

# 深度学习新进展： Alpha因子的再挖掘

安宁宁 S0260512020003

邮箱: [anningning@gf.com.cn](mailto:anningning@gf.com.cn)

广发证券金融工程

2017年6月



# 03

| 策略与实证分析 |

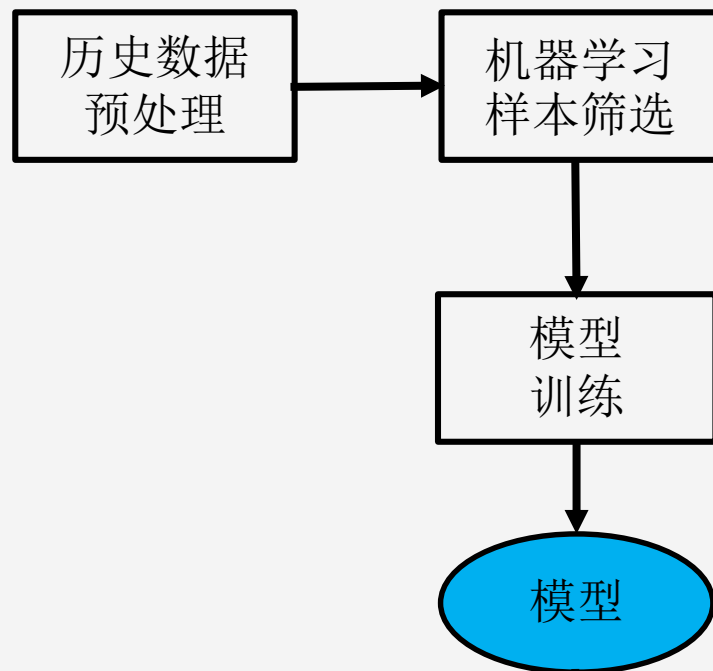


## 回测参数设置

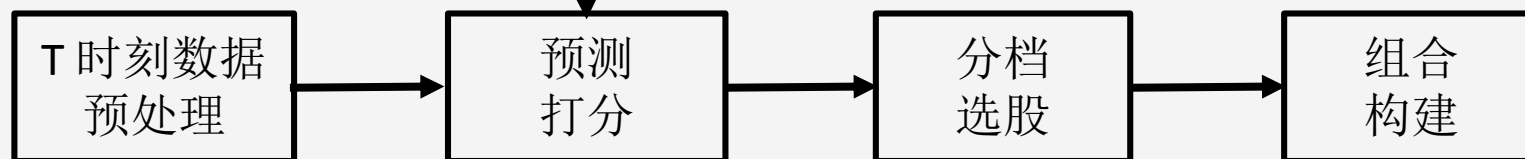
- **调仓周期**：20个交易日
- **股票池**：全市场选股，剔除上市交易时间不满一年的股票，剔除ST股票，剔除交易日停牌和涨停、跌停的股票
- **超配组合**：调仓时分十档，等权买入深度学习模型打分最高的一档
- **对冲基准**：中证500指数
- **原始因子数据**：估值因子、规模因子、反转因子、流动性因子、波动性因子、技术指标，共计128个因子
- **深度学习模型训练期**：2007年1月-2010年12月
- **策略回测**：2011年1月-2017年4月

## 策略流程

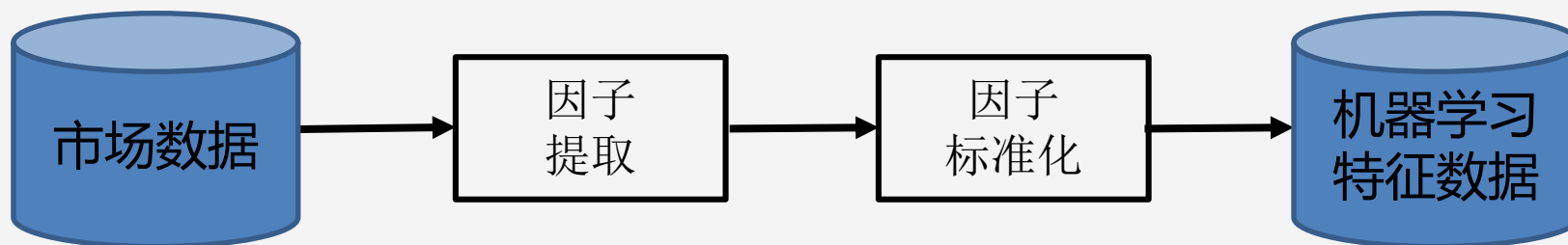
### 模型训练



### 选股交易



## 数据预处理

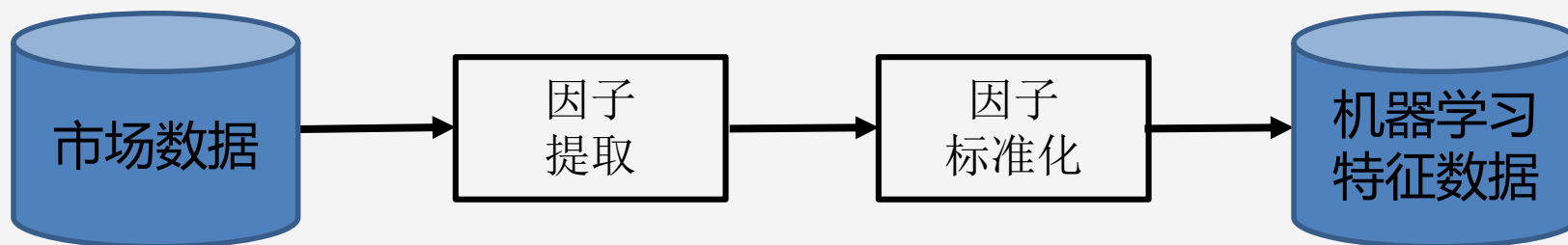


因子提取：从Wind终端提取数据，计算因子

一般而言，结合专业领域知识，提取合适的特征，有利于提高机器学习模型的性能

- 常用的选股因子，如规模因子、反转因子、PB、PE等本身就是比较有效的选股因子
- 技术指标选股也可以获取一定的超额收益（参见广发金工报告《Alpha因子何处寻 掘金海量技术指标》）

