



清华大学  
Tsinghua University

# 通过机器学习进行的实证资产定价

Empirical Asset Pricing via Machine Learning



汇报人：蒋子函、文家伟

# 文章结构

## 0. 为什么使用机器学习？

### 1. 方法论

1.1 通过验证进行样本拆分和调整

1.2 简单线性

1.3 惩罚线性

1.4 降维：PCR和PLS

1.5 广义线性

1.6 强化回归树和随机森林

1.7 神经网络

1.8 效果评价

1.9 变量重要性和边际关系

### 2. 美国股票的实证研究

2.1 数据和总体模型

2.2 个股截面研究

2.3 哪些协变量重要

2.4 组合预测

### 3. 结论

2020年Review of Financial Studies刊出，作者中有两位是在Booth的华人顾诗颢、修大成，另一位则是在耶鲁和 AQR任职的Bryan Kelly。

文章对使用机器学习做实证资产定价的经典问题（即测度资产的风险溢价）进行了可比较的分析，表明使用机器学习的投资者可获得巨大的经济收益，甚至可比现有文献中基于回归的策略表现高出一倍。该文确定出**最佳的模型（树和神经网络）**，并追踪到它们预测的增量收益来自于预测因子的交互，这恰恰是其他方法所错失的部分。所有的方法找出的**最佳预测信号集（因子）**是一致的，其中包含了动量、流动性、波动性的相关变量。

将机器学习应用于金融市场，在量化投资领域中是一个老生常谈的问题，它的有效性和可靠性也一直存在不少争论。而在金融学术界，之前只有寥寥数篇文章涉及到一部分机器学习中的方法，常态化地使用机器学习则是在最近才出现的苗头。从学术角度来说，此文章是比较有代表性的一篇，基本上把机器学习所有的方法全部尝试了一遍。

# 0. 为什么使用机器学习?



首先，是实证资产定价领域本身的特点，让人将它和机器学习进行联想：

## 本质是预测问题

测度一项资产的风险溢价，本质上是一个**预测问题**——风险溢价就是未来实现的超额收益率的**条件期望**；

## 变量集合大

对风险溢价来说，备选的**变量集合非常大**；

## 模型形式不确定

高维预测因子进入风险溢价的函数**形式是不确定的**。

### 机器学习特点

- 1: **多样性**。机器学习有各种不同的算法，可以搜索很大的函数空间；
- 2: **灵活性**。机器的各种算法可以对复杂的非线性关系进行近似；
- 3: **标准化**。机器学习具有参数惩罚和模型选择标准，哪怕函数形式非常广泛，也可以避免过拟合偏差和错误发现。



### 研究目的

一项资产的超额收益可表述成一个加性的预测误差模型 (additive prediction error model) :

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}$$

其中:  $E_t(r_{i,t+1}) = g^*(z_{i,t})$

$i$ : 股票

$t$ : 时间 (月数)

文章的目的是, 分离出一个  $E_t(r_{i,t+1})$  的表示形式, 它是预测因子的函数, 并对  $r_{i,t+1}$  有最大的样本外解释力 (测试集准确性)。其中预测因子是高维的  $z_{i,t}$ , 函数形式可以是灵活的条件期望收益率函数  $g^*(\cdot)$ , 在这里假设函数形式与  $i$  和  $t$  均无关。

风险溢价度量相当于条件期望  $E(r_{i,t+1} | F_t)$  的最佳近似, 其中,  $r_{i,t+1}$  是资产超过无风险利率的回报, 而  $F_t$  是市场参与者真实的和不可观察的信息集。



数据共有60年：1957年-2016年

（训练样本）训练集：18年（1957年 - 1974年）

（验证样本）验证集：12年（1975年 - 1986年）

（样本外测试）测试集：30年（1987年 - 2016年）

# 线性模型

---

$$g(z_{i,t}; \theta) = z'_{i,t} \theta$$

简单线性模型规定，条件期望可以由原始预测变量和参数向量  $\theta$  的线性函数近似。

$$\mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (r_{i,t+1} - g(z_{i,t}; \theta))^2$$

OLS线性回归的基本原则：最优拟合曲线应该使各点到直线的距离的平方和（即残差平方和，简称RSS）最小。  
实际应用经常会用均方差（MSE）。

# 线性模型优化

---

首先优化的是损失函数：

$$\mathcal{L}_H(\theta) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T H(r_{i,t+1} - g(z_{i,t}; \theta), \xi)$$

其中：

$$H(x; \xi) = \begin{cases} x^2, & \text{if } |x| \leq \xi; \\ 2\xi|x| - \xi^2, & \text{if } |x| > \xi. \end{cases}$$

$\xi$ 是HuberLoss的参数，当预测偏差小于 $\xi$ 时，它采用平方误差，当预测偏差大于 $\xi$ ，采用线性误差。相比于最小二乘的线性回归，Huber Loss降低了对异常点的惩罚程度，是一种常用的robust regression的损失函数。

# 线性模型优化——正则化

在众多的自变量面前，简单的线性模型注定会失败。  
因此给使用弹性网络模型加上罚项：

$$\mathcal{L}(\theta; \cdot) = \mathcal{L}(\theta) + \phi(\theta; \cdot)$$

其中：

$$\phi(\theta; \lambda, \rho) = \lambda(1 - \rho) \sum_{j=1}^P |\theta_j| + \frac{1}{2} \lambda \rho \sum_{j=1}^P \theta_j^2.$$

Elastic Net将Lasso和Ridge组成一个具有两种惩罚因素的单一模型：一个与L1范数成比例，另外一个与L2范数成比例。

使用这种方式方法所得到的模型就像纯粹的Lasso回归一样可以提取特征，但同时具有与岭回归提供的一样的正则化能力。



# 线性模型优化——降维

## 主成分回归

用一系列低维的“成分”来表示高维的预测因子，过程中尽可能不损失原来的信息。再用得到的一系列成分对目标变量做回归。PCR仅仅是在预测因子一侧做变换，要求尽可能多地保留预测因子的信息，没有考虑要预测的目标变量，因此也没有考虑找出的“成分”的预测能力。

## 偏最小二乘

PLS考虑了预测因子的预测能力，它先用每个预测因子对目标变量做单变量预测回归，再以得到的系数为权重，将所有的预测变量做加权平均，得到第一个成分，再接着，把预测因子和目标变量都对之前得到的成分做正交化，再不断重复之前的过程，最终也得到一系列成分。

## 广义线性模型

相较于线性模型，广义线性模型更具实用性。模型将预测误差分解为近似误差+估计误差+固有误差

$$r_{i,t+1} - \hat{r}_{i,t+1} = \underbrace{g^*(z_{i,t}) - g(z_{i,t}; \theta)}_{\text{approximation error}} + \underbrace{g(z_{i,t}; \theta) - g(z_{i,t}; \hat{\theta})}_{\text{estimation error}} + \underbrace{\epsilon_{i,t+1}}_{\text{intrinsic error}}$$

广义线性模型的期望函数为：

$$g(z; \theta, p(\cdot)) = \sum_{j=1}^P p(z_j)' \theta_j$$

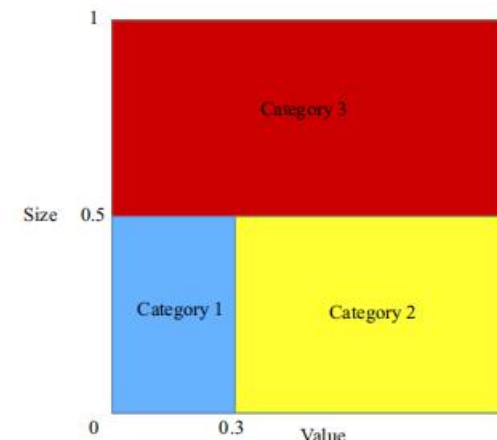
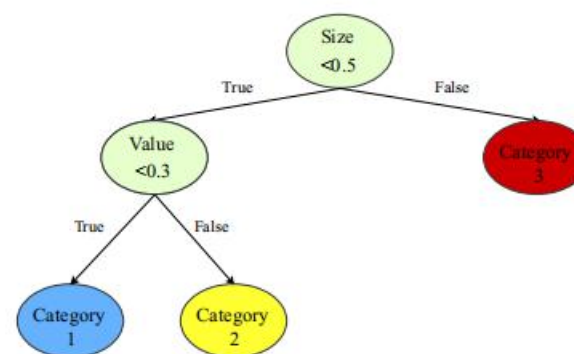
惩罚函数则选择了“Group LASSO”的形式：

$$\phi(\theta; \lambda, K) = \lambda \sum_{j=1}^P \left( \sum_{k=1}^K \theta_{j,k}^2 \right)^{1/2}$$

