

收益率可预测性的实证检验

蒋志强

zqjiang.ecust@qq.com



股票收益率的预测性



ARTICLE | [Full Access](#)

Does Academic Research Destroy Stock Return Predictability?

R. DAVID MCLEAN, JEFFREY PONTIFF

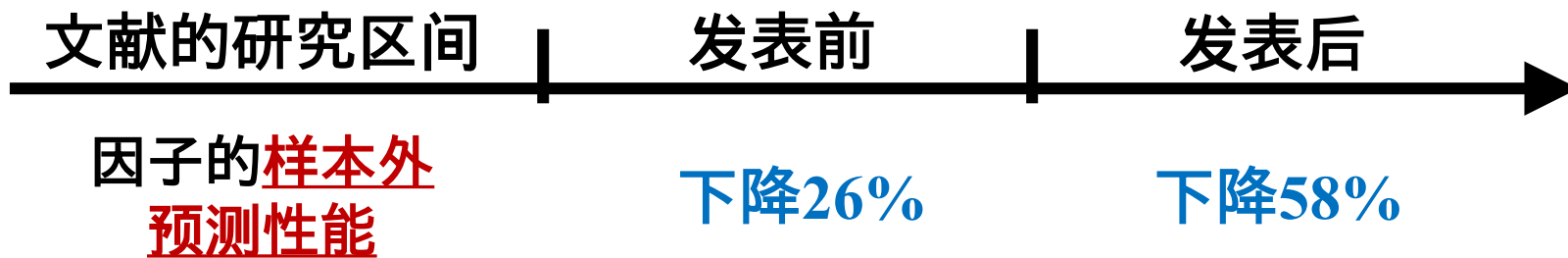


Volume 71, Issue 1

February 2016

Pages 5-32

收集整理了历史文献报道的97个预测因子



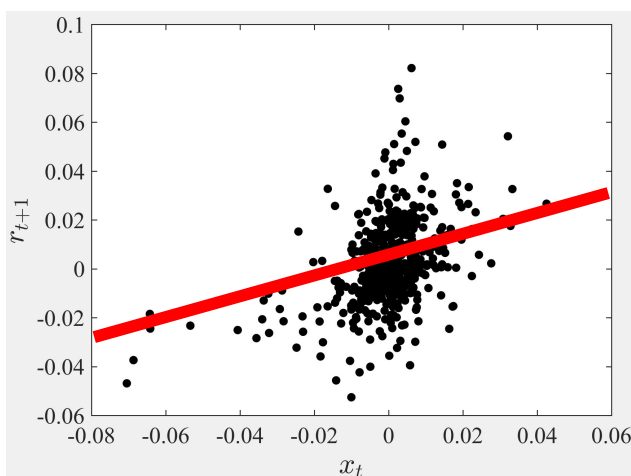
结论: 投资者会关注关于资产错误定价的学术研究, 并应用于投资实践, 纠正市场的错误定价。

收益率预测性的模型

单因子模型(双变量预测模型):

收益率 = 预测因子

$$r_{t+1} = \beta x_t + \alpha + \varepsilon_{t+1}$$



预测因子: 油价、技术分析指标、流动性、波动性、长短期利差、股息价格比

收益率预测性的思路和方法

单因子模型(双变量预测模型):

$$\text{收益率} \quad r_{t+1} = \text{预测因子} \quad \beta x_t + \alpha + \varepsilon_{t+1}$$

$$\begin{array}{l} r_2, r_3, r_4, \dots, r_{i-1}, r_i, r_{i+1}, r_{i+2}, \dots, r_n \\ x_1, x_2, x_3, \dots, x_{i-2}, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1} \end{array}$$

In-sample tests

- ✓ 最小二乘估计参数 α 和 β
- ✓ 参数显著性检验
- ✓ 计算 R^2 , 很小, 上限 **5%** (Huang and Zhou, 2017 JFQA)

收益率预测性的思路和方法

单因子模型(双变量预测模型):

收益率 = 预测因子

$$r_{t+1} = \beta x_t + \alpha + \varepsilon_{t+1}$$

$$\begin{array}{c} r_2, r_3, r_4, \dots, r_{i-1}, r_i, r_{i+1}, r_{i+2}, \dots, r_n \\ x_1, x_2, x_3, \dots, x_{i-2}, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1} \end{array}$$