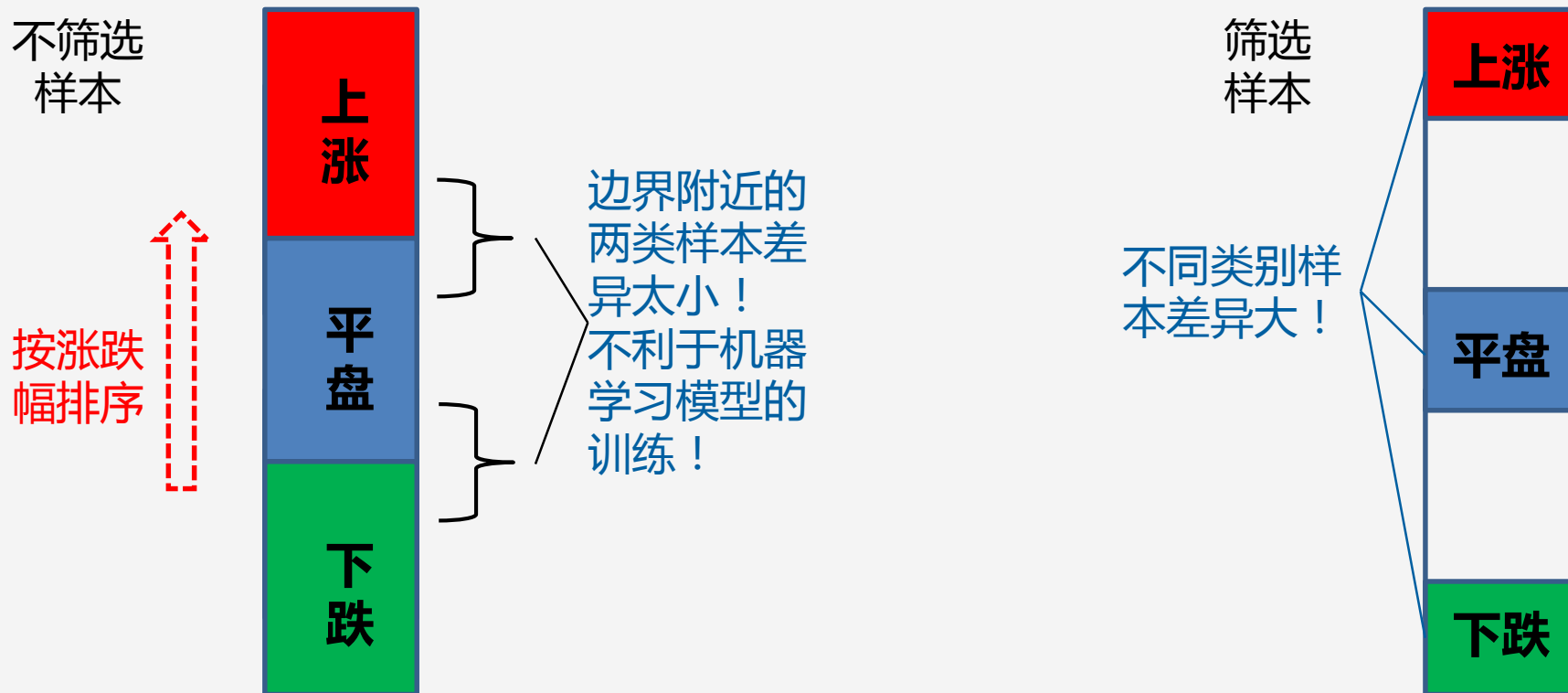
## 数据预处理



### 因子标准化：

- 1、异常值、缺失值处理
- 2、极值压边界处理
- 3、沿时间方向的因子标准化
- 4、沿截面的因子标准化
- 5、按照机器学习模型来调整因子分布（应该服从正态分布，还是[0,1]区间分布？）

