

# 多周期机器学习选股模型

## 人工智能研究报告

### 报告摘要:

#### ● 选股模型的时效性

信息具有时效性。选股因子对股票收益率的预测能力会随着时间的延后而衰减。机器学习股票收益预测模型的目标是将股票因子与股票未来收益率关联起来。股票因子蕴含的信息决定了模型的预测能力,包括预测准确度和预测窗口长度。如果机器学习模型所用的股票因子中包含的是市场短期情绪面的信息,那么训练出来的机器学习模型可能对市场短期走势的预测能力较强;如果机器学习模型所用的因子包含的是市场中长期的价格扭曲信息,那么训练出来的机器学习模型可能对市场中长期的预测能力较强。

#### ● 模型构建

本报告按照因子在不同预测窗口长度的 IC 将选股因子分成不同的组别,并针对不同的股票收益预测问题,分别构建机器学习预测模型,包括短线预测模型、中线预测模型和长线预测模型。

#### ● 实证分析

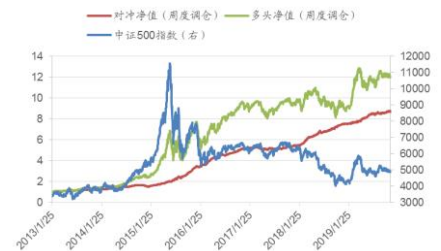
实证分析表明机器学习模型在不同预测窗口长度的股票收益率预测上都能够取得不错的效果,基于不同机器学习模型构建的选股策略均表现出不错的性能。机器学习选股因子的 IC 和 ICIR 会随时间衰减,在高频调仓频率的选股策略中,短线预测机器学习模型表现较好;在低频调仓频率的选股策略中,长线预测机器学习模型表现较好。由于机器学习模型的性能随着预测窗口的增长而减弱,提高换仓频率一般会获得更好的收益表现。在千分之三的换仓成本下,相对月频或季频调仓,周频调仓是机器学习模型更合适的一个调仓频率。

不同机器学习选股因子的相关性不高,因此可以将不同模型的选股因子集成进行多模型选股。在周频调仓下,模型 S、模型 M 和模型 L 三模型集成选股策略的年化超额收益为 38.97%,夏普比率为 3.89,显著优于单模型的选股表现,而且模型的换手率也比中短线预测模型的换手率要低。

#### ● 风险提示

策略模型并非百分百有效,市场结构及交易行为的改变以及类似交易参与者的增多有可能使得策略失效。

图 1: 多模型集成策略收益曲线



数据来源: Wind, 广发证券发展研究中心

表 1: 多模型集成策略选股表现

指标	多模型集成策略
年化超额收益	38.97%
最大回撤	-10.00%
夏普比率	3.89
年换手率	32.27

数据来源: Wind, 广发证券发展研究中心

分析师: 文巧钧



SAC 执证号: S0260517070001

SFC CE No. BNI358



0755-88286935



wenqiaojun@gf.com.cn

分析师: 安宁宁



SAC 执证号: S0260512020003

SFC CE No. BNW179



0755-23948352



anningning@gf.com.cn

分析师: 罗军



SAC 执证号: S0260511010004

020-66335128



luojun@gf.com.cn

请注意, 罗军并非香港证券及期货事务监察委员会的注册持牌人, 不可在香港从事受监管活动。

### 相关研究:

机器学习模型在因子选股上 2019-05-12

的比较分析:人工智能研究报告

## 目录索引

一、问题背景 .....	4
二、策略原理与流程 .....	4
2.1 特定训练窗口长度的机器学习模型 .....	4
2.2 极限梯度提升树 .....	6
2.3 机器学习股票收益率预测模型 .....	7
2.4 策略回测设置 .....	8
三、实证分析 .....	8
3.1 机器学习模型因子筛选 .....	8
3.2 不同训练窗口长度的模型表现 .....	12
3.3 策略在不同调仓周期下的表现 .....	18
3.4 多模型集成策略的表现 .....	22
四、总结与展望 .....	24
风险提示 .....	24

## 图表索引

图 1: 不同训练窗口的选股模型 .....	5
图 2: 机器学习选股框架 .....	6
图 3: XGBoost 模型求解示意图 .....	7
图 4: 不同训练窗口长度的模型预测准确度 .....	13
图 5: 模型 S 的 IC .....	13
图 6: 模型 M 的 IC .....	14
图 7: 模型 L 的 IC .....	14
图 8: 模型 IC 的衰减情况 .....	15
图 9: 模型年化 ICIR 的衰减情况 .....	15
图 10: 模型 S 周度调仓选股表现 .....	16
图 11: 模型 M 月度调仓选股表现 .....	17
图 12: 模型 L 季度调仓选股表现 .....	18
图 13: 模型 S 不同调仓周期选股表现 .....	19
图 14: 模型 M 不同调仓周期选股表现 .....	20
图 15: 模型 L 不同调仓周期选股表现 .....	21
图 16: 机器学习多模型集成策略选股表现 .....	22
图 17: 多模型集成策略的 IC 与预测窗口长度的关系 .....	23
表 1: 候选风格因子 .....	8
表 2: 不同预测周期的因子 IC .....	10
表 3: 模型 S 周度调仓策略分年度表现 .....	16
表 4: 模型 M 月度调仓策略分年度表现 .....	17
表 5: 模型 L 季度调仓策略分年度表现 .....	18
表 6: 模型 S 不同周期策略表现汇总 .....	19
表 7: 模型 M 不同周期策略表现汇总 .....	20
表 8: 模型 L 不同周期策略表现汇总 .....	21
表 9: 不同调仓频率下模型夏普比率比较 .....	21
表 10: 不同模型股票收益率预测打分相关性 .....	22
表 11: 多模型集成周度调仓策略分年度表现 .....	23
表 12: 不同周期机器学习选股模型对冲表现 .....	23

## 一、问题背景

信息具有时效性。一般而言，选股因子对股票收益率的预测能力会随着时间的延后而不断衰减。由于不同因子中蕴含的股票信息有差异，不同因子对股票未来收益率的最佳预测窗口的长度也不一样。一般而言，高频价量信息反映股票短期的走势，对股票短期表现的预测能力较强，而估值、盈利等基本面信息反映股票长期的价格扭曲，对股票中长期表现的预测能力较强。

机器学习股票收益预测模型的目标是将股票因子与股票未来收益率关联起来。股票因子蕴含的信息决定了模型的预测能力，包括预测准确度和预测窗口长度。如果机器学习模型所用的股票因子中包含的是市场短期情绪面的信息，那么训练出来的机器学习模型可能对市场短期走势的预测能力较强；如果机器学习模型所用的因子包含的是市场中长期的价格扭曲信息，那么训练出来的机器学习模型可能对市场长期的预测能力较强。

本报告主要解决以下问题：

### 1) 如何构建适用于特定预测窗口长度的机器学习选股模型。

本报告考察了常用的风格因子和技术指标，尝试按照因子在不同预测窗口长度的IC将选股因子分成不同的组别，并针对不同的股票收益率预测问题，分别采用XGBoost模型构建机器学习预测模型。实证分析表明机器学习模型在不同预测窗口长度的股票收益率预测上都能够取得不错的效果，基于不同机器学习模型构建的选股策略表现出不错的性能。

### 2) 机器学习股票收益预测模型选股性能的时效性。

本报告对于不同预测窗口长度的机器学习模型，研究选股因子的IC和ICIR随着时间衰减的情况，考察机器学习模型在不同调仓频率下的性能。实证分析表明，机器学习选股因子的IC和ICIR会随时间衰减，在高调仓频率的选股策略中，短线预测模型表现较好，在低调仓频率的选股策略中，长线预测模型表现较好。