### Boosted regression trees and random forests

树模型的灵活度非常高，很容易发生过拟合，在树模型的基础上，可以用集成（ensemble）方法提高稳健性，有两种常用的集成方法：

- Boosting法。将很多“极其简单”的树加总在一起，文章用的是比较流行的梯度提升回归树（GBRT）；
- Bagging法。对数据集不断进行 Bootstrap 抽样，针对每次抽样出的数据子集训练一棵树，最后将所有的树综合在一起，这就是随机森林（random forest）。



**Figure 1**  
**Regression tree example**

This figure presents the diagrams of a regression tree (left) and its equivalent representation (right) in the space of two characteristics (size and value). The terminal nodes of the tree are colored in blue, yellow, and red. Based on their values of these two characteristics, the sample of individual stocks is divided into three categories.

**Algorithm 2: Classification and Regression Tree**

Initialize the stump.  $C_1(0)$  denotes the range of all covariates,  $C_l(d)$  denote the  $l$ -th node of depth  $d$ .  
**for**  $d$  from 1 to  $L$  **do**  
  **for**  $i$  in  $\{C_l(d-1), l=1, \dots, 2^{d-1}\}$  **do**  
    i) For each feature  $j=1, 2, \dots, P$ , and each threshold level  $\alpha$ , define a split as  $s=(j, \alpha)$ , which divides  $C_l(d-1)$  into  $C_{left}$  and  $C_{right}$ :  

$$C_{left}(s) = \{z_j \leq \alpha\} \cap C_l(d-1); \quad C_{right}(s) = \{z_j > \alpha\} \cap C_l(d-1),$$
  
    where  $z_j$  denotes the  $j$ th covariate.  
    ii) Define the impurity function:  

$$\mathcal{L}(C, C_{left}, C_{right}) = \frac{|C_{left}|}{|C|} H(C_{left}) + \frac{|C_{right}|}{|C|} H(C_{right}), \text{ where}$$

$$H(C) = \frac{1}{|C|} \sum_{z_{i,t} \in C} (r_{i,t+1} - \theta)^2, \quad \theta = \frac{1}{|C|} \sum_{z_{i,t} \in C} r_{i,t+1},$$
  
    and  $|C|$  denotes the number of observations in set  $C$ .  
    iii) Select the optimal split:  

$$s^* \leftarrow \underset{s}{\operatorname{argmin}} \mathcal{L}(C(s), C_{left}(s), C_{right}(s)).$$
  
    iv) Update the nodes:  

$$C_{2l-1}(d) \leftarrow C_{left}(s^*), \quad C_{2l}(d) \leftarrow C_{right}(s^*).$$
  
  **end**  
**end**

**Result:** The output of a regression tree is given by:

$$g(z_{i,t}; \theta, L) = \sum_{k=1}^{2^L} \theta_k \mathbf{1}\{z_{i,t} \in C_k(L)\}, \text{ where } \theta_k = \frac{1}{|C_k(L)|} \sum_{z_{i,t} \in C_k(L)} r_{i,t+1}.$$

For a single binary complete regression tree  $\mathcal{T}$  of depth  $L$ , the VIP for the covariate  $z_j$  is

$$\text{VIP}(z_j, \mathcal{T}) = \sum_{d=1}^{L-1} \sum_{i=1}^{2^{d-1}} \Delta \text{im}(C_i(d-1), C_{2i-1}(d), C_{2i}(d)) \mathbf{1}\{z_j \in \mathcal{T}(i, d)\},$$

where  $\mathcal{T}(i, d)$  represents the covariate on the  $i$ -th (internal) node of depth  $d$ , which splits  $C_i(d-1)$  into two sub-regions  $\{C_{2i-1}(d), C_{2i}(d)\}$ , and  $\Delta \text{im}(\cdot, \cdot, \cdot)$  is defined by:

$$\Delta \text{im}(C, C_{left}, C_{right}) = H(C) - \mathcal{L}(C, C_{left}, C_{right}).$$

**Algorithm 3: Random Forest**

**for**  $b$  from 1 to  $B$  **do**  
  Generate Bootstrap samples  $\{(z_{i,t}, r_{i,t+1}), (i, t) \in \text{Bootstrap}(b)\}$  from the original dataset, for which a tree is grown using Algorithm 2. At each step of splitting, use only a random subsample, say  $\sqrt{P}$  or any specific number, of all features. Write the resulting  $b$ th tree as:

$$\hat{g}_b(z_{i,t}; \hat{\theta}_b, L) = \sum_{k=1}^{2^L} \theta_b^{(k)} \mathbf{1}\{z_{i,t} \in C_k(L)\}.$$

**end**

**Result:** The final random forest output is given by the average of the outputs of all  $B$  trees.

$$\hat{g}(z_{i,t}; L, B) = \frac{1}{B} \sum_{b=1}^B \hat{g}_b(z_{i,t}; \hat{\theta}_b, L).$$

**Algorithm 4: Gradient Boosted Tree**

Initialize the predictor as  $\hat{g}_0(\cdot) = 0$ ;

**for**  $b$  from 1 to  $B$  **do**

  Compute for each  $i=1, 2, \dots, N$  and  $t=1, 2, \dots, T$ , the negative gradient of the loss function  $l(\cdot, \cdot)$ :

$$\varepsilon_{i,t+1} \leftarrow -\frac{\partial l(r_{i,t+1}, g)}{\partial g} \Big|_{g=\hat{g}_{b-1}(z_{i,t})}.$$

  Grow a (shallow) regression tree of depth  $L$  with dataset  $\{(z_{i,t}, \varepsilon_{i,t+1}) : \forall i, \forall t\}$

$$\hat{f}_b(\cdot) \leftarrow g(z_{i,t}; \theta, L).$$

  Update the model by

$$\hat{g}_b(\cdot) \leftarrow \hat{g}_{b-1}(\cdot) + \nu \hat{f}_b(\cdot),$$

  where  $\nu \in (0, 1]$  is a tuning parameter that controls the step length.

**end**

**Result:** The final model output is

$$\hat{g}_B(z_{i,t}; B, \nu, L) = \sum_{b=1}^B \nu \hat{f}_b(\cdot).$$

### Neural Networks

---

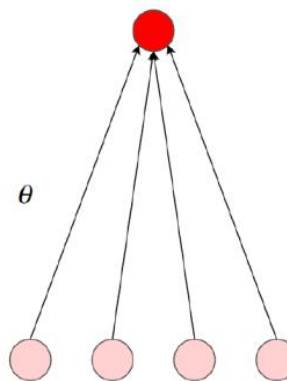
文章使用了含 1-5 层隐藏层的全连接前馈网络结构，其中 NN1的隐含层有32个节点，NN2分别有32、16个节点，以此类推，NN5有32、16、8、4、2个节点。

ReLU作为激活函数。

由于神经网络的高度非线性和非凸性，需使用随机梯度下降算法（SGD）进行训练。此外，文章在加入了 L1正则项的同时，还用了4种其他手段进行正则化：learning rate shrinkage, early stopping, batch normalization, 以及 ensembles。

Output Layer

Input Layer



Output Layer

Hidden Layer

Input Layer

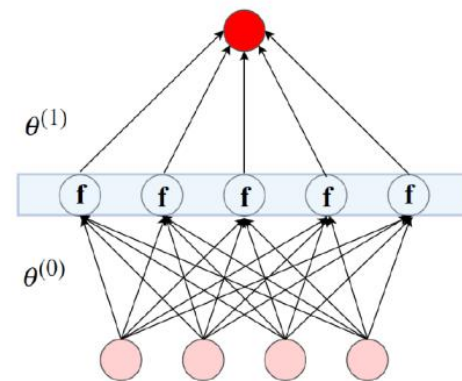


Figure 2  
Neural networks



---

**Algorithm 6:** Early Stopping

---

Initialize  $j = 0$ ,  $\epsilon = \infty$  and select the patience parameter  $p$ .

