



证券研究报告

# 基于模式聚类的短线选股模型

国信证券经济研究所 金融工程部  
量化投资策略小组

证券分析师 董艺婷  
证券投资咨询执业资格证书编码: S0980510120055  
联系人 郑亚斌  
2011.9.13



**国信证券经济研究所**

Guosen Securities Economic Research Institute

# 目录



**模式聚类的提出**



**基于模式聚类的短线选股模型**



**成交价格、成交量模式组合**



**模型应用分析**



**应用和扩展**

# 研究框架

## 问题的提出？

模式是否存在？

模式是否可以预测？

技术分析理论：历史会重演

技术分析派：量价关系

## 模式聚类

成交价格模式

成交量模式

量价模式组合会发生什么？

## 模式组合

Top模式组合

Bottom模式组合

# 目录



模式聚类的提出



**基于模式聚类的短线选股模型**



成交价格、成交量模式组合



模型应用分析



应用和扩展



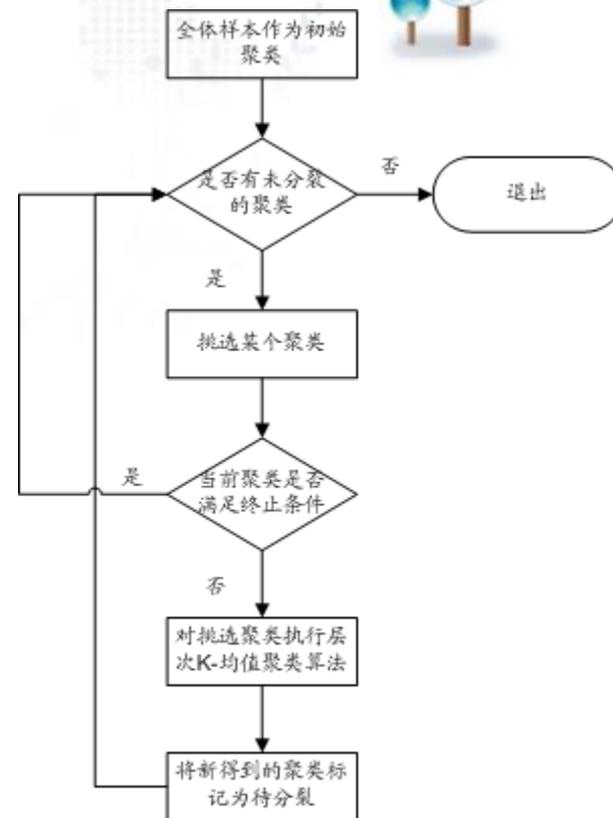
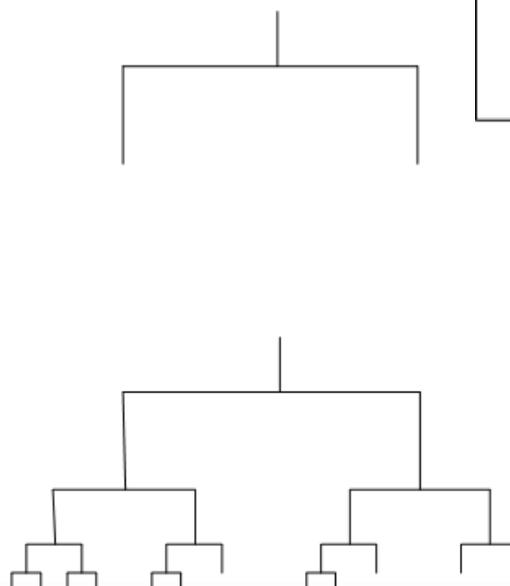
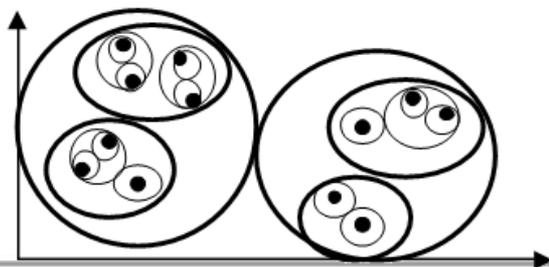
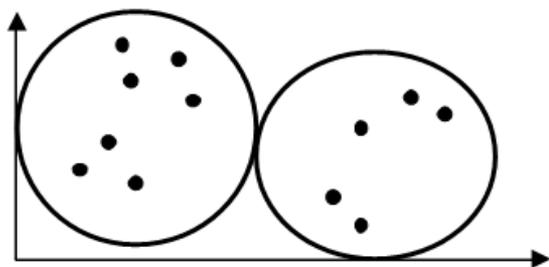
# 层次K-均值聚类算法



K-均值聚类算法：

1. 选择K个聚类中心；
2. 将样本划分到与其最近的聚类中心；
3. 更新聚类中心完成一次迭代；
4. 重复第2, 3步直至收敛。

层次K-均值聚类算法：每一层执行K-均值聚类

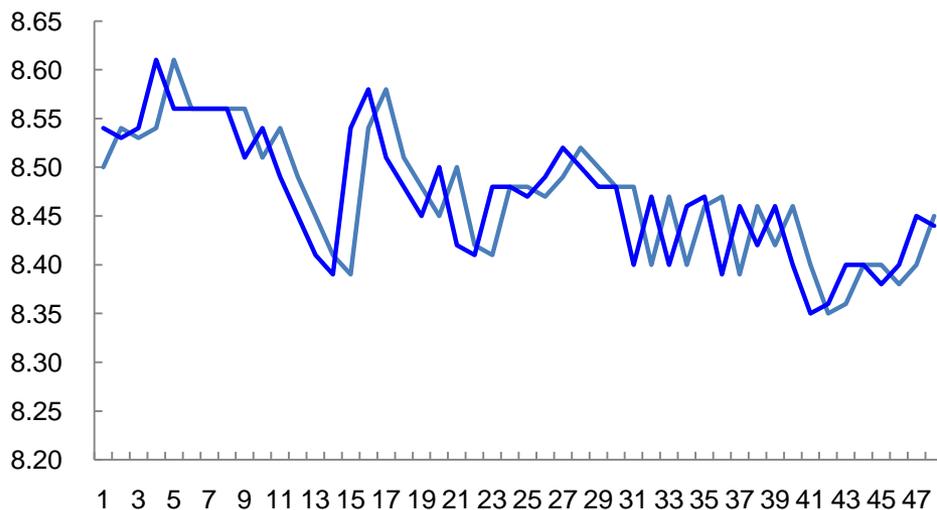


资料来源：Google图库，国信证券经济研究所整理

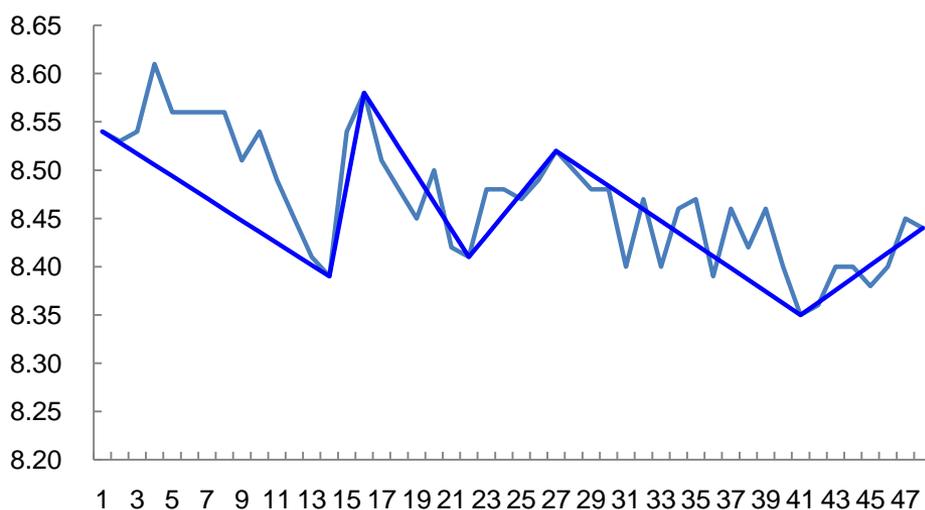
# 股价和成交量模式的处理

- 中证800历史成分股，2005.3.28 —— 2011.3.14
- 三个交易日内15分钟线数据
- 以第二个交易日开盘后1小时VWAP买入持有3-10个小时
- 聚类算法中的“距离”定义：时间序列关键点提取