### 3) 不同预测窗口长度的机器学习模型的结合。

不同机器学习选股因子的相关性不高，因此可以将不同模型的选股因子结合起来进行多模型选股，实证表明，多模型集成选股具有更好的表现。

## 二、策略原理与流程

### 2.1 特定训练窗口长度的机器学习模型

因子选股是通过对股票的收益率进行预测，寻找能够产生超额收益的股票。机器学习通过对股票历史数据的学习，建立股票收益率的预测模型。模型训练时，如果预测股票收益率的时间窗口越小，训练出来的机器学习模型倾向于预测股票的短期走势；如果预测股票收益率的时间窗口较长，则机器学习模型倾向于预测中长线的股票走势。为了避免混淆，本报告将训练机器学习模型时，股票收益率的计算窗口称为“训练窗口”。

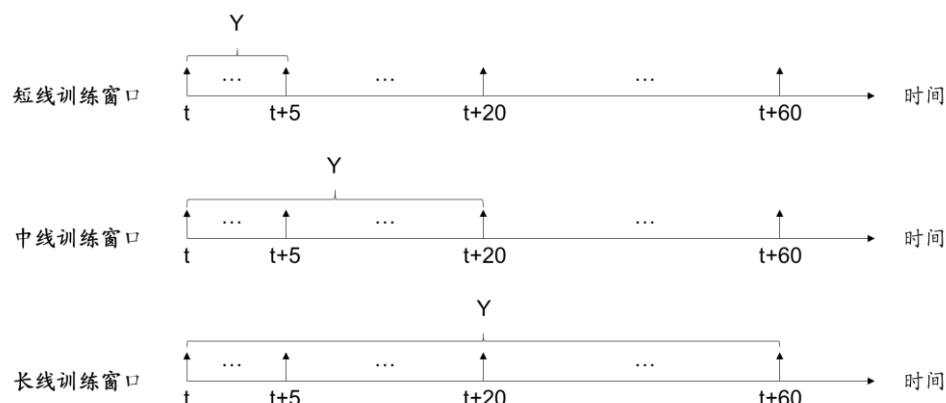
本报告针对不同的训练窗口长度，构建了三组不同的机器学习预测模型：

1) “短线预测模型”：简称模型 S，训练窗口为 5 个交易日（即 1 周）。在模型训练时，预测目标是未来 5 个交易日的股票收益率。

2) “中线预测模型”：简称模型 M，训练窗口为 20 个交易日（约 1 个月）。在模型训练时，预测目标是未来 20 个交易日的股票收益率。

3) “长线预测模型”：简称模型 L，训练窗口为 60 个交易日（约 1 个季度）。在模型训练时，预测目标是未来 60 个交易日的股票收益率。

图 1：不同训练窗口的选股模型



数据来源：广发证券发展研究中心

为了建立不同训练窗口长度的选股模型，本报告首先从候选因子池中考察不同因子的 IC，根据因子 IC 筛选出适合该训练窗口长度模型的因子。

为了训练模型 S，在因子筛选时考察候选因子与未来 5 个交易日收益率的秩相关系数，筛选预测能力最强的因子。类似的，模型 M 的因子筛选时考察候选因子与未来 20 个交易日收益率的相关性，模型 L 的因子筛选时考察候选因子与未来 60 个交易日收益率的相关性。

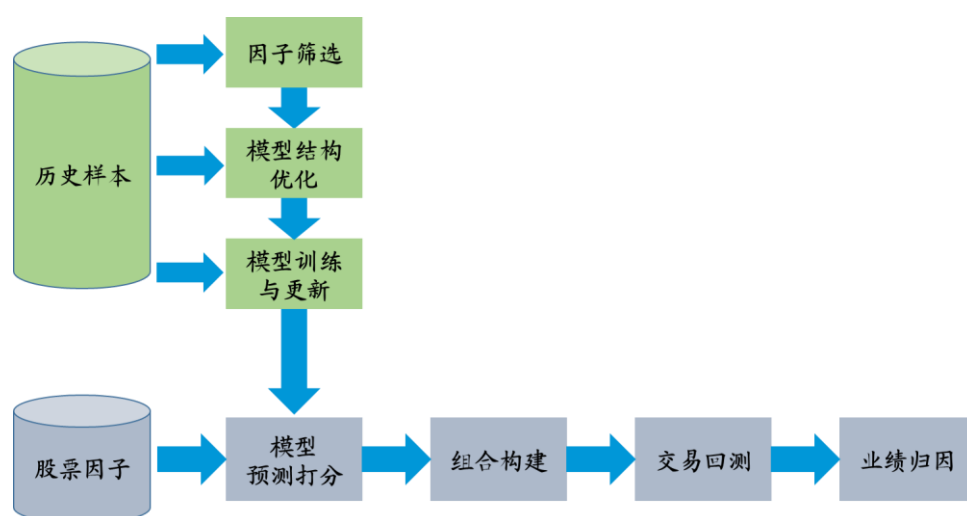
通过筛选出来的因子，优化机器学习模型超参数，训练机器学习模型，并对模型进行定期的迭代更新。本报告通过交叉验证获取模型的最优超参数，每半年进行模型的更新迭代，每次更新时采用过去 6 年的股票样本训练模型。

在机器学习模型选股时，采用最新一期的机器学习模型预测股票的收益，并根据预测结果筛选股票，构建组合。

机器学习因子选股的流程如图 2 所示。



图 2: 机器学习选股框架



数据来源：广发证券发展研究中心

## 2.2 极限梯度提升树

在我们此前的报告《机器学习模型在因子选股上的比较分析:人工智能研究报告》中，探讨了不同的机器学习模型用于因子选股的表现，实证表明，不同机器学习模型的表现具有较高的相关性，其中，极限梯度提升树（eXtreme Gradient Boosting, XGBoost）和深层神经网络（Deep Neural Networks, DNN）是选股表现较好的两种模型。本报告采用XGBoost训练机器学习模型。

XGBoost是近年来机器学习领域非常热门的一种算法，由华盛顿大学的陈天奇博士于发起，是梯度提升树的一种高效实现，在Kaggle等数据大赛中获得了很大的成功。

XGBoost是一种集成学习算法，可以以决策树为基学习器，也可以以线性分类器为基学习器。本报告中，采用决策树为XGBoost模型的基学习器。XGBoost可表示为基学习器的加法模型，随机森林的每个决策树都可以单独进行分类，最后将分类结果平均，而XGBoost的需要将不同决策树汇总起来进行输出预测。 $T$ 个决策树构建的XGBoost模型的输出值为：

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i)$$

XGBoost模型训练时，采用前向分步法，每一步只学习一个基学习器，逐步逼近优化目标，如下图所示。首先确定初始提升树 $\hat{y}_i^0 = f_0(x_i) = 0$ ，然后依次求解获得 $f_1(x_i)$ ， $f_2(x_i)$ ，……，和 $f_T(x_i)$ 。第 $t$ 步求解的模型是：

$$\hat{y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i)$$

其中， $\hat{y}_i^t$ 表示总共 $t$ 棵决策树组成的模型， $f_t(x_i)$ 表示第 $t$ 棵决策树的输出。实际上， $f_t(x_i)$ 用来拟合前 $t-1$ 棵决策树的输出 $\hat{y}_i^{t-1}$ 与真实值之间的残差。通过不断生成新的决策树模型，XGBoost可以逐步逼近拟合目标。

图 3: XGBoost模型求解示意图

$$\begin{aligned}
 &\text{初始设置: } \hat{y}_i^0 = 0 \\
 &\text{第1步: } \hat{y}_i^1 = f_1(x_i) = \hat{y}_i^0 + f_1(x_i) \quad \longrightarrow \quad f_1(x_i) \\
 &\text{第2步: } \hat{y}_i^2 = f_1(x_i) + f_2(x_i) = \hat{y}_i^1 + f_2(x_i) \quad \longrightarrow \quad f_2(x_i) \\
 &\dots\dots \\
 &\text{第t步: } \hat{y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i) \quad \longrightarrow \quad f_t(x_i)
 \end{aligned}$$