样本筛选：减少噪声影响（**仅用于训练机器学习模型**）



训练时，根据未来20日后的股票涨跌幅来给样本贴“标签”：上涨、下跌、平盘

同一时刻，按照涨跌幅进行排序，确定样本的输出标签

模型训练：通过训练样本，确定模型结构，优化模型参数

预测输出 Y 的维度：3

输入特征 X 的维度：156（128个因子+28个行业）

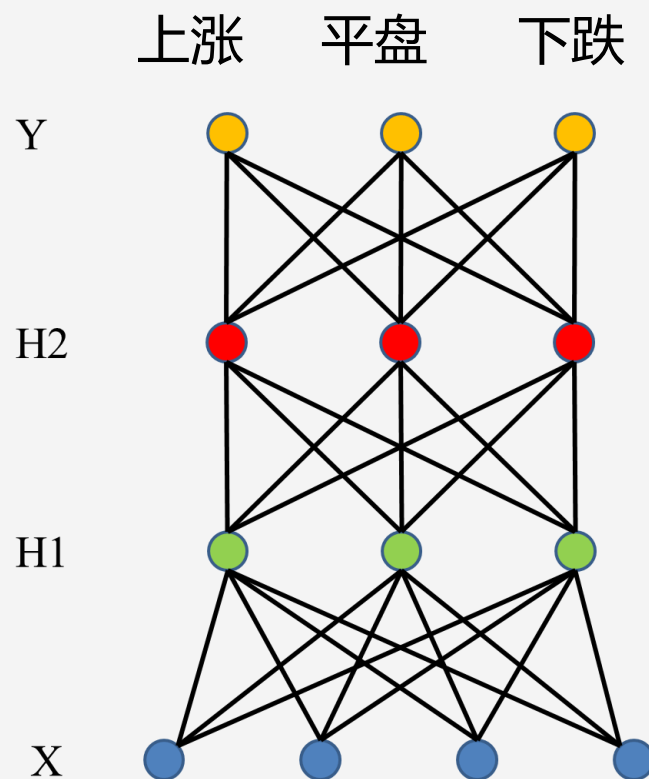
通过网格搜索获取最优的模型结构

选取模型结构为：

156（输入层）-512-200-200-200-128-3（输出层）

即一共包含5个隐层

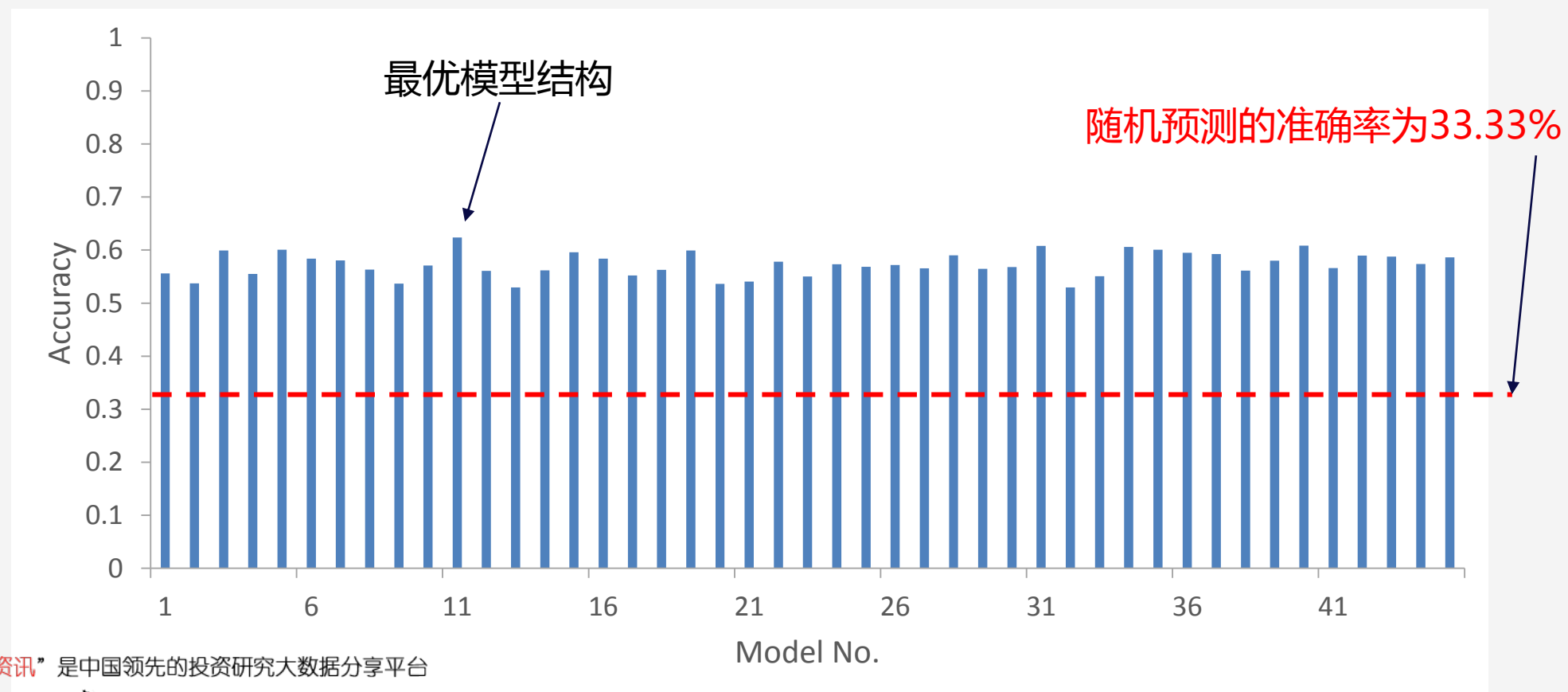
隐层节点数依次为：512（隐层1）、200（隐层2）、  
200（隐层3）、200（隐层4）、128（隐层5）





## 模型训练：

通过网格搜索获取最优的模型结构（5个隐层的结构下，不同模型的预测准确率基本上都在50%以上）



模型性能

训练样本数量：37万

验证集样本数量：5万

训练集预测准确率 = 67.84%

验证集预测准确率 = 62.32%

		预测			
		上涨	平盘	下跌	合计
实际	上涨	12403	3087	1182	16672
	平盘	6052	9103	2248	17403
	下跌	3762	2535	9700	15997
合计		22217	14725	13130	50072