— 时间序列a — 时间序列b



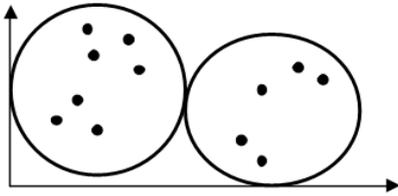
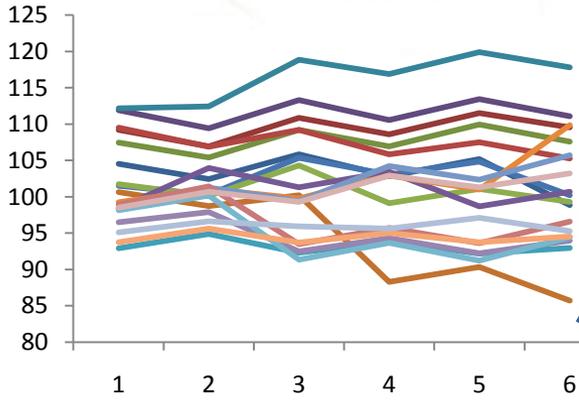
— 原始时间序列 — 折线段拟合时间序列



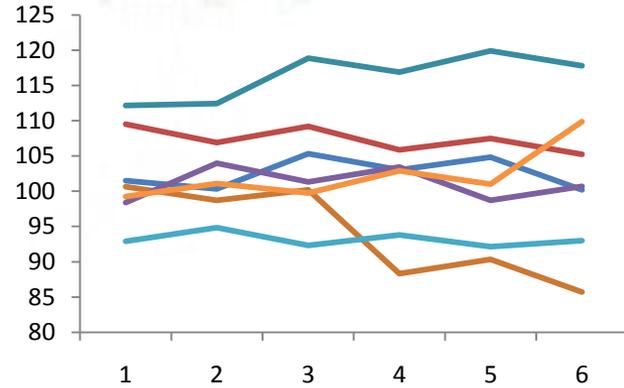
资料来源：天软科技，国信证券经济研究所整理

# 成交价格、成交量模式聚类

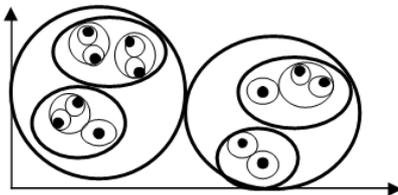
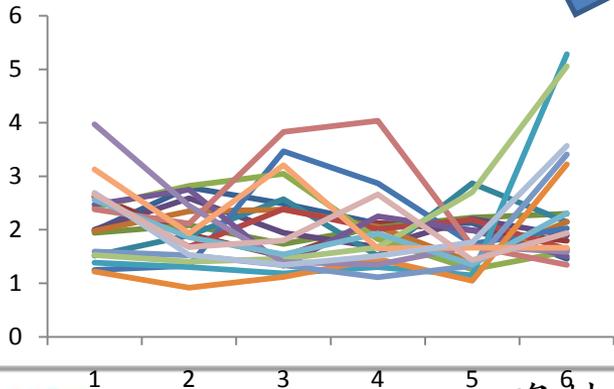
## 成交价格样本