数据来源：广发证券发展研究中心

求解第t棵决策树时，最小化以下目标函数：

$$\begin{aligned}
 Obj^t &= \sum_{i=1}^N l(y_i, \hat{y}_i^t) + \sum_{i=1}^t \Omega(f_i) \\
 &= \sum_{i=1}^N l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) + constant
 \end{aligned}$$

其中， $\Omega(f_i)$  表示决策树的复杂度。通过对上述目标函数在 $\hat{y}_i^{t-1}$ 附近作二阶泰勒展开，以获得决策树模型的最优参数。

## 2.3 机器学习股票收益率预测模型

本报告训练了三组不同的机器学习模型：模型S、模型M和模型L。这三种模型的区别主要在于选取因子的差异和训练窗口长度的差异。

以模型S为例，训练窗口长度为5天，为了训练短周期预测模型，在股票样本构建时，本报告提取每一天股票池内的全体股票，剔除涨停、跌停股票和ST股票之后，根据未来5个交易日后的股票涨跌幅给不同的股票样本贴“标签”：“上涨”、“下跌”和“平盘”。其中，标记为上涨的股票是指相对强势的股票，即未来5个交易日收益率前10%的股票；标记为下跌的股票是指相对弱势的股票，即未来5个交易日收益率后10%的股票；标记为平盘的股票是指未来5个交易日收益率处于中间的10%的股票（分位数45%至55%区间内）。

股票预测模型的目标是寻找能够产生超额收益的股票。因此，机器学习模型在训练时需要建立起股票因子和未来涨跌标签之间的关系，构建一个以股票因子为输入，股票收益率涨跌标签为输出的模型。股票因子标记为

$$\mathbf{x} = [x_1 \quad \dots \quad x_m]^T$$

其中， $m$ 表示股票的因子个数，包括选股因子和行业0-1属性，在本报告中，三组不同的模型采用同样数量的因子， $m=78$ ，包括50个从候选因子池筛选的选股因子和28个行业0-1变量。股票涨跌标签记为 $y$ ， $y=1$ （上涨），2（下跌），3（平盘），表示三种不同的股票涨跌类别。

本报告中采用的机器学习模型是分类模型，将股票样本 $\mathbf{x}_i$ 分类为第 $k$ 类（ $k=1, 2,$

3) 的预测概率为

$$p(y_i = k | x_i) = f(x_i; w)$$

其中， $w$ 表示模型参数。 $p(y_i = k | x_i)$ 越大，表示股票样本 $x_i$ 属于第 $k$ 类的概率越大。

选股策略选择最有可能产生超额收益的股票构建多头组合，因此，机器学习的目标也就是识别出 $p(y_i = 1 | x_i)$ 最大的股票，本报告将 $p(y_i = 1 | x_i)$ 定义为机器学习模型打分或机器学习选股因子。

## 2.4 策略回测设置

在每个调仓日，根据机器学习模型的打分，筛选打分靠前的 10%的股票构建组合，进行策略的回测。相关参数如下：

调仓周期： 分别考察不同的调仓周期，包括 5 个交易日（周频换手）、20 个交易日（月频换手）、60 个交易日（季频换手）；

股票池： 全市场选股、剔除涨停、跌停的股票，停牌股票和 ST 股票；

超配组合： 机器学习打分前 10%的股票；

对冲基准： 中证 500 指数；

因子数据： 候选因子池包括估值因子、规模因子、反转因子、流动性因子、波动性因子等常用的 56 个风格因子和 101 个技术指标因子，通过因子 IC 筛选出不同模型的股票因子；

机器学习模型训练： 每半年滚动更新模型，每次模型更新时采用最近 6 年的样本作为训练集；

组合构建： 行业中性；

策略回测： 2013 年 1 月 1 日-2019 年 11 月 29 日；

交易成本： 千分之三。

## 三、实证分析

### 3.1 机器学习模型因子筛选

本报告以常用的 56 个风格因子和 101 个技术指标因子作为候选的因子，其中，56 个风格因子如下表所示，101 个技术指标来自 Zura Kakushadze 于 2015 年发表的论文《101 Formulaic Alphas》，分别标记为 Alpha1、Alpha2、.....、Alpha101。

表 1：候选风格因子

因子类别	因子名称
盈利因子	净资产收益率
盈利因子	销售净利率
质量因子	总资产周转率
质量因子	财务费用比例
质量因子	营业费用比例
质量因子	流动比率
规模因子	流通市值



估值因子	净市率
估值因子	现市率
估值因子	盈市率
估值因子	销市率
杠杆因子	资产负债率
杠杆因子	长期负债比率
杠杆因子	流动负债率
成长因子	每股收益增长率
成长因子	主营业务收入增长率
成长因子	净利润增长率
成长因子	净资产收益率增长率
波动因子	平均真实波幅
波动因子	真实波幅
波动因子	波动率（半月）
波动因子	波动率（月）
波动因子	波动率（周）
beta 因子	120 日 beta（相对 300 指数）
beta 因子	120 日 beta（相对 500 指数）
beta 因子	120 日 beta（相对 800 指数）
流动因子	成交额
流动因子	换手率（半月）
流动因子	换手率（月）
流动因子	换手率（周）
流动因子	日均成交量（半月）
流动因子	日均成交量（月）
流动因子	日均成交量（周）
流动因子	成交量标准差（半月）
流动因子	成交量标准差（月）
流动因子	成交量标准差（周）
流动因子	换手率
流动因子	成交量
技术因子	反转因子（日）
技术因子	反转因子（周）
技术因子	反转因子（半月）
技术因子	反转因子（月）
技术因子	日振幅
技术因子	动态买卖气指标
技术因子	动态买卖气指标均值
技术因子	乖离率（6 日）
技术因子	乖离率（12 日）
技术因子	乖离率（24 日）
技术因子	MACD 离差
技术因子	MACD 离差平均值

技术因子	MACD 值
技术因子	KDJ 之 K 值
技术因子	KDJ 之 D 值
技术因子	KDJ 之 J 值
技术因子	慢速随机指标之 K 值
技术因子	慢速随机指标之 D 值

数据来源：广发证券发展研究中心

为了训练短线、中线和长线机器学习模型，本报告考察不同的预测周期下选股因子的 IC。根据周收益率 IC、月收益率 IC 和季收益率 IC 分别筛选最显著的 50 个因子，筛选出来的因子在三个不同预测周期下的 IC 如下表所示。可以看到，不同因子适宜的预测周期有明显差异：规模、估值和 beta 等因子在较长周期的收益预测上表现较好，而 101 技术因子一般在短周期的收益预测上表现较好。即使属于相同大类的因子，也会因为计算周期和构建方法的差别而有不同的预测效果，例如，周反转因子在周收益率预测上表现较好，而月反转因子在月收益率预测上表现较好。

