针对大盘股的高频数据进行分析对机构者选股可以起到更为积极的作用。

我们通过样本外检验可以看出:应用高频数据并接合模式识别方法能够获得相对市场更为稳健超额收益。收益波动相对较低。

未来应用设想 经典图库法:我们可以利用高频数据与历史市场表现建立经典图库,然后运用图库中各种经典走势去识别当前标的证券市场走势,以此方法来捕捉投资机会,来获得超额收益。

### 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

<b>宏观</b>		<b>策略</b>		<b>交通运输</b>	
周炳林	0755-82133339	赵 谦	021-60933153	郑 武	0755- 82130422
林松立	010-82254212	崔 嵘	021-60933159	陈建生	0755- 82130422
		廖 喆	021-60933162	岳 鑫	0755- 82130422
		黄学军	021-60933142	高 健	0755-82130678
<b>银行</b>		<b>房地产</b>		<b>机械</b>	
邱志承	021-68864597	方 焱	0755-82130648	余爱斌	0755-82133400
黄 飙	0755-82133476	区瑞明	0755-82130678	李筱筠	010-82254205
谈 焯	010- 82254212	黄道立	0755-82130833	黄海培	021-60933150
				陈 玲	0755-82133400
				杨 森	0755-82133343
<b>汽车及零配件</b>		<b>钢铁</b>		<b>商业贸易</b>	
李 君	021-60933156	郑 东	010-82254160	孙菲菲	0755-82133400
左 涛	021-60933164	秦 波	010-82252922	吴美玉	010-82252911
<b>基础化工</b>		<b>医药</b>		<b>石油与石化</b>	
邱 伟	0755-82133263	贺平鸽	0755-82133396	李 晨	021-60875160
张栋梁	0755-82130532	丁 丹	0755-82130678	严蓓娜	021-60933165
		陈 栋	021-60933147		
<b>电力设备与新能源</b>		<b>传媒</b>		<b>有色金属</b>	
皮家银	021-60933160	陈财茂	021-60933163	彭 波	0755-82133909
<b>电力与公用事业</b>		<b>非银行金融</b>		<b>通信</b>	
徐颖真	021-60875162	邵子钦	0755- 82130468	严 平	021-60875165
谢达成	021-60933161	田 良	0755-82130513	程 峰	021-60933167
		童成敦	0755-82130513		
<b>造纸</b>		<b>家电</b>		<b>计算机</b>	
李世新	0755-82130565	王念春	0755-82130407	凌 晨	021-60933157
<b>电子元器件</b>		<b>纺织服装</b>		<b>农业</b>	
		方军平	021-60933158	张 如	021-60933151
<b>旅游</b>		<b>食品饮料</b>		<b>建材</b>	
廖绪发	021-60875168	黄 茂	0755-82133476	杨 昕	021-60933168
刘智景	021-60933148	谢鸿鹤	0755-82130646		
<b>煤炭</b>		<b>建筑</b>		<b>中小股票</b>	
李 然	0755-82130681	邱 波	0755-82133390	陈 健	0755-82133476
苏绍许	021-60933144	李遵庆	0755-82133343	陈爱华	0755-82133397
				祝 彬	0755-82132518
				王一峰	010-82250828
				邵 达	0755- 82132098
<b>固定收益</b>		<b>投资基金</b>		<b>量化投资</b>	
李怀定	021-60933152	杨 涛	0755-82133339	葛新元	0755-82133332
张 旭	010-82254210	黄志文	0755-82133928	董艺婷	021-60933155
高 宇	0755-82133528	刘舒宇	0755-82131822	戴 军	021-60933166
蔺晓熠	021-60933146	彭怡萍		林晓明	021-60933154
侯慧娣	021-60875161			秦国文	0755-82133528
				程景佳	021- 60933166
				赵斯尘	021- 60875174
<b>指数与产品设计</b>					
焦 健	0755-82131822				
赵学昂	0755-82131822				
王军清	0755-82133297				
阳 瑾	0755-82131822				
周 琦	0755-82131822				
彭甘霖					

**国信证券机构销售团队**

华南区			华东区			华北区		
万成水	0755-82133147 13923401205 wancs@guosen.com.cn		盛建平	021-60875169 15821778133 shengjp@guosen.com.cn		王立法	010-82252236 13910524551 wanglf@guosen.com.cn	
邵燕芳	0755-82133148 13480668226 shaoyf@guosen.com.cn		马小丹	021-60875172 13801832154 maxd@guosen.com.cn		王晓建	010-82252615 13701099132 wangxj@guosen.com.cn	
林 莉	0755- 82133197 13824397011 Linli2@guosen.com.cn		郑 毅	021-60875171 13795229060 zhengyi@guosen.com.cn		谭春元	010-82254209 13810118116 tancy@guosen.com.cn	
王昊文	0755-82130818 18925287888 wanghaow@guosen.com.cn		黄胜蓝	021-60875173 13761873797 huangsl@guosen.com.cn		焦 戢	010-82254202 13601094018 jiaojian@guosen.com.cn	
甘 墨	0755-82133456 15013851021 ganmo@guosen.com.cn		刘 塑	021-60875177 13817906789 liusu@guosen.com.cn		李锐	010-82254212 13691229417 lirui2@guosen.com.cn	
			叶琳菲	021-60875178 13817758288 yelf@guosen.com.cn		徐文琪	010-82254210 13811271758 xuwq@guosen.com.cn	
			许娅	021- 60875176 13482495069				
			江智俊	021-60875175 15221772073				
			孔华强	021-60875170 13681669123				