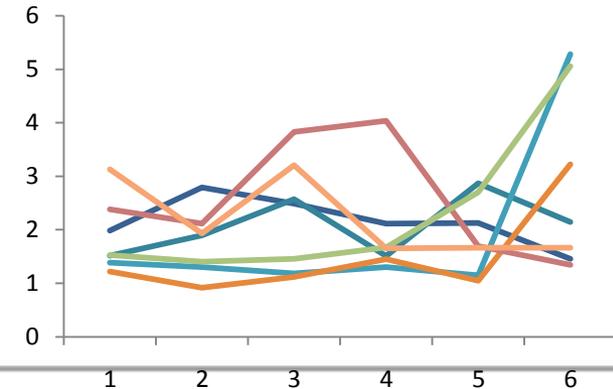
## 成交价格模式



## 成交量样本

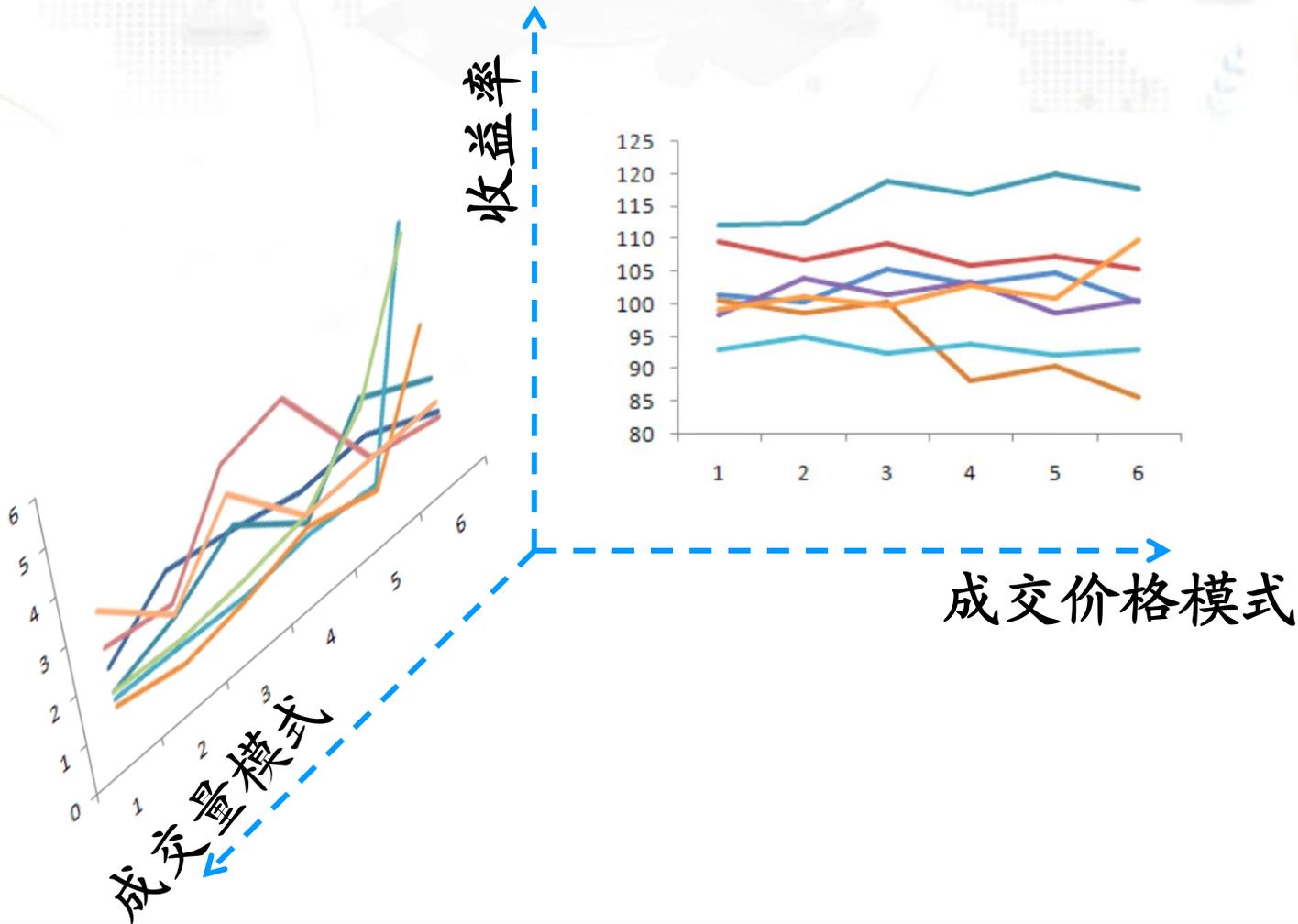


## 成交量模式



资料来源：国信证券经济研究所

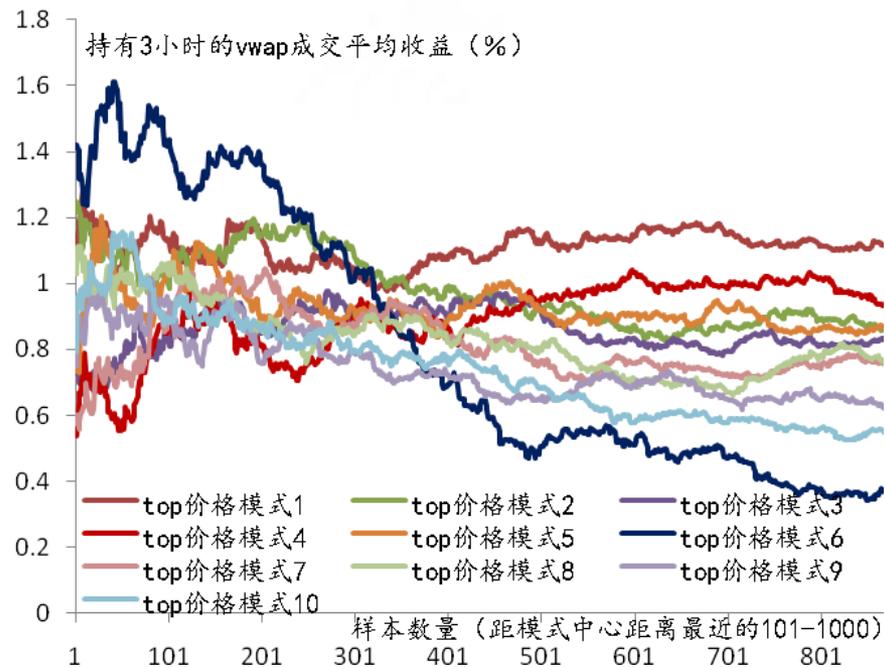
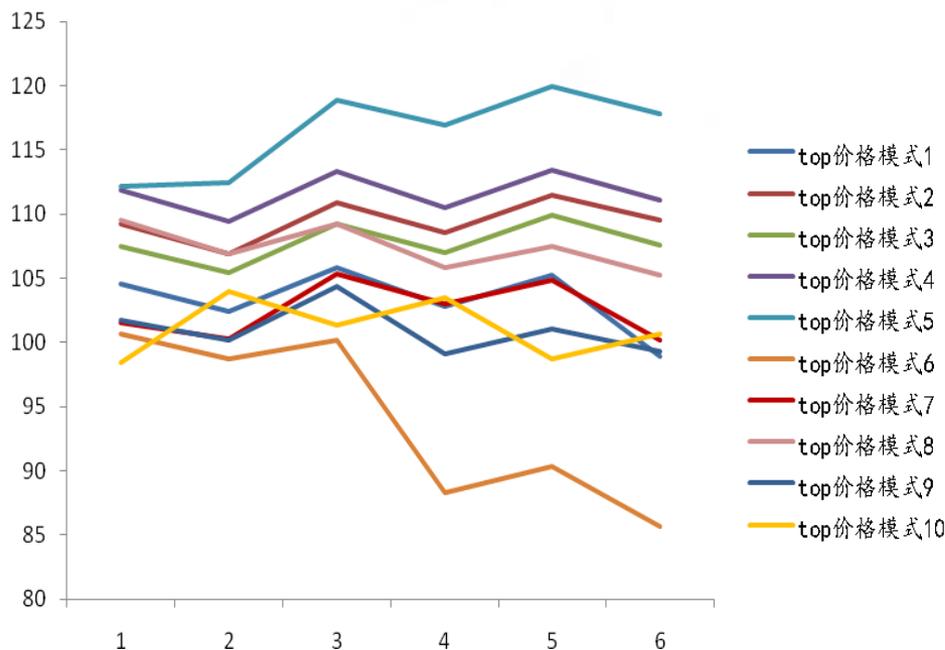
# 成交价格、成交量模式聚类分析



资料来源：国信证券经济研究所

# 股价模式中心的聚类分析

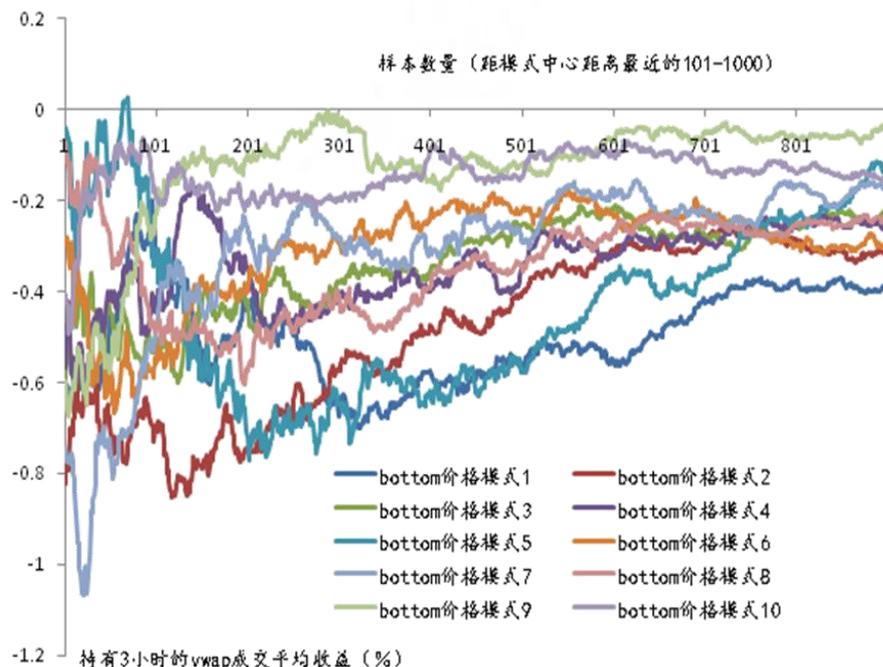
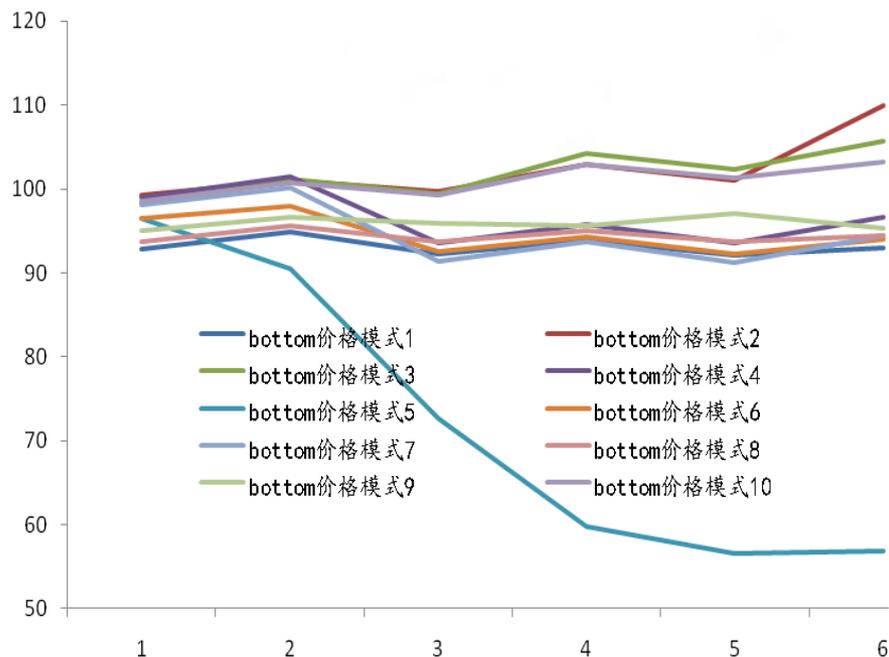
- 198个价格聚类中心（成交价格的不同模式）
- 预期上涨的前10种成交价格模式：



资料来源：天软科技，国信证券经济研究所整理

# 股价模式中心的聚类分析

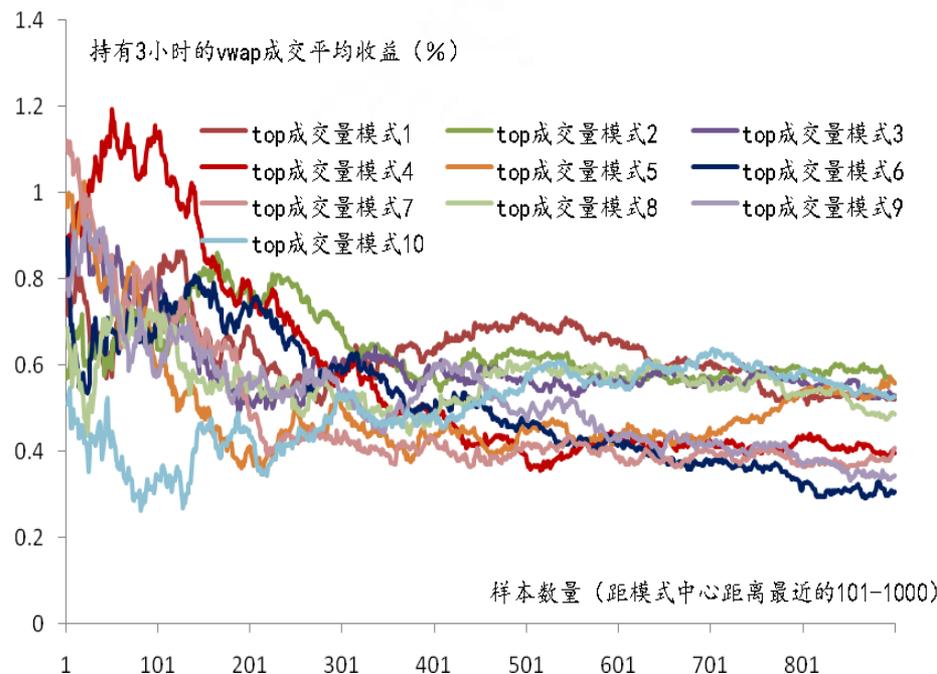
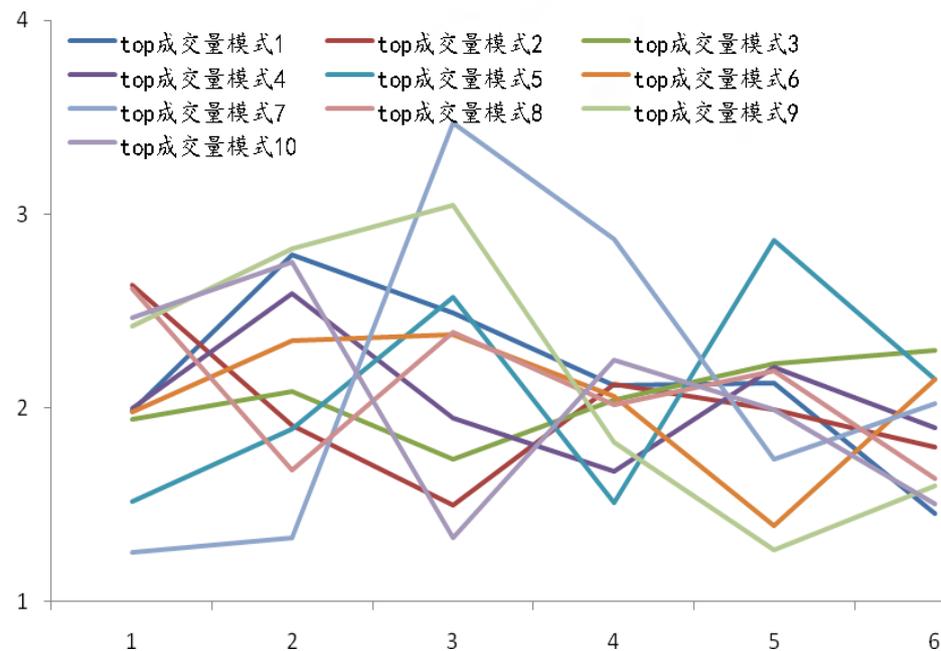
- 预期下跌的前10种成交价格模式:



资料来源: 天软科技, 国信证券经济研究所整理

# 成交量模式中心的聚类分析

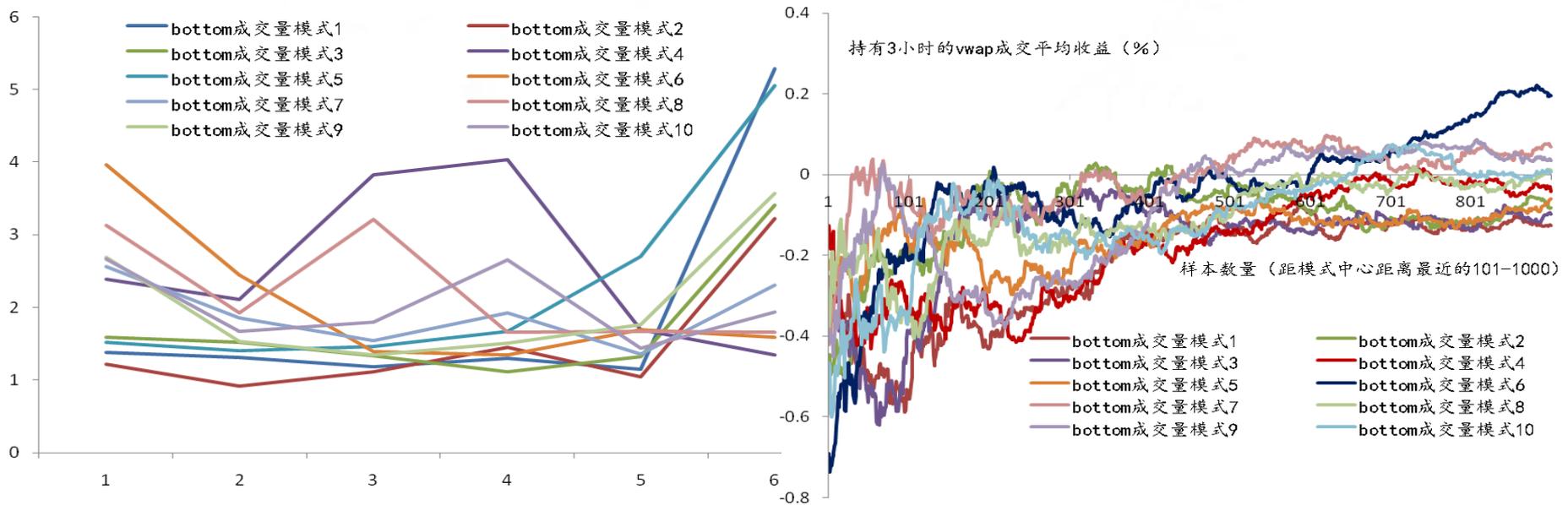
- 194个价格聚类中心（成交量的不同模式）
- 预期上涨的前10种成交量模式：



资料来源：天软科技，国信证券经济研究所整理

# 成交量模式中心的聚类分析

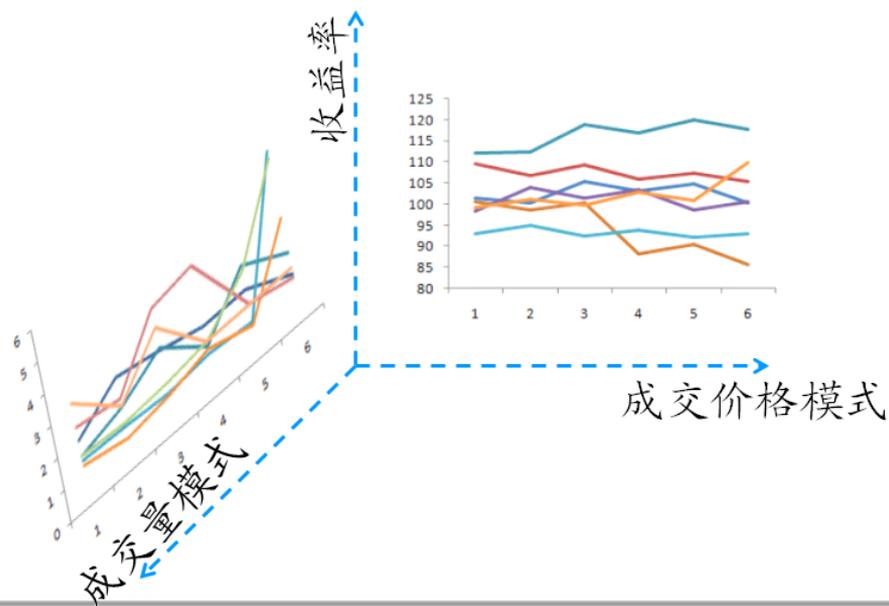
- 预期下跌的前10种成交量模式:



资料来源：天软科技，国信证券经济研究所整理

# 成交价格、成交量模式聚类分析

- 不同的成交价格、成交量模式预测效果不同
- 预期上涨的价格模式中心，收益普遍较为稳定
- 预期上涨的成交量模式稳定性次之
- 预期下跌形态的价格、成交量模式中心解释力相对较弱