表 2：不同预测周期的因子 IC

因子名称	IC (周收益率)	IC (月收益率)	IC (季收益率)
流通市值	-2.63%	-4.17%	-6.00%
现市率	1.83%	3.16%	4.74%
销市率	1.88%	3.16%	4.71%
盈市率	2.44%	3.73%	4.71%
净市率	3.18%	5.07%	7.03%
反转因子 (周)	-7.06%	-4.02%	-3.16%
反转因子 (半月)	-5.35%	-4.96%	-4.13%
反转因子 (月)	-6.26%	-7.18%	-5.75%
乖离率 (6 日)	-7.20%	-3.52%	-2.78%
乖离率 (12 日)	-6.92%	-4.79%	-3.92%
乖离率 (24 日)	-6.99%	-6.79%	-5.47%
换手率	-8.69%	-10.10%	-11.27%
换手率 (周)	-7.86%	-10.25%	-11.61%
换手率 (半月)	-7.10%	-10.02%	-11.53%
换手率 (月)	-6.47%	-9.48%	-11.15%
平均真实波幅	-3.50%	-5.73%	-7.92%
成交额	-9.62%	-12.50%	-15.22%
成交量	-7.66%	-9.47%	-11.32%
日均成交量 (周)	-6.59%	-9.20%	-11.26%
日均成交量 (半月)	-5.82%	-8.79%	-11.01%
日均成交量 (月)	-5.24%	-8.20%	-10.61%
日振幅	-5.66%	-5.62%	-6.70%
真实波幅	-4.66%	-6.14%	-7.94%
波动率 (周)	-2.85%	-4.55%	-6.25%
波动率 (半月)	-3.72%	-5.67%	-7.73%

波动率（月）	-4.06%	-6.10%	-8.63%
成交量标准差（周）	-8.15%	-9.97%	-11.09%
成交量标准差（半月）	-7.63%	-10.15%	-11.52%
成交量标准差（月）	-7.08%	-9.75%	-11.45%
120 日 beta（300 指数）	-0.23%	-1.83%	-4.27%
120 日 beta（800 指数）	-0.16%	-1.83%	-4.48%
动态买卖气指标	-4.93%	-5.52%	-4.58%
动态买卖气指标均值	-3.47%	-4.82%	-4.10%
KDJ 之 J 值	-5.07%	-1.82%	-1.48%
MACD 离差	-4.73%	-5.93%	-4.99%
MACD 离差平均值	-3.92%	-5.29%	-4.52%
Alpha1	-4.22%	-1.39%	-1.26%
Alpha12	3.58%	3.45%	3.46%
Alpha13	6.28%	6.16%	5.05%
Alpha15	4.52%	4.37%	3.75%
Alpha16	6.17%	5.97%	4.76%
Alpha18	4.66%	3.36%	3.70%
Alpha19	4.63%	3.18%	2.40%
Alpha23	5.04%	2.84%	2.37%
Alpha24	4.76%	3.76%	3.85%
Alpha26	4.05%	4.37%	3.93%
Alpha27	4.06%	4.53%	3.90%
Alpha29	3.89%	4.51%	4.45%
Alpha3	3.73%	4.71%	3.87%
Alpha39	4.58%	3.16%	3.01%
Alpha4	6.54%	4.33%	3.40%
Alpha40	5.89%	7.68%	8.17%
Alpha42	5.29%	5.34%	5.88%
Alpha44	5.88%	5.96%	4.92%
Alpha46	4.39%	4.03%	3.71%
Alpha5	4.93%	2.10%	0.46%
Alpha50	5.00%	5.26%	4.39%
Alpha55	4.59%	5.13%	4.76%
Alpha66	4.30%	2.99%	2.31%
Alpha8	4.24%	1.90%	1.54%
Alpha83	4.41%	2.70%	1.85%
Alpha84	6.16%	3.96%	2.87%
Alpha87	4.60%	2.71%	2.07%
Alpha88	7.14%	7.18%	7.53%
Alpha94	4.20%	4.78%	5.87%

数据来源：Wind，广发证券发展研究中心

根据不同周期收益率的因子 IC，选择特定的选股因子构建机器学习预测模型，

不同机器学习模型所用的因子如下所示：

#### 1) 模型 S 所用因子：

反转因子(周)、反转因子(半月)、反转因子(月)、乖离率(6日)、乖离率(12日)、乖离率(24日)、换手率、换手率(周)、换手率(半月)、换手率(月)、成交额、成交量、日均成交量(周)、日均成交量(半月)、日均成交量(月)、日振幅、真实波幅、波动率(月)、成交量标准差(周)、成交量标准差(半月)、成交量标准差(月)、动态买卖气指标、KDJ之J值、MACD 离差；

Alpha101 因子：1、4、5、8、13、15、16、18、19、23、24、26、27、39、40、42、44、46、50、55、66、83、84、87、88、94；

行业 0-1 变量因子。

#### 2) 模型 M 所用因子：

流通市值、盈市率、净市率、反转因子(周)、反转因子(半月)、反转因子(月)、乖离率(6日)、乖离率(12日)、乖离率(24日)、换手率、换手率(周)、换手率(半月)、换手率(月)、平均真实波幅、成交额、成交量、日均成交量(周)、日均成交量(半月)、日均成交量(月)、日振幅、真实波幅、波动率(周)、波动率(半月)、波动率(月)、成交量标准差(周)、成交量标准差(半月)、成交量标准差(月)、动态买卖气指标、动态买卖气指标均值、MACD 离差、MACD 离差平均值；

Alpha101 因子：3、4、12、13、15、16、24、26、27、29、40、42、44、46、50、55、84、88、94；

行业 0-1 变量因子。

#### 3) 模型 L 所用因子：

流通市值、现市率、销市率、盈市率、净市率、反转因子(半月)、反转因子(月)、乖离率(12日)、乖离率(24日)、换手率、换手率(周)、换手率(半月)、换手率(月)、平均真实波幅、成交额、成交量、日均成交量(周)、日均成交量(半月)、日均成交量(月)、日振幅、真实波幅、波动率(周)、波动率(半月)、波动率(月)、成交量标准差(周)、成交量标准差(半月)、成交量标准差(月)、120 日 beta (300 指数)、120 日 beta (800 指数)、动态买卖气指标、动态买卖气指标均值、MACD 离差、MACD 离差平均值；

Alpha101 因子：3、13、15、16、18、24、26、27、29、40、42、44、46、50、55、88、94；

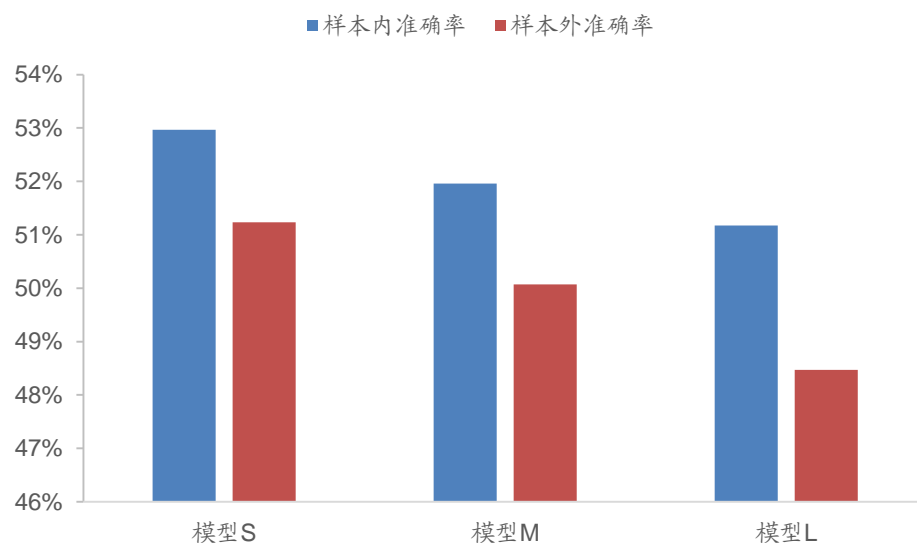
行业 0-1 变量因子。

采用 2007 年至 2012 年的股票训练样本对模型 S、模型 M 和模型 L 的超参数分别进行交叉验证，获得最佳超参数。交叉验证主要优化的模型超参数包括 XGBoost 中决策树最大深度、样本采样比例和特征采样比例。