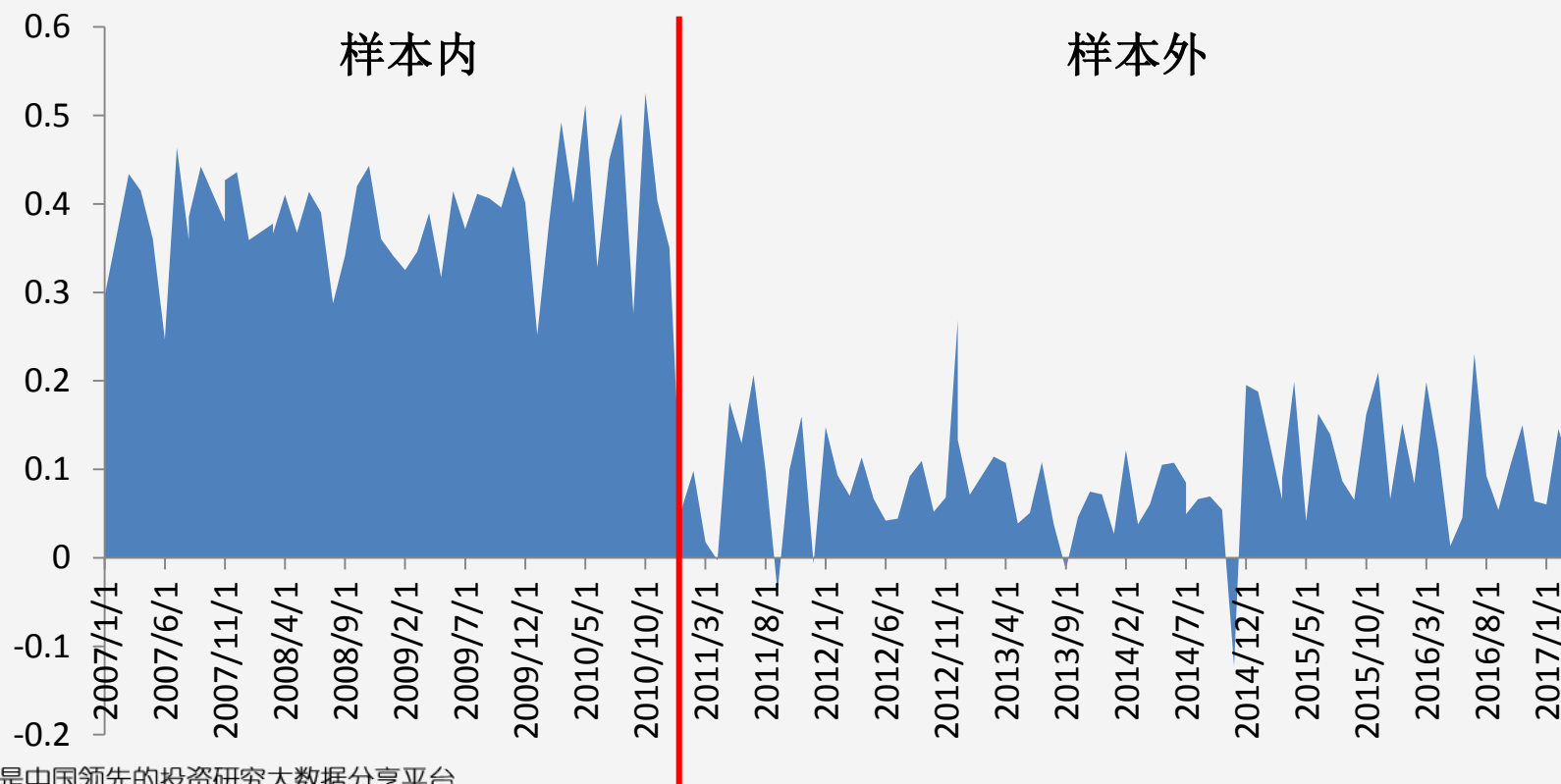
预测上涨的样本中，有55.8%属于上涨一类，仅有16.9%下跌。

“慧博资讯”是中国领先的投资研究大数据分享平台

数据来源：Wind、广发证券发展研究中心