# 目录



模式聚类的提出



基于模式聚类的短线选股模型



**成交价格、成交量模式组合**

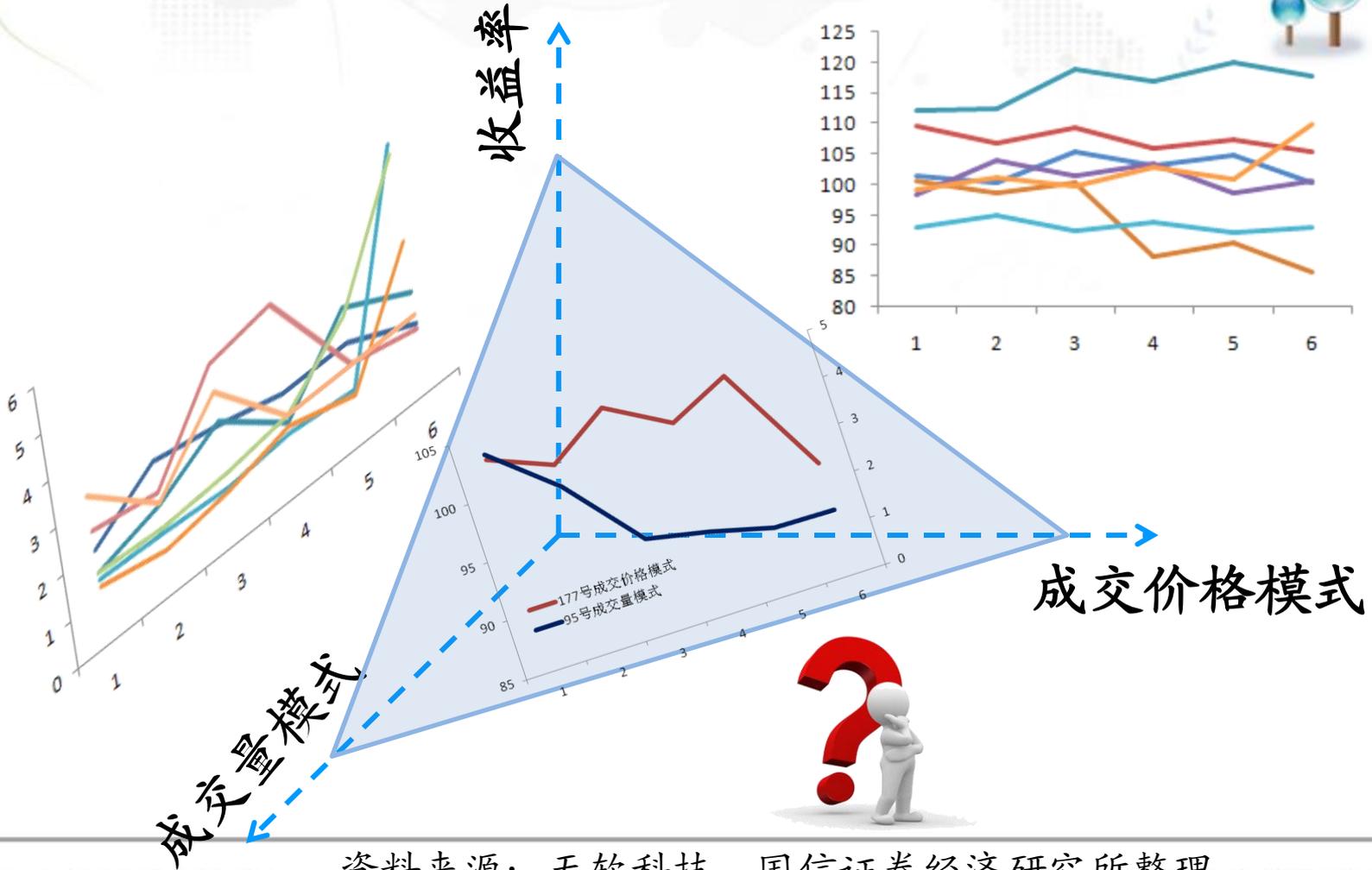


模型应用分析



应用和扩展

# 成交价格、成交量模式组合



资料来源：天软科技，国信证券经济研究所整理

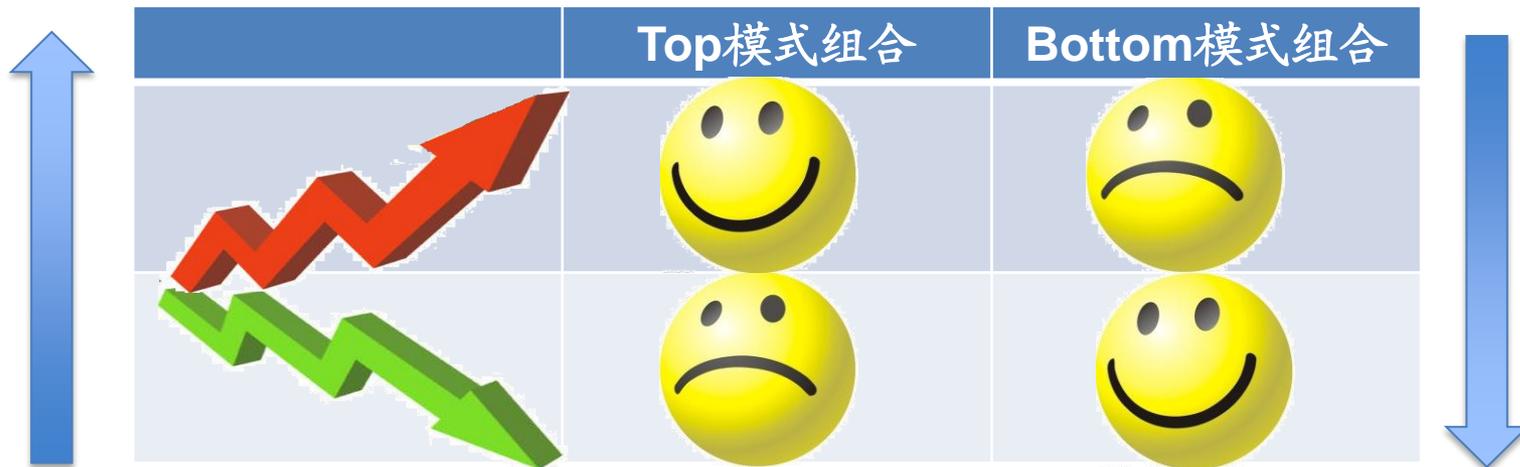
# 成交价格、成交量模式组合

- 选股模型：排序 (ranking)

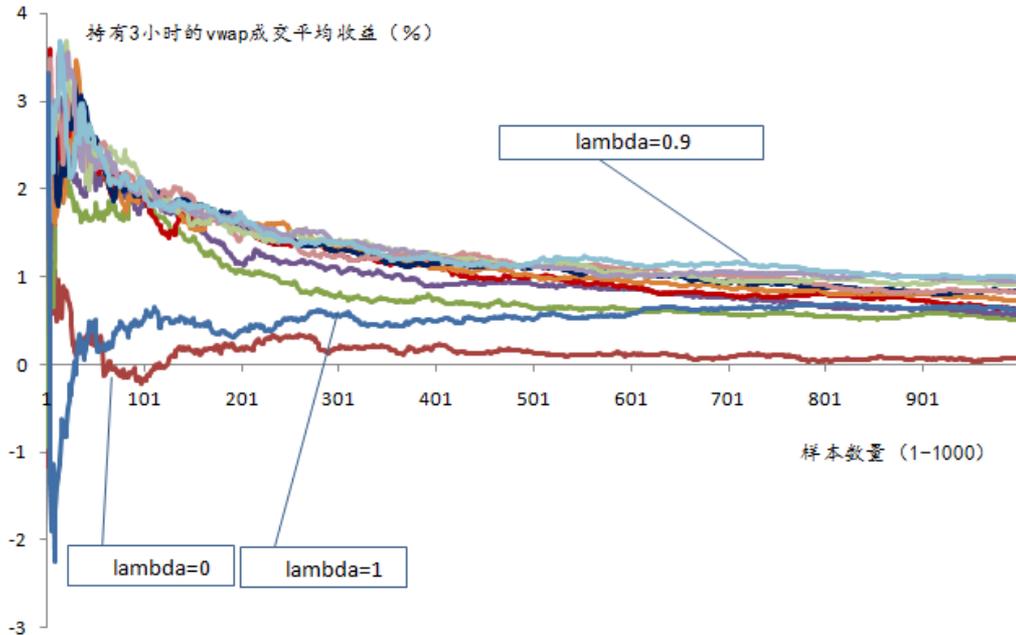
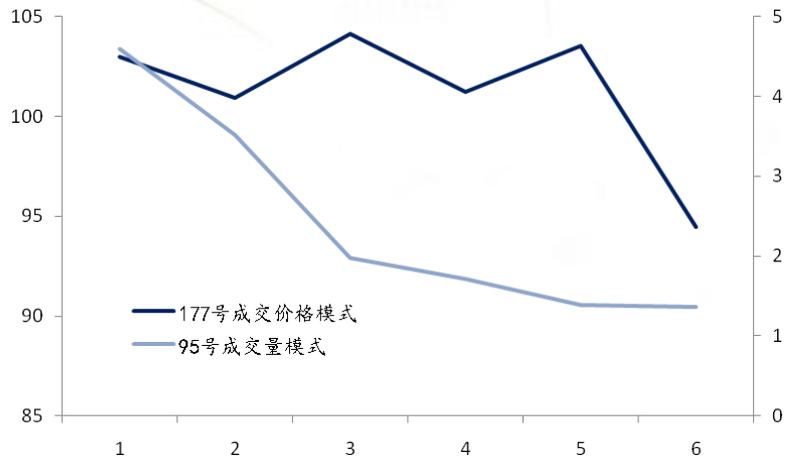
$$Score_k = \lambda \times Rank_{ik} + (1 - \lambda) \times Rank_{jk}, \lambda \in [0, 1]$$