## 3.2 不同训练窗口长度的模型表现

从不同机器学习模型的分类准确度来看，不同训练窗口长度的机器学习模型都有一定的预测能力，样本外预测准确度在50%左右（考虑到三分类问题，随机预测的分类准确度期望值为33%）。从不同的模型来看，短线模型的预测准确率相对较高，中长线模型的预测准确率相对较低。

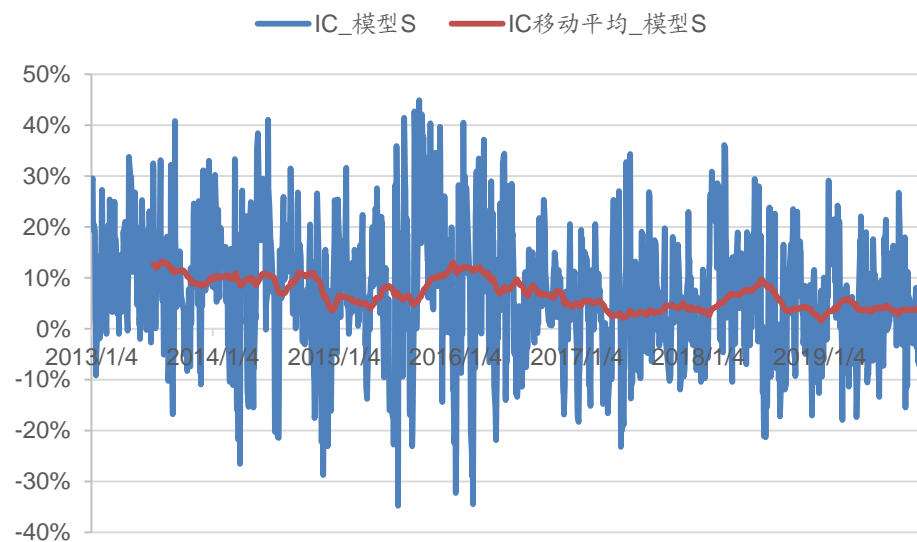
图 4：不同训练窗口长度的模型预测准确度



数据来源：Wind，广发证券发展研究中心

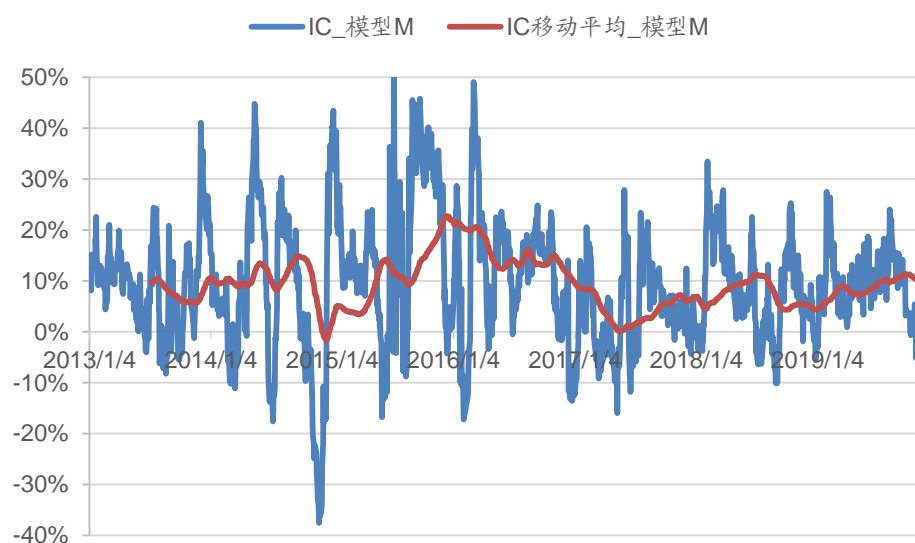
不同模型的 IC 如下图所示。模型 S 的周收 IC 均值为 7.00%，模型 M 的月 IC 均值为 9.46%，模型 L 的季 IC 均值为 11.78%。

图 5：模型 S 的 IC



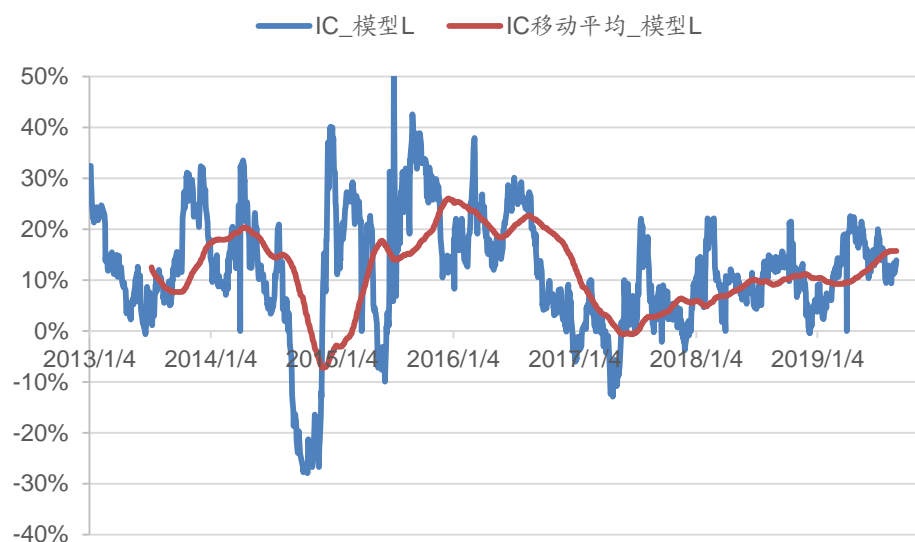
数据来源：Wind，广发证券发展研究中心

图 6: 模型M的IC



数据来源: Wind, 广发证券发展研究中心

图 7: 模型L的IC

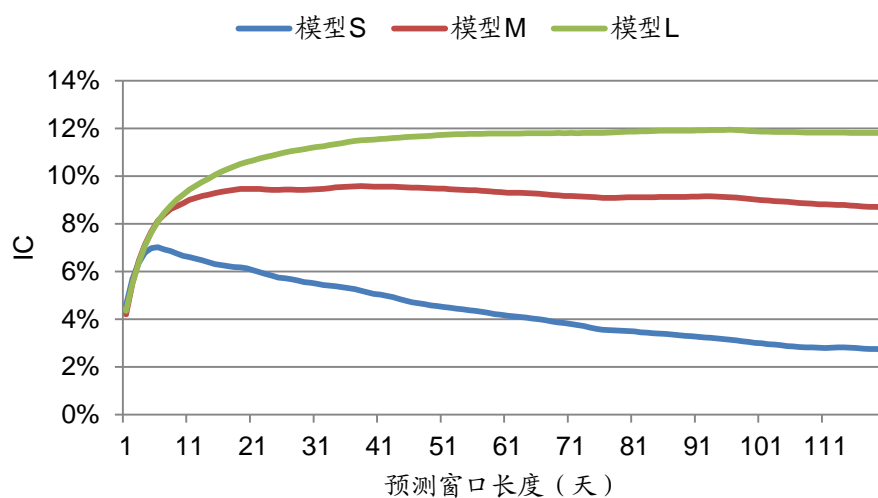


数据来源: Wind, 广发证券发展研究中心

下图展示了不同训练窗口长度机器学习模型的因子IC随股票收益率预测窗口长度变化的情况。可以看到,当预测窗口长度 $T$ 小于模型训练窗口长度时,模型的IC随着 $T$ 的增大而增大。当预测窗口长度为 $T=6$ 时,模型S的IC最高,随着 $T$ 的继续增大,模型S的IC出现了明显的衰减。当 $T=38$ 时,模型M的IC最高,此后,模型M的IC出现了衰减,但衰减不明显。当 $T=96$ 时,模型L的IC最高,而且一直到 $T=120$ 时,模型L的IC都没有明显的衰减。