**while**  $j < p$  **do**

    Update  $\theta$  using the training algorithm (e.g., the steps inside the while loop of Algorithm 5 for  $h$  steps).

    Calculate the prediction error from the validation sample, denoted as  $\epsilon'$ .

**if**  $\epsilon' < \epsilon$  **then**

$j \leftarrow 0$ .

$\epsilon \leftarrow \epsilon'$ .

$\theta' \leftarrow \theta$ .

**else**

$j \leftarrow j + 1$ .

**end**

**end**

**Result:** The final parameter estimate is  $\theta'$ .

---

---

**Algorithm 5:** Adam for Stochastic Gradient Descent (SGD)

---

Initialize the parameter vector  $\theta_0$ . Set  $m_0 = 0, v_0 = 0, t = 0$ .

**while**  $\theta_t$  *not converged* **do**

$t \leftarrow t + 1$ .

$g_t \leftarrow \nabla_{\theta} \mathcal{L}_t(\theta; \cdot) |_{\theta=\theta_{t-1}}$ .

$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$ .

$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t \odot g_t$ .<sup>*a*</sup>

$\hat{m}_t \leftarrow m_t / (1 - (\beta_1)^t)$ .

$\hat{v}_t \leftarrow v_t / (1 - (\beta_2)^t)$ .

$\theta_t \leftarrow \theta_{t-1} - \alpha \hat{m}_t \oslash (\sqrt{\hat{v}_t} + \epsilon)$ .

**end**

**Result:** The final parameter estimate is  $\theta_t$ .

---

---

**Algorithm 6:** Early Stopping

---

Initialize  $j = 0$ ,  $\epsilon = \infty$  and select the patience parameter  $p$ .

**while**  $j < p$  **do**

    Update  $\theta$  using the training algorithm (e.g., the steps inside the while loop of Algorithm 5 for  $h$  steps).

    Calculate the prediction error from the validation sample, denoted as  $\epsilon'$ .

**if**  $\epsilon' < \epsilon$  **then**

$j \leftarrow 0$ .

$\epsilon \leftarrow \epsilon'$ .

$\theta' \leftarrow \theta$ .

**else**

$j \leftarrow j + 1$ .

**end**

**end**

**Result:** The final parameter estimate is  $\theta'$ .

---

---

**Algorithm 7:** Batch Normalization (for one Activation over one Batch)

---

Input: Values of  $x$  for each activation over a batch  $\mathcal{B} = \{x_1, x_2, \dots, x_N\}$ .

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{N} \sum_{i=1}^N x_i$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{N} \sum_{i=1}^N (x_i - \mu_{\mathcal{B}})^2$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta := \text{BN}_{\gamma, \beta}(x_i)$$

**Result:**  $\{y_i = \text{BN}_{\gamma, \beta}(x_i) : i = 1, 2, \dots, N\}$ .

---

### 数据准备

---

纽交所、美交所和纳斯达克上市的所有公司的每月总个人股本回报。样本开始于1957年3月（标普500起始日），结束于2016年12月，总共60年。样本中股票数量近30000支，每月平均股票数量超过6200支。以国债利率作为计算个人超额收益的无风险利率。

根据股票收益率截面数据构建预测特征集合。包括94项特征（其中61项每年更新，13项每季度更新，20项每月更新）。此外还包括74个行业虚拟变量，对应于标准工业分类（SIC）代码的前两位数字。

构建了8个宏观经济预测因子，包括股息价格比（dp）、收益价格比（ep）、账面市盈率（bm）、净权益扩张（NTI）、国库券利率（tb1）、期限利差（tms）、违约利差（dfy）和股票方差（svar）。

将94个公司层面特征和8个宏观预测因子分别做交互项，最终有 $94 \times (8+1) + 74 = 920$ 个变量。



## 模型比较

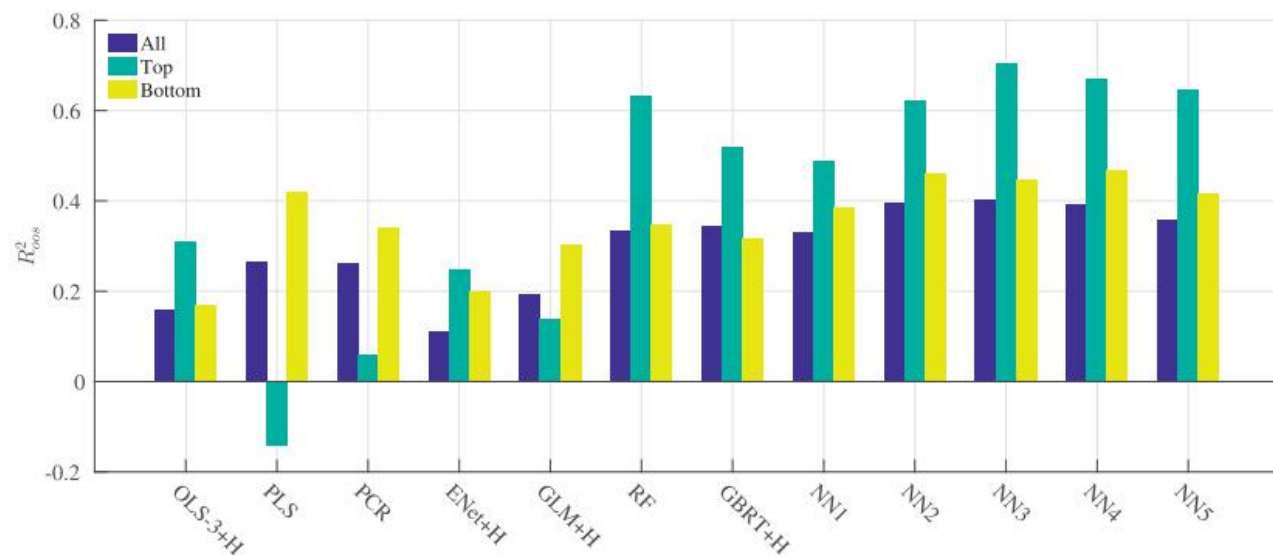
该文通过计算样本外 $R^2$ 评价模型

$$R_{\text{oos}}^2 = 1 - \frac{\sum_{(i,t) \in \mathcal{T}_3} (r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t) \in \mathcal{T}_3} r_{i,t+1}^2}$$

右图展示了各模型的月度 $R_{\text{oos}}^2$ ，并分别在全样本  
市值最大/最小的 1000 个股票中进行检验：

**Table 1****Monthly out-of-sample stock-level prediction performance (percentage  $R_{\text{oos}}^2$ )**

	OLS +H	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
All	-3.46	0.16	0.27	0.26	0.11	0.19	0.33	0.34	0.33	0.39	0.40	0.39	0.36
Top 1,000	-11.28	0.31	-0.14	0.06	0.25	0.14	0.63	0.52	0.49	0.62	0.70	0.67	0.64
Bottom 1,000	-1.30	0.17	0.42	0.34	0.20	0.30	0.35	0.32	0.38	0.46	0.45	0.47	0.42



## 模型比较

该文通过计算样本外 $R^2$ 评价模型

$$R_{\text{oos}}^2 = 1 - \frac{\sum_{(i,t) \in \mathcal{T}_3} (r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t) \in \mathcal{T}_3} r_{i,t+1}^2}$$