$$i = 1, 2, 3, \dots, 198; j = 1, 2, 3, \dots, 194; k = 1, 2, 3, \dots, 121939$$

Average  
Precision

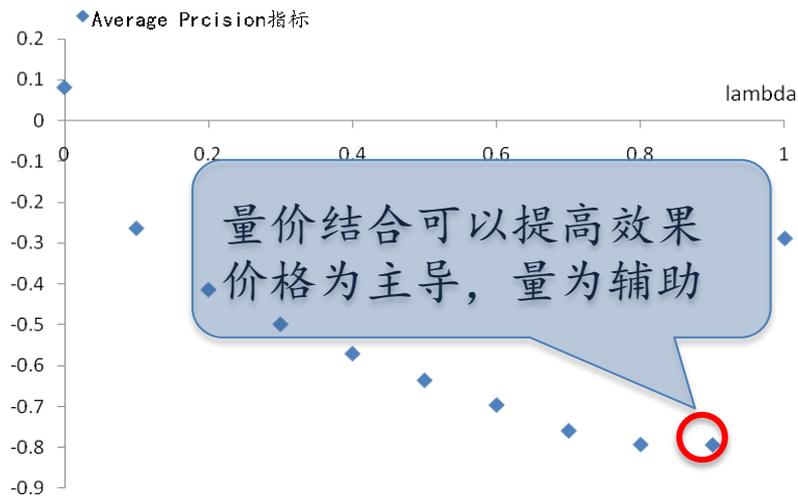
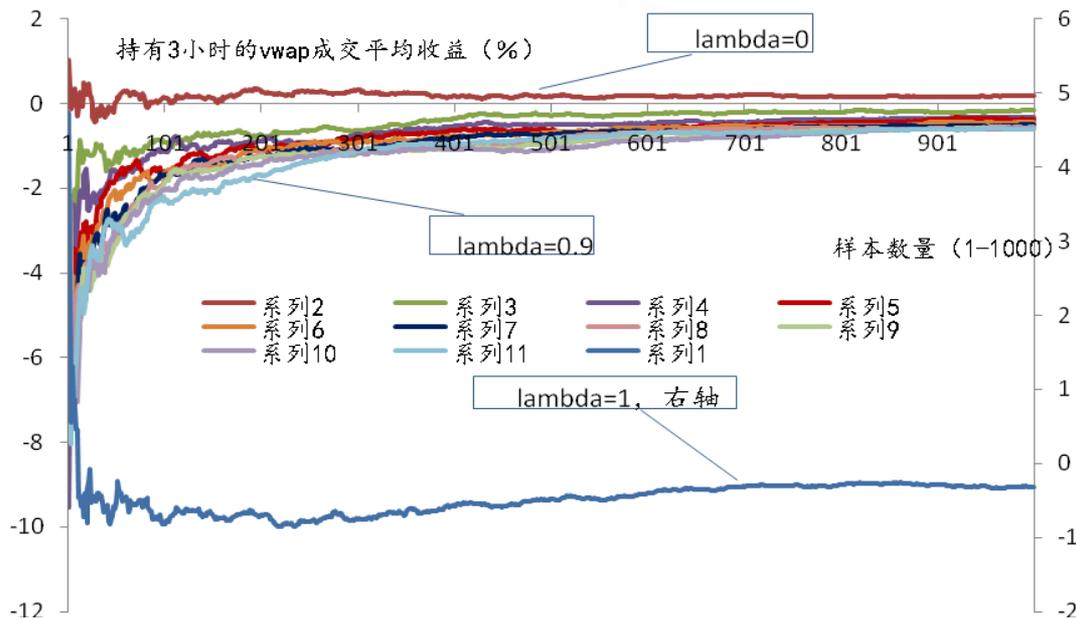
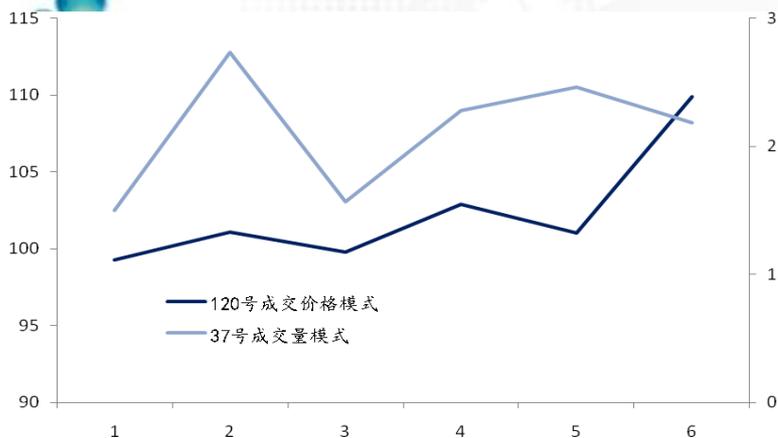