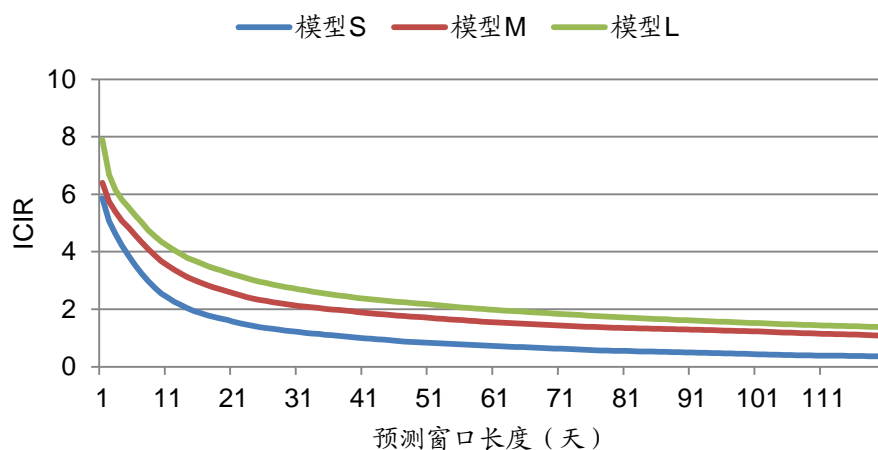
图 8: 模型IC的衰减情况



数据来源: Wind, 广发证券发展研究中心

不同机器学习模型选股因子的ICIR是单调递减的。当股票收益率预测窗口长度较小时, 机器学习因子的年化ICIR高, 随着预测窗口长度的增大, 机器学习因子的年化ICIR不断减小, 说明机器学习因子的预测能力随着预测周期的增大而衰减。

图 9: 模型年化ICIR的衰减情况



数据来源: Wind, 广发证券发展研究中心

模型S周度调仓策略的表现如下图所示, 2013年以来, 策略年化超额收益为30.85%, 最大回撤-10.86%, 信息比为3.23。

图 10: 模型S周度调仓选股表现



数据来源: Wind, 广发证券发展研究中心

模型 S 周度调仓策略每年都获得了正的超额收益, 2019 年以来的年化超额收益为 9.54%。

表 3: 模型S周度调仓策略分年度表现

年份	年化超额收益	最大回撤	多头收益	基准收益	换手率	夏普比率
2013	33.95%	-2.34%	52.12%	16.89%	37.83	4.41
2014	17.42%	-4.00%	62.26%	39.01%	40.96	2.13
2015	81.26%	-10.86%	145.55%	43.12%	40.24	3.76
2016	26.81%	-5.45%	2.68%	-17.78%	40.18	2.80
2017	13.18%	-3.80%	12.62%	-0.20%	39.89	2.25
2018	30.43%	-2.96%	-12.85%	-33.32%	40.64	4.02
2019	9.54%	-2.42%	27.45%	17.44%	35.57	1.87

数据来源: Wind, 广发证券发展研究中心

模型M月度调仓策略的表现如下图所示, 2013年以来, 策略年化超额收益为 15.80%, 最大回撤-10.66%, 信息比为1.70。

图 11: 模型M月度调仓选股表现



数据来源: Wind, 广发证券发展研究中心

除了 2017 年, 模型 M 月度调仓策略在其他年份都获得了正的超额收益, 2019 年以来的年化超额收益为 12.81%。

表 4: 模型M月度调仓策略分年度表现

年份	年化超额收益	最大回撤	多头收益	基准收益	换手率	夏普比率
2013	9.20%	-4.11%	24.73%	16.89%	9.22	1.57
2014	4.18%	-10.16%	45.27%	39.01%	9.54	0.63
2015	63.30%	-10.66%	122.62%	43.12%	10.10	4.25
2016	24.26%	-4.89%	1.23%	-17.78%	9.54	3.39
2017	-1.84%	-6.61%	-2.17%	-0.20%	9.04	-0.30
2018	9.20%	-2.75%	-27.11%	-33.32%	10.57	1.67
2019	12.81%	-3.73%	28.23%	17.44%	8.88	1.65

数据来源: Wind, 广发证券发展研究中心

模型L季度调仓策略的表现如下图所示, 2013年以来, 策略年化超额收益为 11.86%, 最大回撤-12.35%, 信息比为0.98。

图 12: 模型L季度调仓选股表现



数据来源: Wind, 广发证券发展研究中心

表 5: 模型L季度调仓策略分年度表现

年份	年化超额收益	最大回撤	多头收益	基准收益	换手率	夏普比率
2013	8.49%	-2.98%	13.63%	16.89%	2.31	1.79
2014	-0.82%	-11.16%	38.21%	39.01%	3.16	-0.10
2015	45.50%	-12.35%	95.19%	43.12%	3.12	3.63
2016	16.71%	-4.22%	-1.71%	-17.78%	3.03	3.51
2017	-2.41%	-6.90%	-2.62%	-0.20%	3.08	-0.49
2018	11.17%	-3.45%	-25.09%	-33.32%	3.20	2.33
2019	8.52%	-6.00%	24.92%	17.44%	2.97	1.12

数据来源: Wind, 广发证券发展研究中心

从以上三个模型的表现来看, 模型 S 周度调仓策略表现最好, 年化超额收益为 30.85%, 而模型 L 季度调仓表现相对较弱, 但年化超额收益也有 11.86%。从换手率来看, 模型 S 周度调仓策略的年化换手率为 40.52 倍, 而模型 L 季度调仓策略的年化换手率为 3.04 倍。

### 3.3 策略在不同调仓周期下的表现

通过前文对 IC 和 ICIR 的分析, 可以看到, 机器学习模型的训练窗口长度和调仓周期不一定要匹配, 机器学习模型对于训练窗口长度之外的调仓周期也有显著的 IC。以模型 S 为例, 当股票收益窗口长度 T 为 5 天时, IC 为 7.00%, 当窗口长度 T 为 20 天时, IC 为 6.13%, 当窗口长度 T 为 60 天时, IC 为 4.18%。本节考察当调仓周期与模型训练窗口长度不匹配的时候, 机器学习模型的选股性能。

分别按照周频、月频和季频进行调仓, 模型 S 的选股表现如下图所示。其中周频调仓策略的年化超额收益率最高, 为 30.85%; 月频调仓策略的年化超额收益率为 15.51%; 季频调仓策略的年化超额收益率为 5.62%。随着调仓频率的降低, 策略的