## 因子预测能力：IC值

样本外 IC 的平均值为 0.092，标准差为 0.065



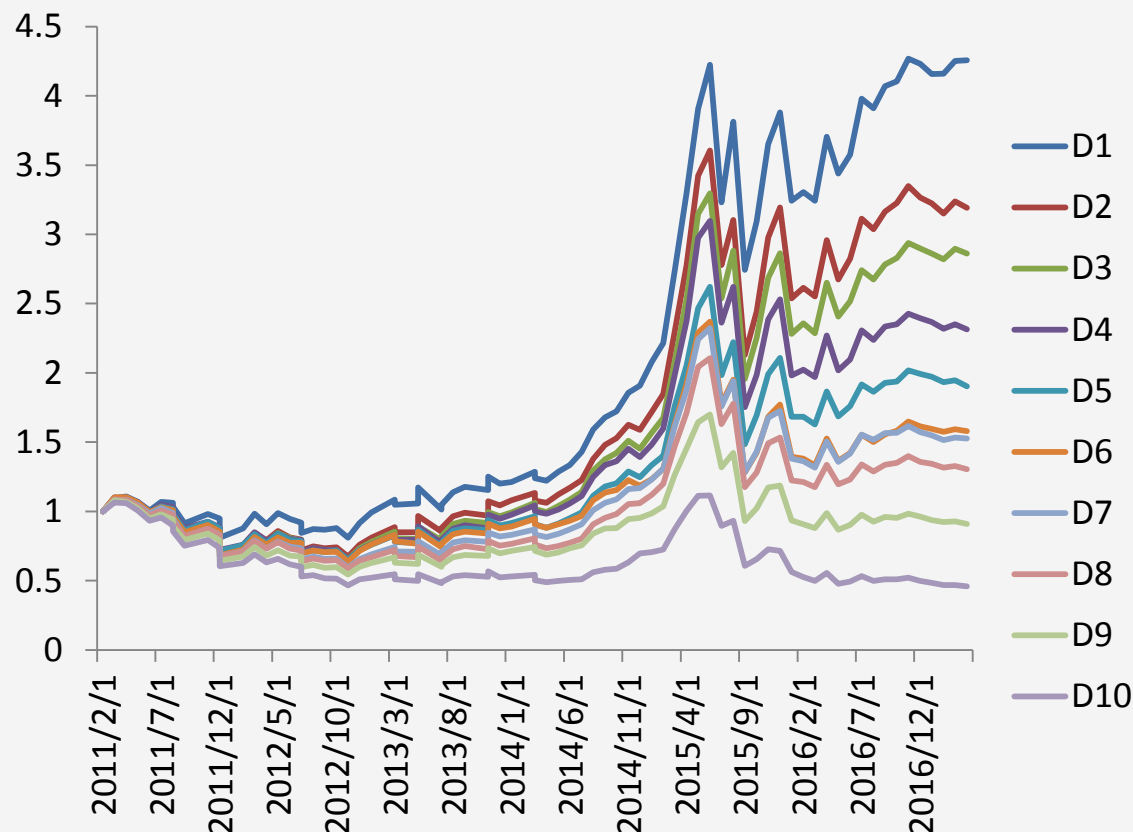
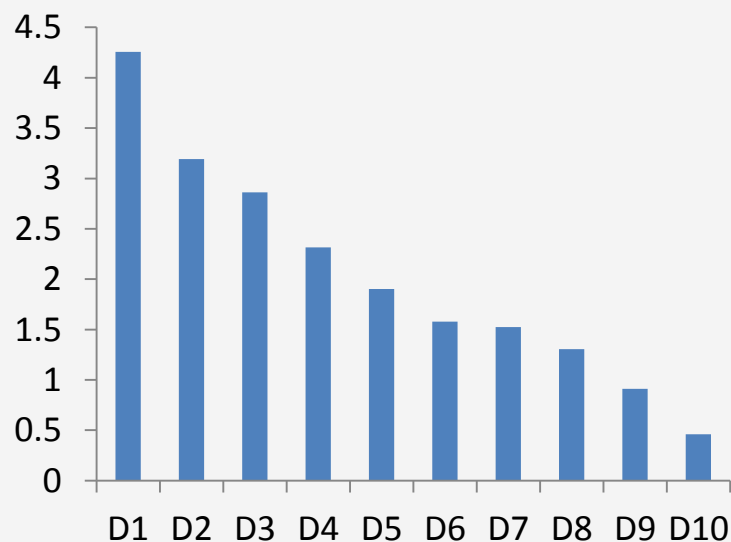
“慧博资讯”是中国领先的投资研究大数据分享平台