以年为时间跨度，再次比较各模型的 $R_{\text{oos}}^2$

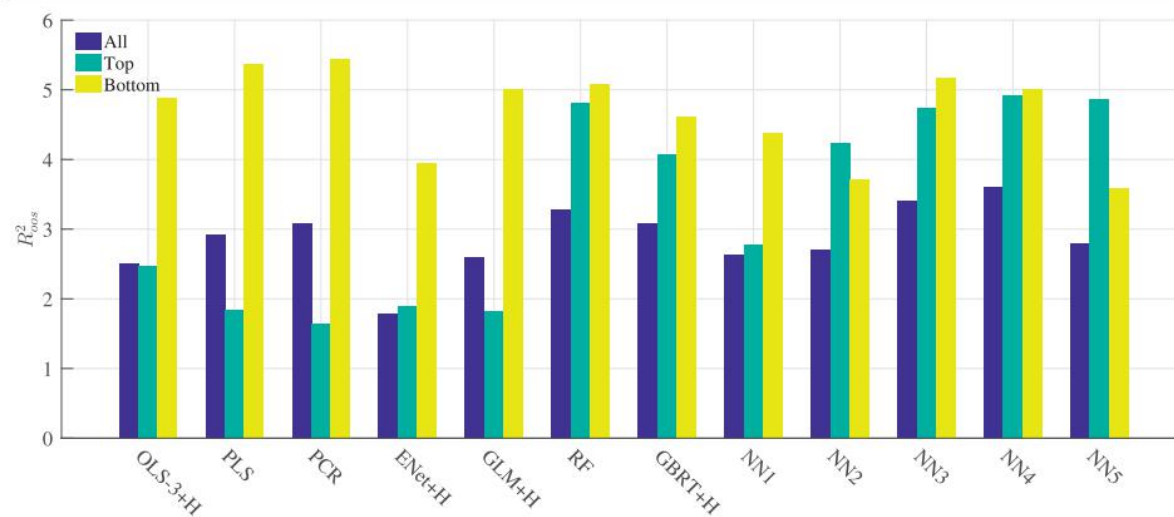
年度 $R^2$ 几乎要大一个数量级。

在预测年度回报方面的成功同样表明，机器学习能够区分在商业周期频率内持续存在的风险溢价。

**Table 2**

Annual out-of-sample stock-level prediction performance (percentage  $R_{\text{oos}}^2$ )

	OLS +H	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
All	-34.86	2.50	2.93	3.08	1.78	2.60	3.28	3.09	2.64	2.70	3.40	3.60	2.79
Top	-54.86	2.48	1.84	1.64	1.90	1.82	4.80	4.07	2.77	4.24	4.73	4.91	4.86
Bottom	-19.22	4.88	5.36	5.44	3.94	5.00	5.08	4.61	4.37	3.72	5.17	5.01	3.58



Annual return forecasting  $R_{\text{oos}}^2$  (see the legend to Table 1).

模型比较

在测试样本的前10年中，模型通常选择少于5个特征。2000年后，选定特征的数量增加，并在20到40之间徘徊。

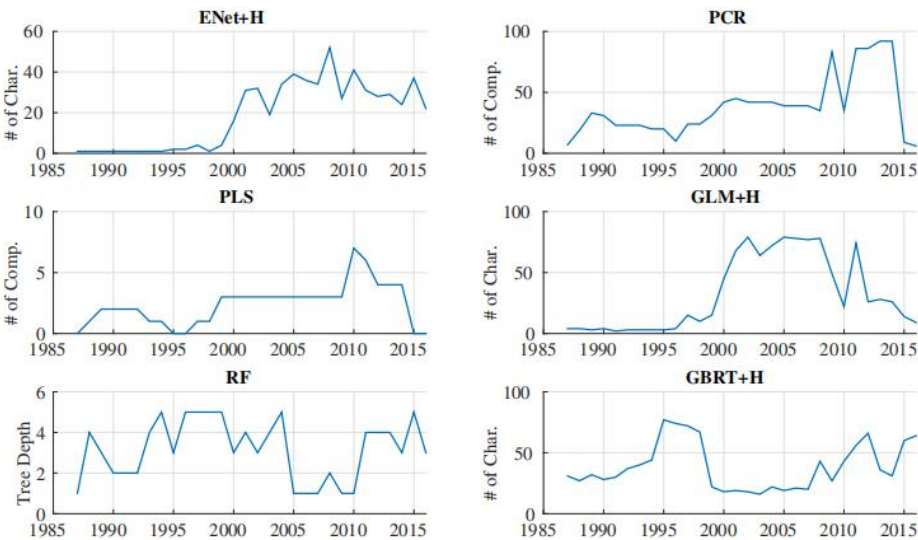


Table 3  
Comparison of monthly out-of-sample prediction using Diebold-Mariano tests

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
OLS+H	<b>3.26*</b>	<b>3.29*</b>	<b>3.35*</b>	<b>3.29*</b>	<b>3.28*</b>	<b>3.29*</b>	<b>3.26*</b>	<b>3.34*</b>	<b>3.40*</b>	<b>3.38*</b>	<b>3.37*</b>	<b>3.38*</b>
OLS-3+H		1.42	<b>1.87</b>	-0.27	0.62	<b>1.64</b>	1.28	1.25	<b>2.13</b>	<b>2.13</b>	<b>2.36</b>	<b>2.11</b>
PLS			-0.19	-1.18	-1.47	0.87	0.67	0.63	1.32	1.37	1.66	1.08
PCR				-1.10	-1.37	0.85	0.75	0.58	1.17	1.19	1.34	1.00
ENet+H					0.64	<b>1.90</b>	1.40	<b>1.73</b>	<b>1.97</b>	<b>2.07</b>	<b>1.98</b>	<b>1.85</b>
GLM+H						<b>1.76</b>	1.22	1.29	<b>2.28</b>	<b>2.17</b>	<b>2.68*</b>	<b>2.37</b>
RF							0.07	-0.03	0.31	0.37	0.34	0.00
GBRT+H								-0.06	0.16	0.21	0.17	-0.04
NN1									0.56	0.59	0.45	0.04
NN2										0.32	-0.03	-0.88
NN3											-0.32	-0.92
NN4												-1.04

This table reports pairwise Diebold-Mariano test statistics comparing the out-of-sample stock-level prediction performance among thirteen models. Positive numbers indicate the column model outperforms the row model. Bold font indicates the difference is significant at 5% level or better for individual tests, and an asterisk indicates significance at the 5% level for 12-way comparisons via our conservative Bonferroni adjustment.

Diebold-Mariano 检验统计量：其中正的数值表示列的模型优于行的模型，黑体表示在 5%的显著性水平上显著。同时，该表还展示了 Bonferroni multiple comparison的结果，星号表示在 5%显著性水平上显著。

## 2. 实证研究



### 哪些变量重要

所有模型的总体特征排名。每列中的颜色渐变显示从最小到最重要（从最亮到最暗）的特定于模型的特征排序。  
(原文图5)