换手率和夏普比率明显下降，策略的最大回撤明显增大。可以看到，当调仓频率从周频降低到季频之后，模型 S 的选股性能降低明显。

图 13: 模型S不同调仓周期选股表现



数据来源: Wind, 广发证券发展研究中心

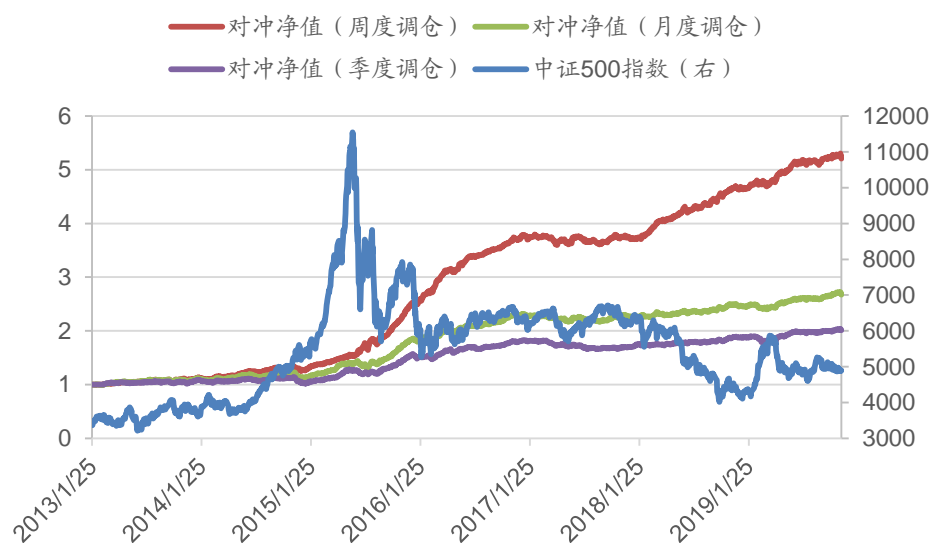
表 6: 模型S不同周期策略表现汇总

	周度调仓	月度调仓	季度调仓
年化收益	30.85%	15.51%	5.62%
最大回撤	-10.86%	-13.56%	-15.13%
夏普比率	3.23	1.58	0.69
年换手率	40.52	10.69	3.52

数据来源: Wind, 广发证券发展研究中心

在不同的调仓周期下，模型 M 的选股表现如下图所示。从结果来看，虽然模型 M 的训练周期长度为 20 天，但模型 M 在周度调仓时具有更好的策略表现。这个和上文对 ICIR 的分析结果相符合：模型 M 的年化 ICIR 是随着调仓频率的降低而不断下降的。

图 14: 模型M不同调仓周期选股表现



数据来源：Wind，广发证券发展研究中心

周度调仓下，模型 M 的表现不如模型 S，但在月度调仓或季度调仓时，模型 M 的表现相比模型 S 都有一定的提升。这是由于模型 S 在预测月度收益率或季度收益率时的预测准确度衰减较快引起的。

表 7: 模型M不同周期策略表现汇总

	周度调仓	月度调仓	季度调仓
年化收益	28.59%	15.80%	11.01%
最大回撤	-7.35%	-10.66%	-11.01%
夏普比率	3.20	1.70	1.15
年换手率	33.64	9.78	3.32

数据来源：Wind，广发证券发展研究中心

在不同的调仓周期下，模型 L 的选股表现如下图所示。与此前的两个模型类似，虽然模型 L 的训练窗口长度为 60 天，但模型 L 在周频调仓时的选股表现明显优于月频调仓和季频调仓下的结果。值得一提的是，在月度调仓下，模型 L 的表现甚至优于模型 M。



图 15: 模型L不同调仓周期选股表现



数据来源: Wind, 广发证券发展研究中心

表 8: 模型L不同周期策略表现汇总

	周度调仓	月度调仓	季度调仓
年化收益	22.89%	18.29%	11.86%
最大回撤	-9.09%	-10.85%	-12.35%
夏普比率	2.74	2.16	0.98
年换手率	27.98	8.57	3.04

数据来源: Wind, 广发证券发展研究中心

总体而言, 模型 S、模型 M 和模型 L 都是在周度调仓下夏普比率最高, 在月度调仓和季度调仓下模型的夏普比率都有明显的衰减。而且, 在周度调仓下, 模型 S 的夏普比率高于模型 M 和模型 L, 也就是说, 周度调仓下短线预测模型表现较好; 在月度调仓和季度调仓下, 模型 M 和模型 L 的表现要优于模型 S, 也就是说, 月度调仓和季度调仓下, 中长线预测模型表现较好。

表 9: 不同调仓频率下模型夏普比率比较

夏普比率	周度调仓	月度调仓	季度调仓
模型 S	3.23	1.58	0.69
模型 M	3.20	1.70	1.15
模型 L	2.74	2.16	0.98

数据来源: Wind, 广发证券发展研究中心

### 3.4 多模型集成策略的表现

不同训练窗口长度的机器学习模型的选股表现具有一定的相关性，如下表所示。可以看到，模型 S 和模型 M 的选股因子相关系数为 0.602，模型 M 和模型 L 的相关系数为 0.623，而模型 S 和模型 L 的差异较大，相关系数低，为 0.360。

表 10: 不同模型股票收益率预测打分相关性

	模型 S	模型 M	模型 L
模型 S	1	0.602	0.360
模型 M		1	0.623
模型 L			1

数据来源: Wind, 广发证券发展研究中心

考虑到不同训练窗口长度的模型在周频调仓时都有不错的选股表现，而且不同模型之间的打分相关性比较低，因此，可以将模型 S、模型 M 和模型 L 集成起来，进行多机器学习模型集成选股。在本报告中，模型 S、模型 M 和模型 L 在集成模型中具有相同的权重。

图 16: 机器学习多模型集成策略选股表现



数据来源: Wind, 广发证券发展研究中心

在周度调仓下，多模型集成策略的分年度表现如下表所示。策略的年化超额收益为 38.97%，最大回撤为 -10.00%，夏普比率为 3.89，年化换手率为 32.27 倍。2019 年以来，多模型集成策略的年化超额收益为 16.22%。

表 11: 多模型集成周度调仓策略分年度表现

年份	年化超额收益	最大回撤	多头收益	基准收益	换手率	夏普比率
2013	24.65%	-3.72%	41.90%	16.89%	31.02	3.87
2014	22.63%	-9.54%	69.68%	39.01%	32.36	2.56
2015	128.00%	-6.34%	215.29%	43.12%	32.38	6.07
2016	50.09%	-3.93%	21.39%	-17.78%	31.27	5.19
2017	4.28%	-5.25%	3.79%	-0.20%	29.72	0.71
2018	38.54%	-3.32%	-7.19%	-33.32%	33.74	5.93
2019	16.22%	-1.91%	35.34%	17.44%	28.19	2.60

数据来源: Wind, 广发证券发展研究中心

在周度调仓下, 模型S、模型M、模型L和多模型集成策略的表现对比如下表所示。与单模型策略相比, 多模型集成策略具有明显更高的年化超额收益和夏普比, 而且换手率比模型S和模型M要低。

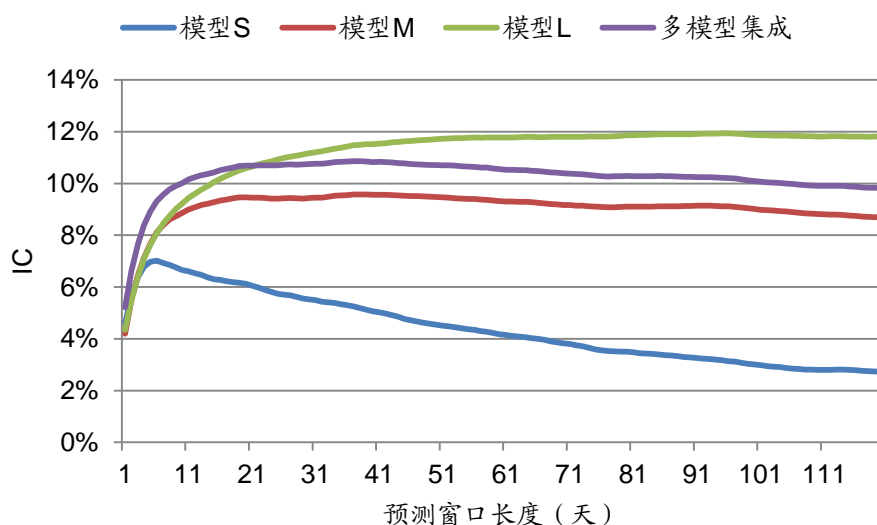
表 12: 不同周期机器学习选股模型对冲表现

	模型 S	模型 M	模型 L	多模型集成
年化超额收益	30.85%	28.59%	22.89%	<b>38.97%</b>
最大回撤	-10.86%	-7.35%	-9.09%	-10.00%
夏普比率	3.23	3.20	2.74	<b>3.89</b>
年换手率	40.52	33.64	27.98	<b>32.27</b>