# Top量价模式组合



资料来源：天软科技，国信证券经济研究所整理

# Bottom量价模式组合



资料来源：天软科技，国信证券经济研究所整理

# 目录



模式聚类的提出



基于模式聚类的短线选股模型



成交价格、成交量模式组合

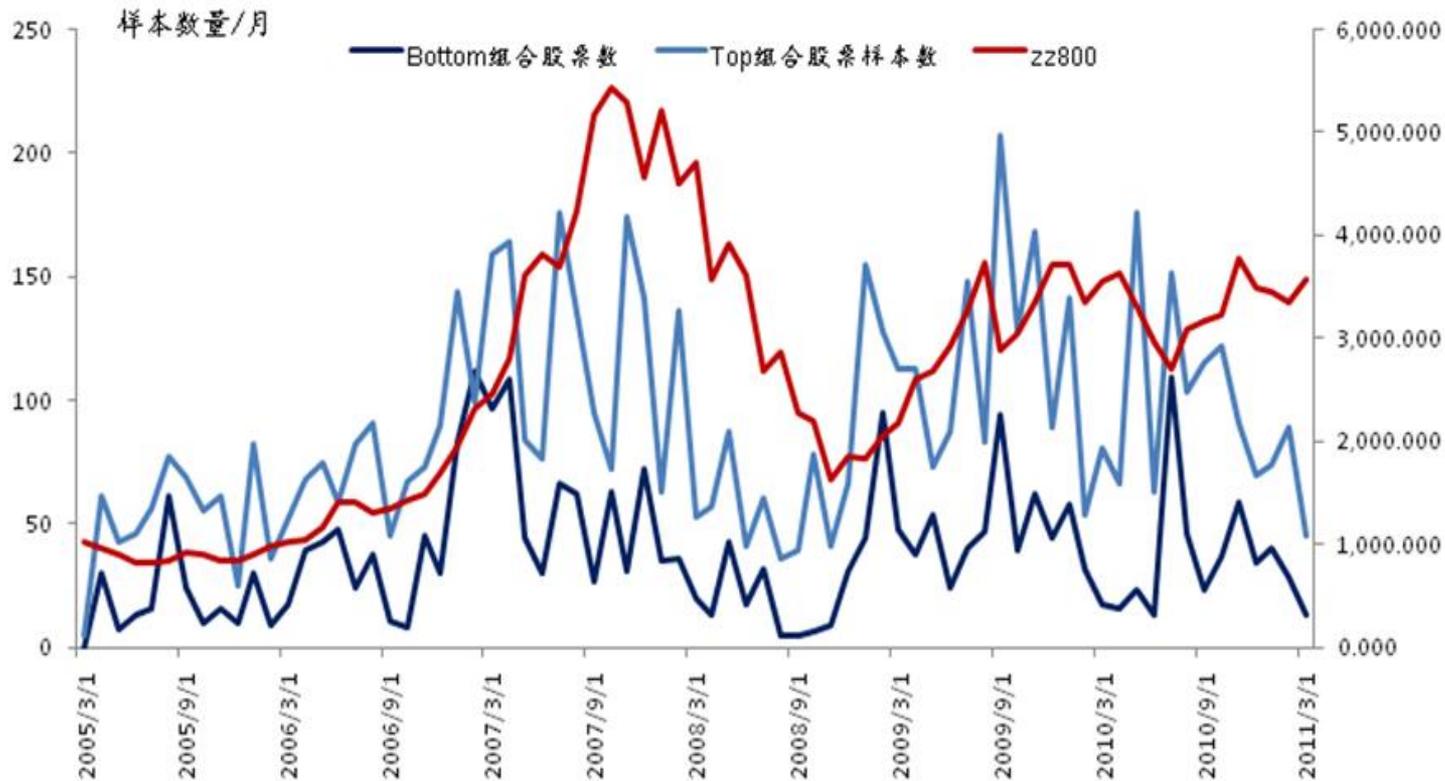


**模型应用分析**



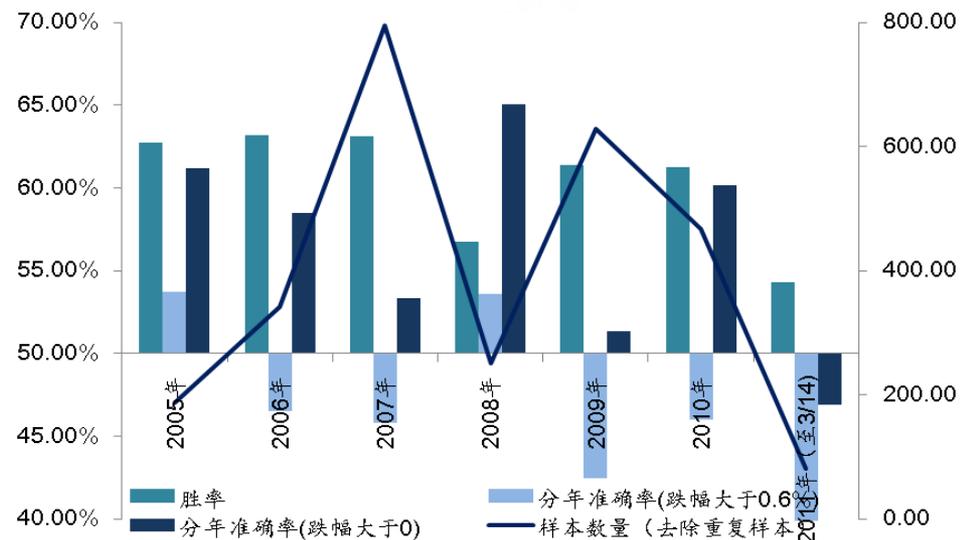
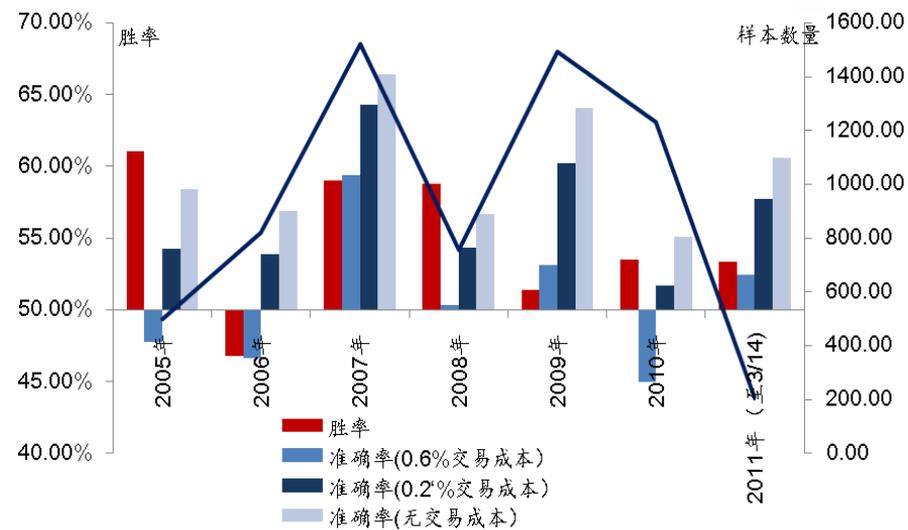
应用和扩展

# Top、Bottom组合样本数分布



资料来源：天软科技，国信证券经济研究所整理

# Top、Bottom 组合胜率

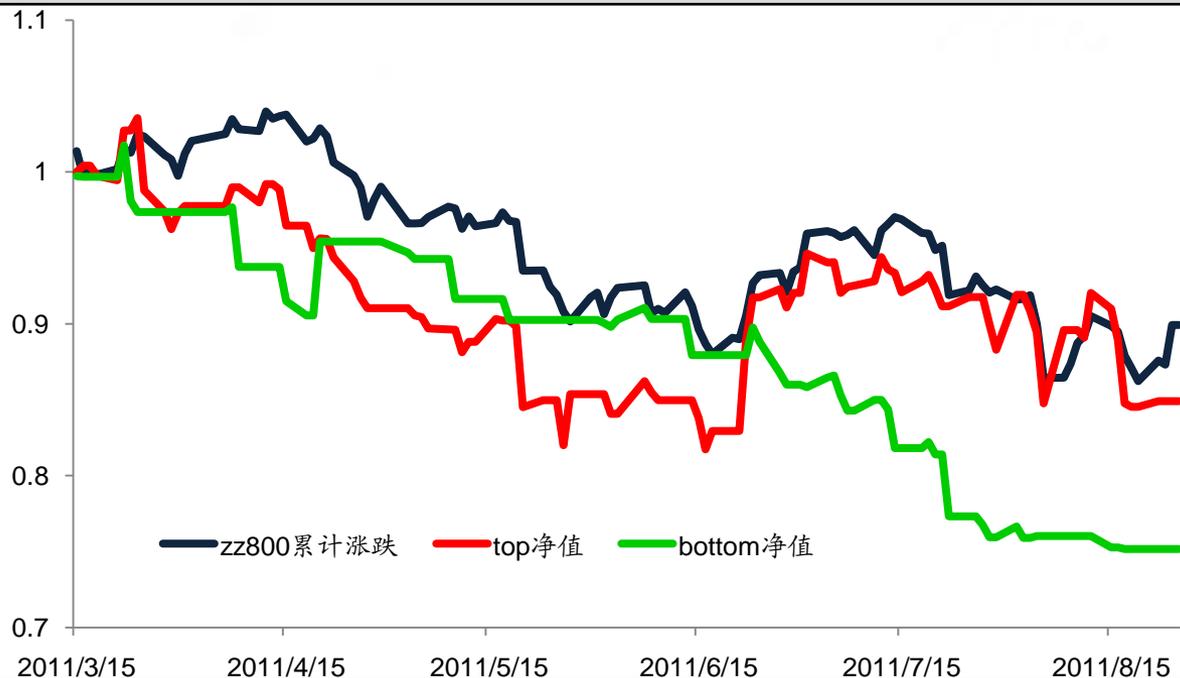


资料来源：天软科技，国信证券经济研究所整理



# Top、Bottom组合样本外区间检验

	样本总数	涨(跌)幅>0.6%	涨(跌)幅>0.2%	涨(跌)幅>0	有样本的交易日数	涨(跌)幅>zz800 指数
top20模式组合	363	105	131	142	78	33
比例	100.00%	28.93%	36.09%	39.12%	100.00%	42.31%
bottom20模式组合	88	44	49	56	42	25
比例	100.00%	50.00%	55.68%	63.64%	100.00%	59.52%



资料来源：天软科技，国信证券经济研究所整理



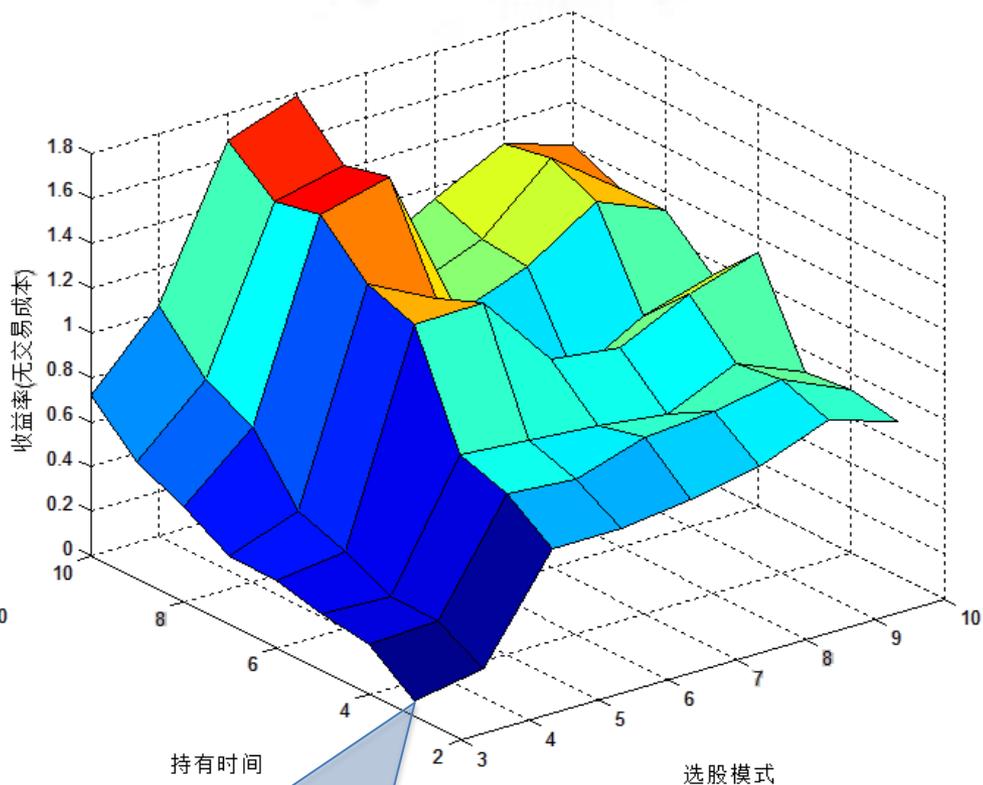
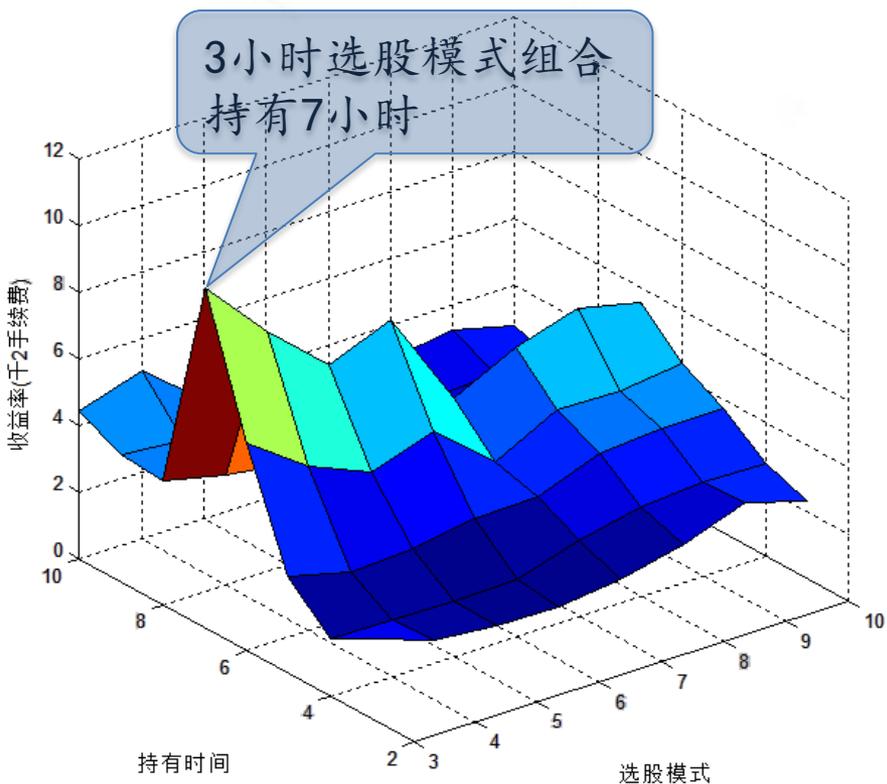
# 持有时间对模式组合的影响