测度：当保持其余所有变量不变，将某个变量的所有的值都设为 0 时，看整个预测 $R^2$ 的下降程度；



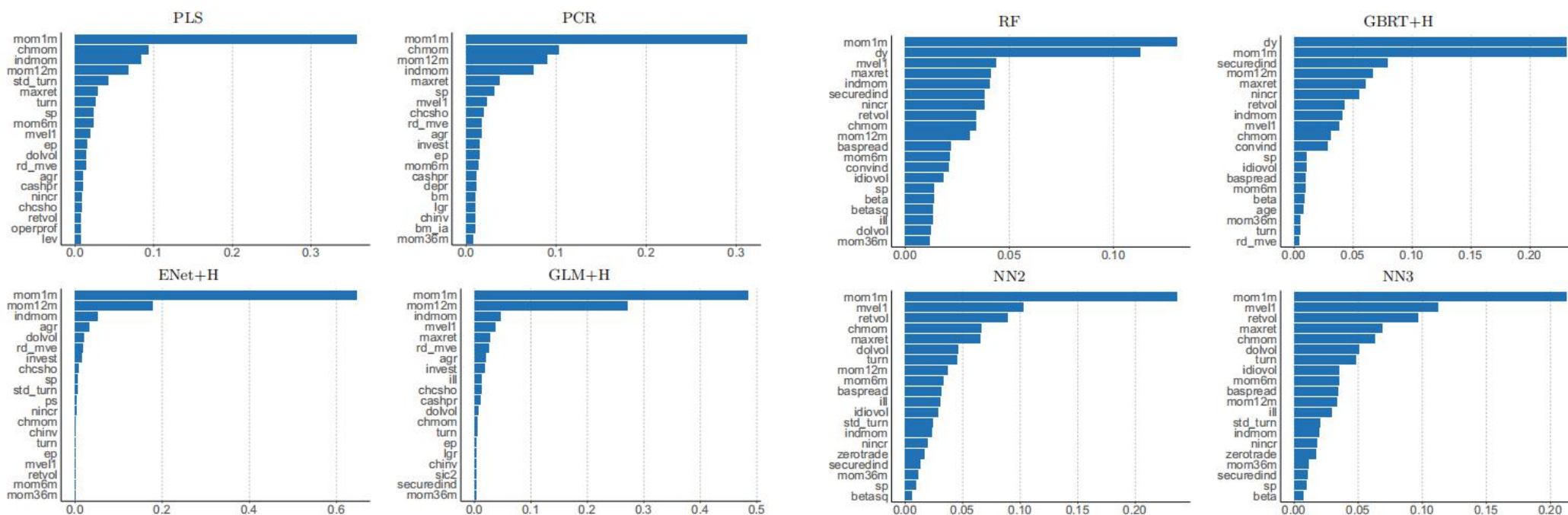


## 2. 实证研究



### 哪些变量重要

分别将每个变量在所有出现的地方都设为0，计算 $R^2$ 的减少。变量的重要性情况如图：（原文图4）





## 2. 实证研究

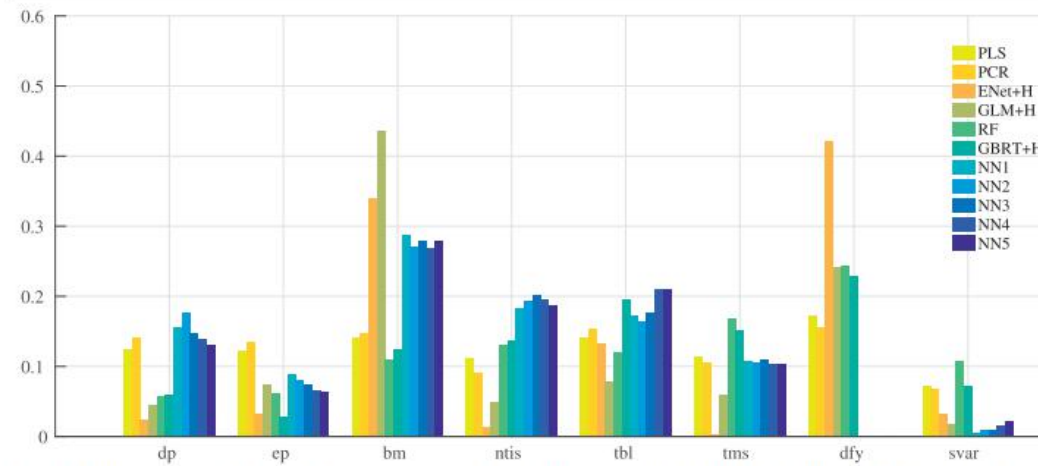
### 哪些变量重要

另一种测度是SSD——偏导数平方和  
(sum of squared partial derivatives, SSD)

该文发现结果SSD与 $R^2$ 减少测度相关性极高，  
在不同模型中，两种测度的Pearson相关系数  
在84.0%到97.7%之间。

**Table 4**  
Variable importance for macroeconomic predictors

	PLS	PCR	ENet+H	GLM+H	RF	GBRT+H	NN1	NN2	NN3	NN4	NN5
dp	12.52	14.12	2.49	4.54	5.80	6.05	15.57	17.58	14.84	13.95	13.15
ep	12.25	13.52	3.27	7.37	6.27	2.85	8.86	8.09	7.34	6.54	6.47
bm	14.21	14.83	33.95	43.46	10.94	12.49	28.57	27.18	27.92	26.95	27.90
ntis	11.25	9.10	1.30	4.89	13.02	13.79	18.37	19.26	20.15	19.59	18.68
tbl	14.02	15.29	13.29	7.90	11.98	19.49	17.18	16.40	17.76	20.99	21.06
tms	11.35	10.66	0.31	5.87	16.81	15.27	10.79	10.59	10.91	10.38	10.33
dfy	17.17	15.68	42.13	24.10	24.37	22.93	0.09	0.06	0.06	0.04	0.12
svar	7.22	6.80	3.26	1.87	10.82	7.13	0.57	0.85	1.02	1.57	2.29



Variable importance for eight macroeconomic variables in each model. Variable importance is an average over all training samples. Variable importance within each model is normalized to sum to one. The lower panel provides a complementary visual comparison of macroeconomic variable importance.

## 2. 实证研究



### 组合预测

**Table 5**  
**Monthly portfolio-level out-of-sample predictive  $R^2$**

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
<i>A. Common factor portfolios</i>												
S&P 500	-0.22	-0.86	-1.55	0.75	0.71	1.37	1.40	1.08	1.13	1.80	1.63	1.17
SMB	0.81	2.09	0.39	1.72	2.36	0.57	0.35	1.40	1.16	1.31	1.20	1.27
HML	0.66	0.50	1.21	0.46	0.84	0.98	0.21	1.22	1.31	1.06	1.25	1.24
RMW	-2.35	1.19	0.41	-1.07	-0.06	-0.54	-0.92	0.68	0.47	0.84	0.53	0.54
CMA	0.80	-0.44	0.03	-1.07	1.24	-0.11	-1.04	1.88	1.60	1.06	1.84	1.31
UMD	-0.90	-1.09	-0.47	0.47	-0.37	1.37	-0.25	-0.56	-0.26	0.19	0.27	0.35
<i>B. Subcomponents of factor portfolios</i>												
Big value	0.10	0.00	-0.33	0.25	0.59	1.31	1.06	0.85	0.87	1.46	1.21	0.99
Big growth	-0.33	-1.26	-1.62	0.70	0.51	1.32	1.19	1.00	1.10	1.50	1.24	1.11
Big neutral	-0.17	-1.09	-1.51	0.80	0.36	1.31	1.28	1.43	1.24	1.70	1.81	1.40
Small value	0.30	1.66	1.05	0.64	0.85	1.24	0.52	1.59	1.37	1.54	1.40	1.30
Small growth	-0.16	0.14	-0.18	-0.33	-0.12	0.71	1.24	0.05	0.42	0.48	0.41	0.50
Small neutral	-0.27	0.60	0.19	0.21	0.28	0.88	0.36	0.58	0.62	0.70	0.58	0.68
Big conservative	-0.57	-0.10	-1.06	1.02	0.46	1.11	0.55	1.15	1.13	1.59	1.37	1.07
Big aggressive	0.20	-0.80	-1.15	0.30	0.67	1.75	2.00	1.33	1.51	1.78	1.55	1.42
Big neutral	-0.29	-1.75	-1.96	0.83	0.48	1.13	0.77	0.85	0.85	1.51	1.45	1.16
Small conservative	-0.05	1.17	0.71	-0.02	0.34	0.96	0.56	0.82	0.87	0.96	0.90	0.83
Small aggressive	-0.10	0.51	0.01	-0.09	0.14	1.00	1.46	0.34	0.64	0.75	0.62	0.71
Small neutral	-0.30	0.45	0.12	0.42	0.35	0.76	-0.01	0.70	0.69	0.83	0.66	0.72

## 组合预测

**Table 5**  
Monthly portfolio-level out-of-sample predictive  $R^2$

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
Big robust	-1.02	-1.08	-2.06	0.55	0.35	1.10	0.33	0.74	0.79	1.28	1.03	0.74
Big weak	-0.12	1.42	1.07	0.89	1.10	1.33	1.77	1.79	1.79	2.05	1.66	1.60
Big neutral	0.86	-1.22	-1.26	0.41	0.13	1.10	0.91	0.84	0.94	1.19	1.15	0.99
Small robust	-0.71	0.35	-0.38	-0.04	-0.42	0.70	0.19	0.24	0.50	0.63	0.53	0.55
Small weak	0.05	1.06	0.59	-0.13	0.44	1.05	1.42	0.71	0.92	0.99	0.90	0.89
Small neutral	-0.51	0.07	-0.47	-0.33	-0.32	0.60	-0.08	0.10	0.25	0.38	0.32	0.41
Big up	0.20	-0.25	-1.24	0.66	1.17	1.18	0.90	0.80	0.76	1.13	1.12	0.93
Big down	-1.54	-1.63	-1.55	0.44	-0.33	1.14	0.71	0.36	0.70	1.07	0.90	0.84
Big medium	-0.04	-1.51	-1.94	0.81	-0.08	1.57	1.80	1.29	1.32	1.71	1.55	1.23
Small up	0.07	0.78	0.56	-0.07	0.25	0.62	-0.03	0.06	0.07	0.21	0.19	0.25
Small down	-0.21	0.15	-0.20	0.15	-0.01	1.51	1.38	0.74	0.82	1.02	0.91	0.96
Small medium	0.07	0.82	0.20	0.59	0.37	1.22	1.06	1.09	1.09	1.18	1.00	1.03