数据来源: Wind, 广发证券发展研究中心

从多模型模型集成的复合因子 IC 来看, 当预测窗口长度不超过 21 天时, 多模型集成的 IC 显著高于单一模型的表现。而当预测窗口长度继续增加时, 多模型集成策略也出现了明显的衰减, 这可能和短线预测模型的预测性能衰减过快有关。

图 17: 多模型集成策略的IC与预测窗口长度的关系



数据来源: Wind, 广发证券发展研究中心

## 四、总结与展望

本报告通过实证研究，分析了机器学习模型的性能与训练窗口长度、股票收益预测窗口长度的关系，研究了机器学习模型在不同调仓周期下选股的可行性。与因子的时效性类似，机器学习选股因子也具有时效性。当调仓频率较高时，采用短线预测的机器学习模型选股性能较好；当调仓频率较低时，采用中长线预测的机器学习模型的选股性能较好。由于机器学习模型的性能随着预测窗口的增长而衰减，增加调仓频率一般会获得更好的收益表现。在千分之三的换仓成本下，相对月频或季频调仓，周频调仓是机器学习模型更合适的一个调仓频率。

采用不同训练窗口长度训练出来的模型有一定的差异，通过将不同训练窗口长度的机器学习模型集成起来进行选股，与单模型选股相比性能有了显著提升。在周频调仓下，模型S、模型M和模型L三模型集成选股策略的年化超额收益为38.97%，夏普比率为3.89，显著优于单模型的选股表现，而且模型的换手率也比中短线预测模型的选股换手率要低。

## 风险提示

策略模型并非百分百有效，市场结构及交易行为的改变以及类似交易参与者的增多有可能使得策略失效。

## 广发金融工程研究小组

罗 军：首席分析师，华南理工大学硕士，从业 14 年，2010 年进入广发证券发展研究中心。

安 宁 宁：联席首席分析师，暨南大学硕士，从业 12 年，2011 年进入广发证券发展研究中心。

史 庆 盛：资深分析师，华南理工大学硕士，从业 8 年，2011 年进入广发证券发展研究中心。

马 普 凡：资深分析师，英国拉夫堡大学硕士，从业 9 年，2014 年进入广发证券发展研究中心。

张 超：资深分析师，中山大学硕士，从业 7 年，2012 年进入广发证券发展研究中心。

文 巧 钧：资深分析师，浙江大学博士，从业 4 年，2015 年进入广发证券发展研究中心。

陈 原 文：资深分析师，中山大学硕士，从业 4 年，2015 年进入广发证券发展研究中心。

樊 瑞 铎：资深分析师，南开大学硕士，从业 4 年，2015 年进入广发证券发展研究中心。

李 豪：资深分析师，上海交通大学硕士，从业 3 年，2016 年进入广发证券发展研究中心。

郭 圳 滨：研究助理，中山大学硕士，2018 年进入广发证券发展研究中心。

## 广发证券—行业投资评级说明

买入：预期未来 12 个月内，股价表现强于大盘 10%以上。

持有：预期未来 12 个月内，股价相对大盘的变动幅度介于-10%~+10%。

卖出：预期未来 12 个月内，股价表现弱于大盘 10%以上。

## 广发证券—公司投资评级说明

买入：预期未来 12 个月内，股价表现强于大盘 15%以上。

增持：预期未来 12 个月内，股价表现强于大盘 5%-15%。

持有：预期未来 12 个月内，股价相对大盘的变动幅度介于-5%~+5%。

卖出：预期未来 12 个月内，股价表现弱于大盘 5%以上。

## 联系我们

	广州市	深圳市	北京市	上海市	香港
地址	广州市天河区马场路 26 号广发证券大厦 35 楼	深圳市福田区益田路 6001 号太平金融大厦 31 层	北京市西城区月坛北 街 2 号月坛大厦 18 层	上海市浦东新区世纪 大道 8 号国金中心一 期 16 楼	香港中环干诺道中 111 号永安中心 14 楼 1401-1410 室
邮政编码	510627	518026	100045	200120	
客服邮箱	gfyf@gf.com.cn				

## 法律主体声明

本报告由广发证券股份有限公司或其关联机构制作，广发证券股份有限公司及其关联机构以下统称为“广发证券”。本报告的分销依据不同国家、地区的法律、法规和监管要求由广发证券于该国家或地区的具有相关合法合规经营资质的子公司/经营机构完成。

广发证券股份有限公司具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管，负责本报告于中国（港澳台地区除外）的分销。

广发证券（香港）经纪有限公司具备香港证监会批复的就证券提供意见（4 号牌照）的牌照，接受香港证监会监管，负责本报告于中国香港地区的分销。

本报告署名研究人员所持中国证券业协会注册分析师资质信息和香港证监会批复的牌照信息已于署名研究人员姓名处披露。

## 重要声明

广发证券股份有限公司及其关联机构可能与本报告中提及的公司寻求或正在建立业务关系，因此，投资者应当考虑广发证券股份有限公司及其关联机构因可能存在的潜在利益冲突而对本报告的独立性产生影响。投资者不应仅依据本报告内容作出任何投资决策。

本报告署名研究人员、联系人（以下均简称“研究人员”）针对本报告中相关公司或证券的研究分析内容，在此声明：（1）本报告的全部分析结论、研究观点均精确反映研究人员于本报告发出当日的关于相关公司或证券的所有个人观点，并不代表广发证券的立场；（2）研究人员的部分或全部的报酬无论在过去、现在还是将来均不会与本报告所述特定分析结论、研究观点具有直接或间接的联系。

研究人员制作本报告的报酬标准依据研究质量、客户评价、工作量等多种因素确定，其影响因素亦包括广发证券的整体经营收入，该等经营收入部分来源于广发证券的投资银行类业务。

本报告仅面向经广发证券授权使用的客户/特定合作机构发送，不对外公开发布，只有接收人才可以使用，且对于接收人而言具有保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。在特定国家或地区传播或者发布本报告可能违反当地法律，广发证券并未采取任何行动以允许于该等国家或地区传播或者分销本报告。

本报告所提及证券可能不被允许在某些国家或地区内出售。请注意，投资涉及风险，证券价格可能会波动，因此投资回报可能会有所变化，过去的业绩并不保证未来的表现。本报告的内容、观点或建议并未考虑任何个别客户的具体投资目标、财务状况和特殊需求，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券认为可靠，但广发证券不对其准确性、完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策，如有需要，应先咨询专业意见。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券的立场。广发证券的销售人员、交易员或其他专业人士可能以书面或口头形式，向其客户或自营交易部门提供与本报告观点相反的市场评论或交易策略，广发证券的自营交易部门亦可能会有与本报告观点不一致，甚至相反的投资策略。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且无需另行通告。广发证券或其证券研究报告业务的相关董事、高级职员、分析师和员工可能拥有本报告所提及证券的权益。在阅读本报告时，收件人应了解相关的权益披露（若有）。

本研究报告可能包括和/或描述/呈列期货合约价格的事实历史信息（“信息”）。请注意此信息仅供用作组成我们的研究方法/分析中的部分论点/依据/证据，以支持我们对所述相关行业/公司的观点的结论。在任何情况下，它并不（明示或暗示）与香港证监会第5类受规管活动（就期货合约提供意见）有关联或构成此活动。

## 权益披露

(1) 广发证券（香港）跟本研究报告所述公司在过去12个月内并没有任何投资银行业务的关系。

## 版权声明

未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。