数据来源：Wind、广发证券发展研究中心

## 因子分档表现

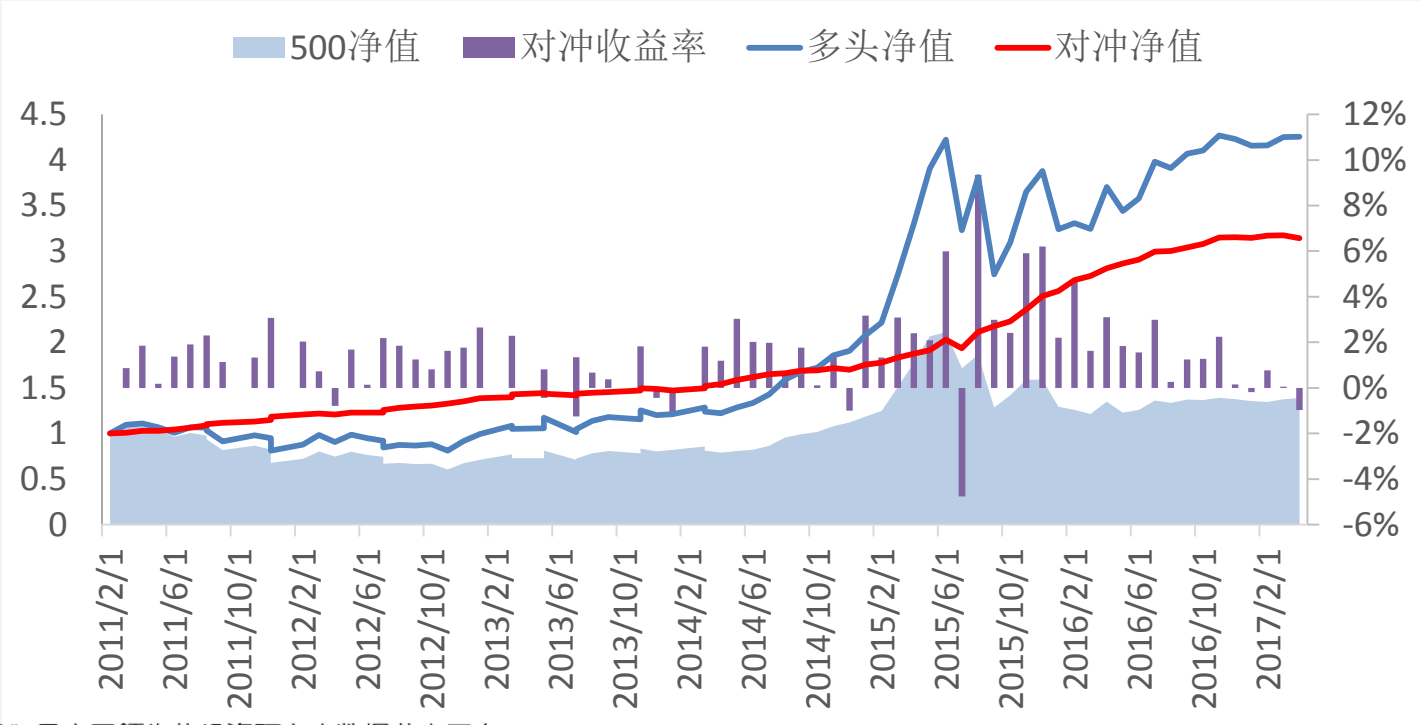
因子表现的单调性好

深度学习打分越高，累积收益率越高



因子预测能力：与中证500指数对冲表现

2011年以来，年化收益率20.3%，最大回撤 -4.77%，月度胜率为 88.0%



年份	累积收益率	最大回撤
2011	21.59%	-1.93%
2012	17.98%	-1.35%
2013	13.00%	-2.53%
2014	18.72%	-3.40%
2015	52.48%	-4.77%
2016	26.43%	-1.79%
2017	0.83%	-1.17%

“慧博资讯”是中国领先的投资研究大数据分享平台

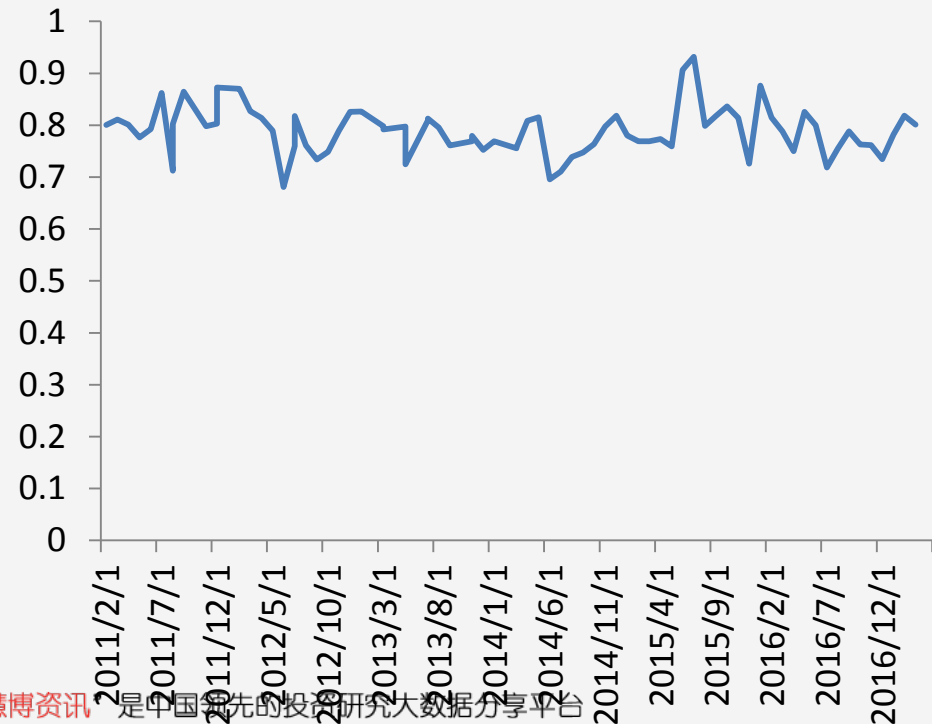
数据来源：Wind、广发证券发展研究中心

换手率和交易成本的影响

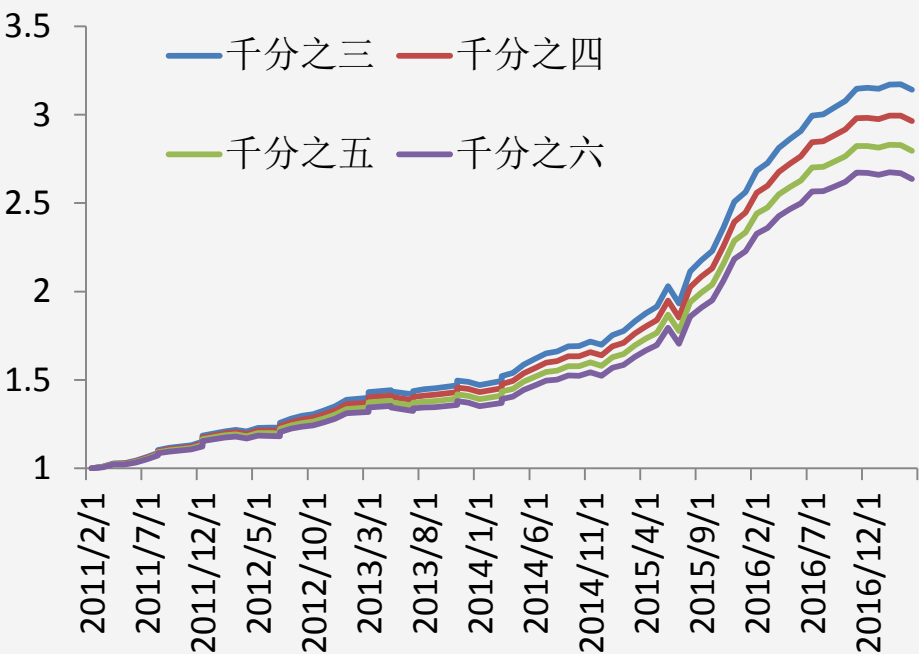
由于技术因子比较多，模型的换手率比较高

交易成本	0.3%	0.4%	0.5%	0.6%
年化收益率	20.28%	19.15%	18.04%	16.93%
最大回撤	-4.77%	-4.86%	-4.95%	-5.04%