## 2. 实证研究



### 组合预测

**Table 6**  
**Market timing Sharpe ratio gains**

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
<i>A. Common factor portfolios</i>												
S&P 500	0.07	0.05	-0.06	0.12	0.19	0.18	0.19	0.22	0.20	0.26	0.22	0.19
SMB	0.06	0.17	0.09	0.24	0.26	0.00	-0.07	0.21	0.18	0.15	0.09	0.11
HML	0.00	0.01	0.04	-0.03	-0.02	0.04	0.02	0.04	0.06	0.04	0.02	0.01
RMW	0.00	-0.01	-0.06	-0.19	-0.13	-0.11	-0.01	-0.03	-0.09	0.01	0.01	-0.07
CMA	0.02	0.02	0	-0.09	-0.05	0.08	-0.01	0.00	0.01	0.05	0.04	0.06
UMD	0.01	-0.06	-0.02	-0.02	-0.07	-0.04	-0.07	-0.04	-0.08	-0.04	-0.10	-0.01
<i>B. Subcomponents of factor portfolios</i>												
Big value	-0.01	0.06	-0.03	0.09	0.06	0.09	0.08	0.11	0.11	0.13	0.10	0.11
Big growth	0.08	-0.01	-0.08	0.10	0.17	0.20	0.21	0.22	0.20	0.26	0.22	0.21
Big neutral	0.06	0.03	-0.06	0.11	0.16	0.13	0.17	0.23	0.21	0.23	0.23	0.21
Small value	-0.04	0.15	0.09	0.01	0.08	0.07	0.08	0.11	0.11	0.10	0.11	0.13
Small growth	0.00	0.03	-0.06	-0.03	-0.05	0.04	0.05	0.02	0.03	0.03	0.02	0.02
Small neutral	0.02	0.09	0.05	0.03	0.04	0.11	0.11	0.09	0.08	0.10	0.09	0.11
Big conservative	0.08	0.02	-0.04	0.08	0.15	0.09	0.13	0.17	0.14	0.19	0.16	0.14
Big aggressive	0.08	-0.01	-0.11	0.01	0.13	0.22	0.18	0.21	0.19	0.23	0.20	0.20
Big neutral	0.04	-0.01	-0.08	0.09	0.11	0.09	0.11	0.13	0.12	0.18	0.18	0.16
Small conservative	0.04	0.17	0.12	0.02	0.05	0.17	0.15	0.11	0.11	0.14	0.13	0.15
Small aggressive	0.01	0.05	-0.06	-0.05	-0.03	0.08	0.06	0.02	0.05	0.06	0.04	0.05
Small neutral	0.01	0.06	0.03	0.01	0.04	0.08	0.09	0.07	0.06	0.08	0.07	0.09

## 2. 实证研究



### 组合预测

**Table 6**  
**Market timing Sharpe ratio gains**

	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
Big conservative	0.08	0.02	-0.04	0.08	0.15	0.09	0.13	0.17	0.14	0.19	0.16	0.14
Big aggressive	0.08	-0.01	-0.11	0.01	0.13	0.22	0.18	0.21	0.19	0.23	0.20	0.20
Big neutral	0.04	-0.01	-0.08	0.09	0.11	0.09	0.11	0.13	0.12	0.18	0.18	0.16
Small conservative	0.04	0.17	0.12	0.02	0.05	0.17	0.15	0.11	0.11	0.14	0.13	0.15
Small aggressive	0.01	0.05	-0.06	-0.05	-0.03	0.08	0.06	0.02	0.05	0.06	0.04	0.05
Small neutral	0.01	0.06	0.03	0.01	0.04	0.08	0.09	0.07	0.06	0.08	0.07	0.09
Big robust	0.10	0.07	-0.07	0.11	0.18	0.17	0.18	0.18	0.16	0.22	0.19	0.16
Big weak	0.05	0.12	0.05	0.09	0.12	0.21	0.17	0.22	0.20	0.21	0.18	0.19
Big neutral	0.09	0.00	-0.04	0.09	0.20	0.19	0.17	0.22	0.21	0.24	0.21	0.20
Small robust	0.09	0.04	-0.03	0.00	0.00	0.10	0.07	0.04	0.05	0.08	0.08	0.08
Small weak	-0.03	0.09	0.00	-0.03	-0.02	0.07	0.07	0.06	0.06	0.06	0.05	0.06
Small neutral	0.04	0.04	-0.03	0.00	0.01	0.11	0.09	0.04	0.04	0.07	0.07	0.08
Big up	0.10	0.05	-0.06	0.10	0.21	0.16	0.14	0.17	0.14	0.17	0.18	0.17
Big down	-0.02	0.09	-0.08	-0.02	0.02	0.08	0.10	0.10	0.07	0.12	0.11	0.09
Big medium	-0.01	0.04	-0.06	0.14	0.09	0.17	0.20	0.22	0.21	0.25	0.22	0.19
Small up	0.08	0.13	0.10	0.05	0.07	0.16	0.12	0.07	0.06	0.08	0.07	0.10
Small down	-0.14	0.04	-0.05	-0.09	-0.05	0.06	0.04	0.01	0.01	0.02	0.01	0.01
Small medium	0.05	0.11	0.07	0.08	0.09	0.13	0.15	0.13	0.12	0.14	0.13	0.15



## 2. 实证研究



### 组合预测——机器学习

Table 7  
Performance of the machine learning portfolios

	OLS-3+H				PLS				PCR			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	Pred	Avg	SD	SR
Low(L)	-0.17	0.40	5.90	0.24	-0.83	0.29	5.31	0.19	-0.68	0.03	5.98	0.02
2	0.17	0.58	4.65	0.43	-0.21	0.55	4.96	0.38	-0.11	0.42	5.25	0.28
3	0.35	0.60	4.43	0.47	0.12	0.64	4.63	0.48	0.19	0.53	4.94	0.37
4	0.49	0.71	4.32	0.57	0.38	0.78	4.30	0.63	0.42	0.68	4.64	0.51
5	0.62	0.79	4.57	0.60	0.61	0.77	4.53	0.59	0.62	0.81	4.66	0.60
6	0.75	0.92	5.03	0.63	0.84	0.88	4.78	0.64	0.81	0.81	4.58	0.61
7	0.88	0.85	5.18	0.57	1.06	0.92	4.89	0.65	1.01	0.87	4.72	0.64
8	1.02	0.86	5.29	0.56	1.32	0.92	5.14	0.62	1.23	1.01	4.77	0.73
9	1.21	1.18	5.47	0.75	1.66	1.15	5.24	0.76	1.52	1.20	4.88	0.86
High(H)	1.51	1.34	5.88	0.79	2.25	1.30	5.85	0.77	2.02	1.25	5.60	0.77
H-L	1.67	0.94	5.33	0.61	3.09	1.02	4.88	0.72	2.70	1.22	4.82	0.88

	ENet+H				GLM+H				RF			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	Pred	Avg	SD	SR
Low(L)	-0.04	0.24	5.44	0.15	-0.47	0.08	5.65	0.05	0.29	-0.09	6.00	-0.05
2	0.27	0.56	4.84	0.40	0.01	0.49	4.80	0.35	0.44	0.38	5.02	0.27
3	0.44	0.53	4.50	0.40	0.29	0.65	4.52	0.50	0.53	0.64	4.70	0.48
4	0.59	0.72	4.11	0.61	0.50	0.72	4.59	0.55	0.60	0.60	4.56	0.46
5	0.73	0.72	4.42	0.57	0.68	0.70	4.55	0.53	0.67	0.57	4.51	0.44
6	0.87	0.85	4.60	0.64	0.84	0.84	4.53	0.65	0.73	0.64	4.54	0.49
7	1.01	0.87	4.75	0.64	1.00	0.86	4.82	0.62	0.80	0.67	4.65	0.50
8	1.16	0.88	5.20	0.59	1.18	0.87	5.18	0.58	0.87	1.00	4.91	0.71
9	1.36	0.80	5.61	0.50	1.40	1.04	5.44	0.66	0.96	1.23	5.59	0.76
High(H)	1.66	0.84	6.76	0.43	1.81	1.14	6.33	0.62	1.12	1.53	7.27	0.73
H-L	1.70	0.60	5.37	0.39	2.27	1.06	4.79	0.76	0.83	1.62	5.75	0.98