## Top模式组合



## Bottom模式组合



# 目录



模式聚类的提出



基于模式聚类的短线选股模型



成交价格、成交量模式组合



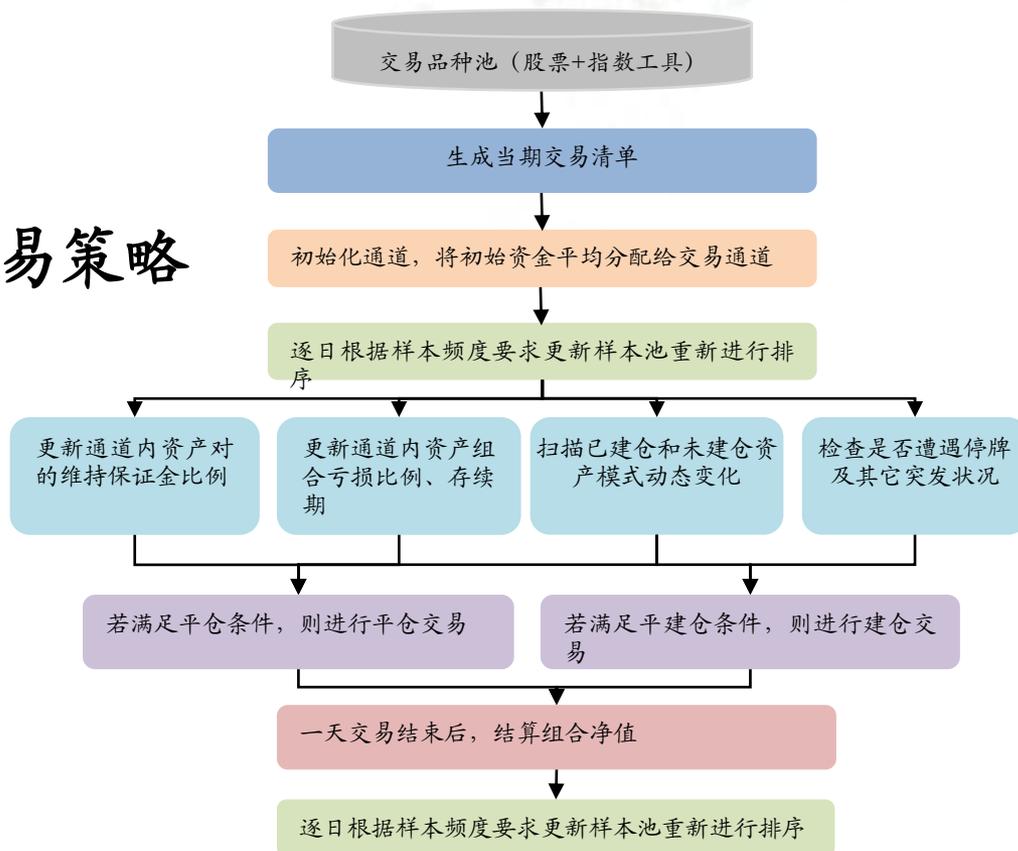
模型应用分析



应用和扩展

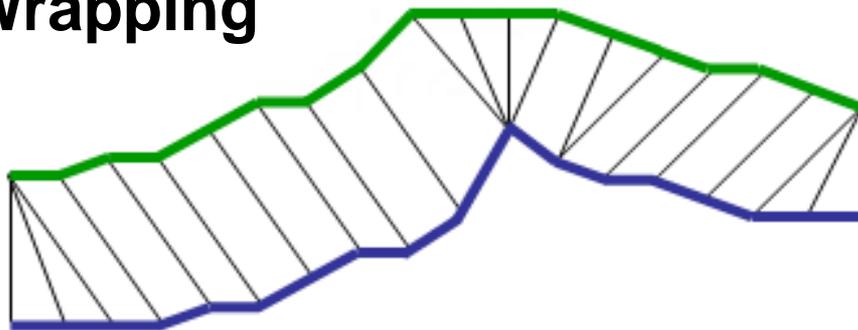
# 应用扩展

- 优化讨论
- 选股模型用于指数投资
- 基于模型选股结果的交易策略

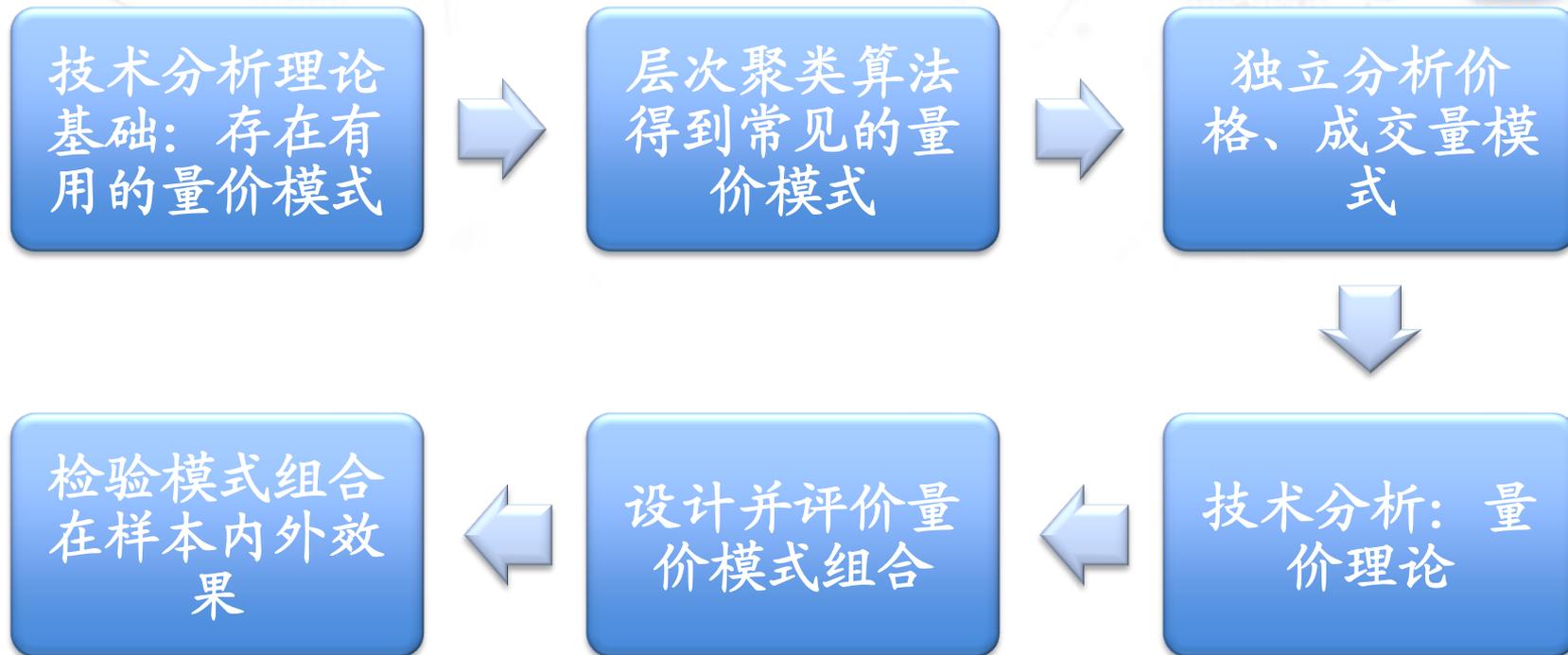


# 应用扩展

- 频度扩展：小时线、日K线
- 模式伸缩：Dynamic Time Wrapping
- 择时应用



# 总结





# 期待您的关注和意见！ 谢谢！

## 风险提示

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。

## 证券投资咨询业务的说明

证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。

证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。



**国信证券经济研究所**  
Guosen Securities Economic Research Institute