超配组合换手率



不同交易成本下策略表现

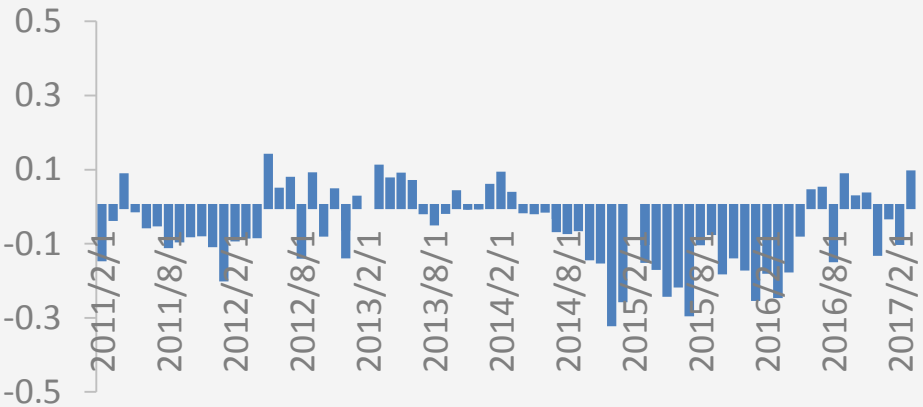


深度学习因子与风格因子的相关性  
与常见风格因子的相关性不高

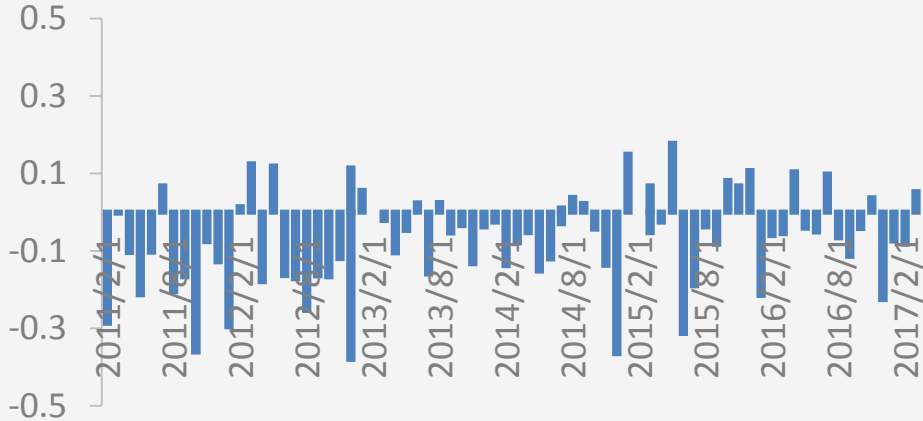
样本外因子的相关性

	深度学习因子
流通市值	-0.060
20日反转	-0.073
20日换手率	-0.119
盈市率	0.017

流通市值



20日反转



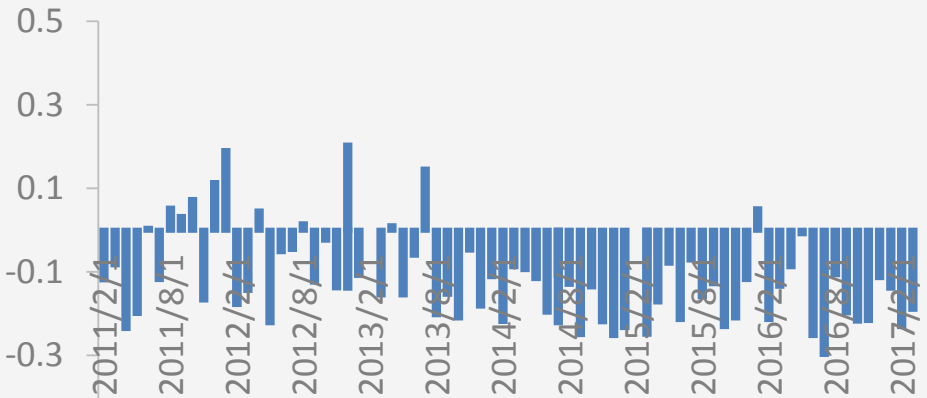
深度学习因子与风格因子的相关性

与常见风格因子的相关性不高

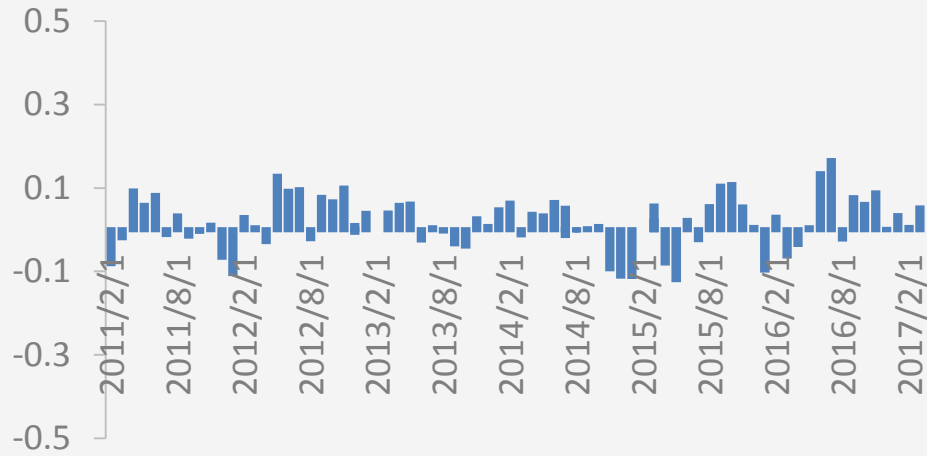
样本外因子的相关性

	深度学习因子
流通市值	-0.060
20日反转	-0.073
20日换手率	-0.119
盈市率	0.017

20日换手率



盈市率





拓展讨论：模型更新

1/1/2006	1/1/2011	4/28/2017
模型2010：样本内	样本外	