	GBRT+H				NN1				NN2			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	Pred	Avg	SD	SR
Low(L)	-0.45	0.18	5.60	0.11	-0.38	-0.29	7.02	-0.14	-0.23	-0.54	7.83	-0.24
2	-0.16	0.49	4.93	0.35	0.16	0.41	5.89	0.24	0.21	0.36	6.08	0.20
3	0.02	0.59	4.75	0.43	0.44	0.51	5.07	0.35	0.44	0.65	5.07	0.44
4	0.17	0.63	4.68	0.46	0.64	0.70	4.56	0.53	0.59	0.73	4.53	0.56
5	0.34	0.57	4.70	0.42	0.80	0.77	4.37	0.61	0.72	0.81	4.38	0.64
6	0.46	0.77	4.48	0.59	0.95	0.78	4.39	0.62	0.84	0.84	4.51	0.65
7	0.59	0.52	4.73	0.38	1.11	0.81	4.40	0.64	0.97	0.95	4.61	0.71
8	0.72	0.72	4.92	0.51	1.31	0.75	4.86	0.54	1.13	0.93	5.09	0.63
9	0.88	0.99	5.19	0.66	1.58	0.96	5.22	0.64	1.37	1.04	5.69	0.63
High(H)	1.11	1.17	5.88	0.69	2.19	1.52	6.79	0.77	1.99	1.38	6.98	0.69
H-L	1.56	0.99	4.22	0.81	2.57	1.81	5.34	1.17	2.22	1.92	5.75	1.16

	NN3				NN4				NN5			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	Pred	Avg	SD	SR
Low(L)	-0.03	-0.43	7.73	-0.19	-0.12	-0.52	7.69	-0.23	-0.23	-0.51	7.69	-0.23
2	0.34	0.30	6.38	0.16	0.30	0.33	6.16	0.19	0.23	0.31	6.10	0.17
3	0.51	0.57	5.27	0.37	0.50	0.42	5.18	0.28	0.45	0.54	5.02	0.37
4	0.63	0.66	4.69	0.49	0.62	0.60	4.51	0.46	0.60	0.67	4.47	0.52
5	0.71	0.69	4.41	0.55	0.72	0.69	4.26	0.56	0.73	0.77	4.32	0.62
6	0.79	0.76	4.46	0.59	0.81	0.84	4.46	0.65	0.85	0.86	4.35	0.68
7	0.88	0.99	4.77	0.72	0.90	0.93	4.56	0.70	0.96	0.88	4.76	0.64
8	1.00	1.09	5.47	0.69	1.03	1.08	5.13	0.73	1.11	0.94	5.17	0.63
9	1.21	1.25	5.94	0.73	1.23	1.26	5.93	0.74	1.34	1.02	6.02	0.58
High(H)	1.83	1.69	7.29	0.80	1.89	1.75	7.51	0.81	1.99	1.46	7.40	0.68
H-L	1.86	2.12	6.13	1.20	2.01	2.26	5.80	1.35	2.22	1.97	5.93	1.15

Table A.9: Performance of Machine Learning Portfolios (Equally Weighted)

	OLS-3+H				PLS				PCR			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.14	0.11	7.99	0.05	-0.83	-0.26	6.41	-0.14	-0.71	-0.65	7.04	-0.32
2	0.17	0.35	6.81	0.18	-0.20	0.19	5.92	0.11	-0.11	0.16	6.23	0.09
3	0.35	0.44	6.09	0.25	0.12	0.40	5.49	0.25	0.19	0.40	5.67	0.25
4	0.49	0.63	5.61	0.39	0.39	0.67	5.06	0.46	0.42	0.58	5.45	0.37
5	0.63	0.73	5.24	0.49	0.62	0.69	5.14	0.47	0.63	0.72	5.11	0.49
6	0.75	0.83	4.88	0.59	0.84	0.77	5.14	0.52	0.81	0.80	4.98	0.55
7	0.88	0.75	4.73	0.55	1.06	0.88	5.12	0.60	1.01	0.98	5.02	0.68
8	1.03	0.80	4.72	0.59	1.32	1.01	5.29	0.66	1.23	1.08	5.02	0.75
9	1.22	1.14	4.73	0.83	1.67	1.28	5.60	0.79	1.52	1.33	5.28	0.88
High(H)	1.60	1.45	5.21	0.96	2.38	1.82	6.16	1.02	2.12	1.81	5.93	1.06
H-L	1.73	1.34	5.59	0.83	3.21	2.08	4.89	1.47	2.83	2.45	4.51	1.89
	ENet+H				GLM+H				RF			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.04	-0.24	6.43	-0.13	-0.49	-0.50	6.81	-0.25	0.26	-0.48	7.16	-0.23
2	0.27	0.44	5.90	0.26	0.01	0.32	5.80	0.19	0.44	0.24	5.67	0.15
3	0.44	0.52	5.27	0.34	0.29	0.56	5.46	0.36	0.53	0.55	5.36	0.36
4	0.59	0.70	4.73	0.51	0.50	0.61	5.22	0.41	0.60	0.62	5.15	0.42
5	0.73	0.71	4.94	0.49	0.68	0.72	5.11	0.49	0.67	0.66	5.11	0.44
6	0.87	0.79	5.00	0.55	0.84	0.78	5.12	0.53	0.73	0.77	5.13	0.52
7	1.01	0.85	5.21	0.56	1.00	0.78	5.06	0.54	0.80	0.74	5.10	0.50
8	1.17	0.88	5.47	0.56	1.18	0.89	5.14	0.60	0.87	0.99	5.29	0.65
9	1.36	0.85	5.90	0.50	1.41	1.25	5.80	0.75	0.97	1.22	5.67	0.74
High(H)	1.72	1.86	7.27	0.89	1.89	1.81	6.57	0.96	1.20	1.90	7.03	0.94
H-L	1.76	2.11	5.50	1.33	2.38	2.31	4.41	1.82	0.94	2.38	5.57	1.48
	GBRT+H				NN1				NN2			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.49	-0.37	6.46	-0.20	-0.45	-0.78	7.43	-0.36	-0.32	-1.01	7.79	-0.45
2	-0.16	0.42	5.80	0.25	0.15	0.22	6.24	0.12	0.20	0.17	6.34	0.09
3	0.02	0.56	5.31	0.36	0.43	0.47	5.55	0.29	0.43	0.52	5.49	0.33
4	0.17	0.74	5.43	0.47	0.64	0.64	5.00	0.45	0.59	0.71	5.02	0.49
5	0.33	0.63	5.31	0.41	0.80	0.80	4.76	0.58	0.72	0.76	4.60	0.57
6	0.46	0.83	5.23	0.55	0.95	0.85	4.63	0.63	0.84	0.81	4.52	0.62
7	0.59	0.67	5.13	0.45	1.12	0.84	4.66	0.62	0.97	0.94	4.61	0.70
8	0.72	0.82	5.08	0.56	1.32	0.88	4.95	0.62	1.14	0.92	4.86	0.66
9	0.88	1.12	5.41	0.72	1.63	1.17	5.62	0.72	1.41	1.10	5.55	0.69
High(H)	1.19	1.77	6.69	0.92	2.43	2.13	7.34	1.00	2.25	2.30	7.81	1.02
H-L	1.68	2.14	4.28	1.73	2.89	2.91	4.72	2.13	2.57	3.31	4.92	2.33
	NN3				NN4				NN5			
	Pred	Avg	Std	SR	Pred	Avg	Std	SR	Pred	Avg	Std	SR
Low(L)	-0.31	-0.92	7.94	-0.40	-0.19	-0.95	7.83	-0.42	-0.08	-0.83	7.92	-0.36
2	0.22	0.16	6.46	0.09	0.29	0.17	6.50	0.09	0.33	0.24	6.64	0.12
3	0.45	0.44	5.40	0.28	0.49	0.45	5.58	0.28	0.51	0.53	5.65	0.32
4	0.60	0.66	4.83	0.48	0.62	0.57	4.94	0.40	0.62	0.59	4.91	0.41
5	0.73	0.77	4.58	0.58	0.72	0.70	4.57	0.53	0.71	0.68	4.56	0.51
6	0.85	0.81	4.47	0.63	0.81	0.75	4.42	0.59	0.80	0.76	4.43	0.60
7	0.97	0.86	4.62	0.64	0.91	0.86	4.47	0.67	0.88	0.88	4.60	0.66
8	1.12	0.93	4.82	0.67	1.04	1.06	4.82	0.76	1.01	0.95	4.90	0.67
9	1.38	1.18	5.51	0.74	1.28	1.24	5.57	0.77	1.25	1.17	5.60	0.73
High(H)	2.28	2.35	8.11	1.00	2.16	2.37	8.03	1.02	2.08	2.27	7.95	0.99
H-L	2.58	3.27	4.80	2.36	2.35	3.33	4.71	2.45	2.16	3.09	4.98	2.15