**滚动更新模型：**训练初始模型之后，固定时间（如，每年）更新一次模型

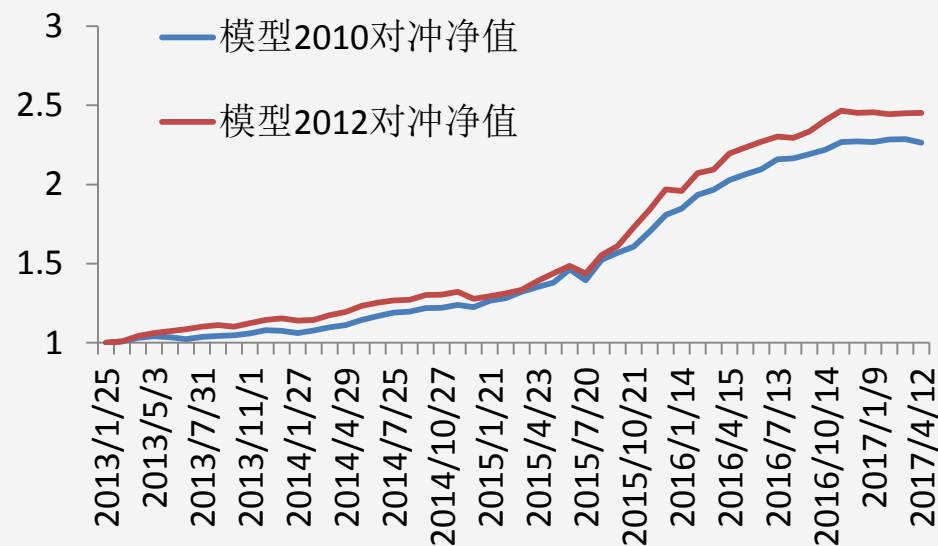
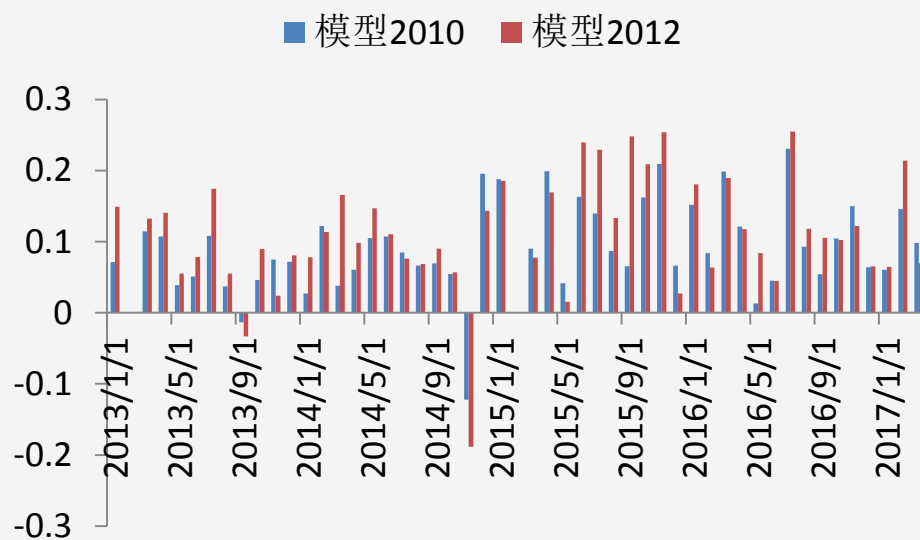
1/1/2006	1/1/2010	1/1/2011	1/1/2012	1/1/2013	.....
模型2010：样本内	样本外				
	模型2011：样本内		样本外		
		模型2012：样本内		样本外	

## 拓展讨论：模型更新

模型2010和模型2012在2013年之后的回测对比

更新后的模型年化收益有明显增强，最大回撤有明显降低

	模型2010	模型2012
年化收益率	22.2%	25.7%
最大回撤	-4.77%	-3.34%





# 04

| 总结 |

>

- 通过实证分析，证实了深度学习交易策略可以用于月频的选股交易上
- 全市场选股，选取十分之一数量的股票进行配置，用中证500指数对冲，从2011年以来，年化收益率为 20.3%，最大回撤为 -4.77%，月度胜率为 88.0%
- 选股因子与常见风格因子（规模、反转、流动性、估值）的相关性不高