## 2. 实证研究



### 组合预测——机器学习

**Table 8**  
**Drawdowns, turnover, and risk-adjusted performance of machine learning portfolios**

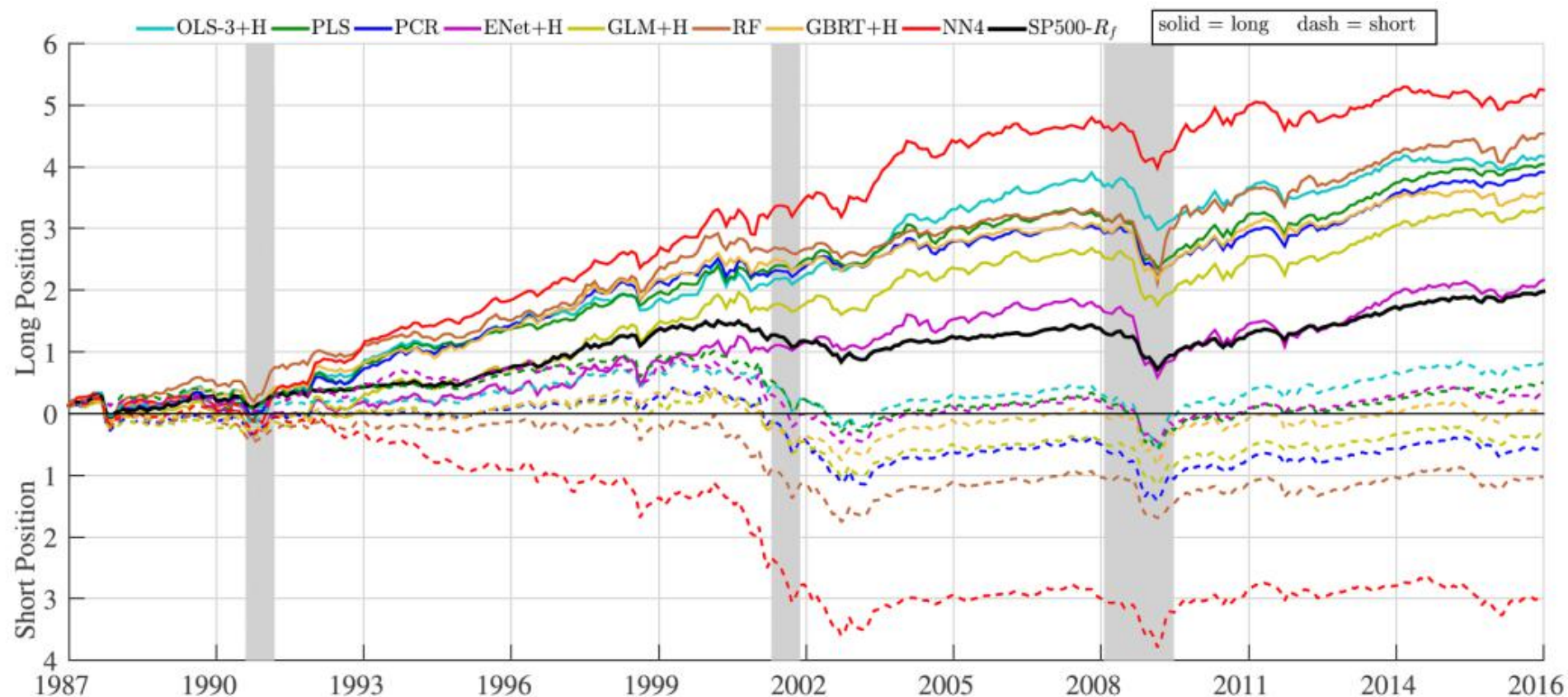
	OLS-3 +H	PLS	PCR	ENet +H	GLM +H	RF	GBRT +H	NN1	NN2	NN3	NN4	NN5
Drawdowns and turnover (value weighted)												
Max DD(%)	69.60	41.13	42.17	60.71	37.09	52.27	48.75	61.60	55.29	30.84	51.78	57.52
Max 1M loss(%)	24.72	27.40	18.38	27.40	15.61	26.21	21.83	18.59	37.02	30.84	33.03	38.95
Turnover(%)	58.20	110.87	125.86	151.59	145.26	133.87	143.53	121.02	122.46	123.50	126.81	125.37
Drawdowns and turnover (equally weighted)												
Max DD(%)	84.74	32.35	31.39	33.70	21.01	46.42	37.19	18.25	25.81	17.34	14.72	21.78
Max 1M loss(%)	37.94	32.35	22.33	32.35	15.74	34.63	22.34	12.79	25.81	12.50	9.01	21.78
Turnover(%)	57.24	104.47	118.07	142.78	137.97	120.29	134.24	112.35	112.43	113.76	114.17	114.34
Risk-adjusted performance (value weighted)												
Mean ret.	0.94	1.02	1.22	0.60	1.06	1.62	0.99	1.81	1.92	2.12	2.26	1.97
FF5+Mom $\alpha$	0.39	0.24	0.62	-0.23	0.38	1.20	0.66	1.20	1.33	1.52	1.76	1.43
$t(\alpha)$	2.76	1.09	2.89	-0.89	1.68	3.95	3.11	4.68	4.74	4.92	6.00	4.71
$R^2$	78.60	34.95	39.11	28.04	30.78	13.43	20.68	27.67	25.81	20.84	20.47	18.23
IR	0.54	0.21	0.57	-0.17	0.33	0.77	0.61	0.92	0.93	0.96	1.18	0.92
Risk-adjusted performance (equally weighted)												
Mean ret.	1.34	2.08	2.45	2.11	2.31	2.38	2.14	2.91	3.31	3.27	3.33	3.09
FF5+Mom $\alpha$	0.83	1.40	1.95	1.32	1.79	1.88	1.87	2.60	3.07	3.02	3.08	2.78
$t(\alpha)$	6.64	5.90	9.92	4.77	8.09	6.66	8.19	10.51	11.66	11.70	12.28	10.68
$R^2$	84.26	26.27	40.50	20.89	21.25	19.91	11.19	13.98	10.60	9.63	11.57	14.54
IR	1.30	1.15	1.94	0.93	1.58	1.30	1.60	2.06	2.28	2.29	2.40	2.09



## 2. 实证研究



### 组合预测——机器学习



**Figure 9**  
Cumulative return of machine learning portfolios

对机器学习中的方法进行了比较分析。

- 1、机器学习方法有助于提高我们对资产价格的实证理解。
- 2、神经网络，以及在较小程度上回归树，是表现最好的方法。
- 3、“浅”学习优于“深”学习，可能是由于资产定价问题中相对缺乏数据和低信噪比。
- 4、机器学习方法对于预测更大、更具流动性的股票回报和投资组合最有价值。
- 5、所有方法都表明一组相当小的主导预测信号，其中最强大的预测与价格趋势相关，包括收益逆转和动量。下一个最有力的预测指标是股票流动性、股票波动性和估值比率。





清华大学  
Tsinghua University

# 通过机器学习进行的实证资产定价

Empirical Asset Pricing via Machine Learning



汇报人：蒋子函、文家伟