Out-of-sample tests 统计显著性检验

- 计算 $R_{OS}^2 = 1 - \frac{\sum_{i=s}^n (r_i - \hat{r}_i)^2}{\sum_{i=s}^n (r_i - \bar{r}_i)^2}$, $R_{OS}^2 > 0$ 有样本外预测效果
- MSFE-adjusted 统计量: 回归 d_i 和 i , 常数项的 t 统计量

$$d_i = (r_i - \bar{r}_i)^2 - [(r_i - \hat{r}_i)^2 - (\bar{r}_i - \hat{r}_i)^2], \quad i = s, \dots, n$$

$$H_0 : R_{OS}^2 \leq 0, \quad H_1 : R_{OS}^2 > 0$$

收益率预测性的思路和方法

Out-of-sample tests 经济显著性检验

根据模型**预测结果**进行**投资决策**:

投资者风险偏好 γ , 投资风险资产和无风险资产

1. 在 t 时刻, 风险资产的配置比例:

$$\hat{\omega}_t = \frac{1}{\gamma} \frac{\hat{r}_{t+1}}{\hat{\sigma}_{t+1}^2}$$

2. 下一个时刻的组合收益率:

$$\hat{r}_{t+1}^p = r_{t+1}^f + \hat{\omega}_t r_{t+1}$$

3. 计算投资效用:

$$\hat{U} = \hat{\mu} - 0.5\gamma\hat{\sigma}^2$$

考虑历史均值基准模型:

1. 在 t 时刻, 风险资产的配置比例:

$$\bar{\omega}_t = \frac{1}{\gamma} \frac{\bar{r}_{t+1}}{\bar{\sigma}_{t+1}^2}$$

2. 下一个时刻的组合收益率:

$$\bar{r}_{t+1}^p = r_{t+1}^f + \bar{\omega}_t r_{t+1}$$

3. 计算投资效用:

$$\bar{U} = \bar{\mu} - 0.5\gamma\bar{\sigma}^2$$

*****检验 \hat{r} 的均值不为0, 计算utility gain:**

$$\Delta U = \hat{U} - \bar{U}$$

投资人为获得超过历史平均预测而愿意支付的年化管理费用

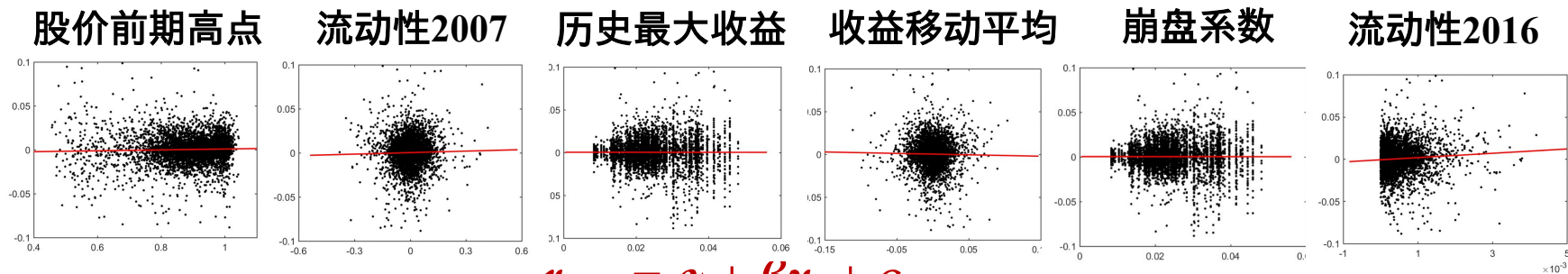
中国股市收益率的预测性研究

选取6个预测因子 管理科学学报、管理世界、经济研究、金融研究

1. 2005 股价前期高点 = 收盘价与前120天的最高收盘价的比值 《金融研究》
2. 2007 流动性 = 价格振幅 / $\lg(\text{成交额})$ 《管理世界》
3. 2013 历史最大收益 = 过去一个月3个最大日收益率的平均值 《经济研究》
4. 2014 收益移动平均 = 前10天的收益的移动平均与当天收益的差 《金融研究》
5. 2016 崩盘系数 = 200天的收益分布的5%分位数 《金融研究》
6. 2016 流动性 = $|\text{收益率} / \lg(\text{成交额})|$ 《管理科学学报》

数据:1997-2017年, 上证指数日度数据, 无风险利率

中国股市收益率的预测性研究



$$r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$$

	股价前期 高点	流动性 2007	历史最大 收益	收益移动 平均	崩盘 系数	流动性 2016
α	-0.0043	2.46×10^{-4}	-2.30×10^{-4}	2.53×10^{-4}	2.77×10^{-4}	-8.87×10^{-4}
β	0.0051**	0.0057*	0.0220	-0.0165	-9.84×10^{-4}	2.5663***
R^2	0.1169%	0.0532%	0.0252%	0.03%	0.0001%	0.5650%
R^2_{os}	0.0046%	-0.0337%	-0.1135%	-0.0793%	-0.1078%	0.3838%

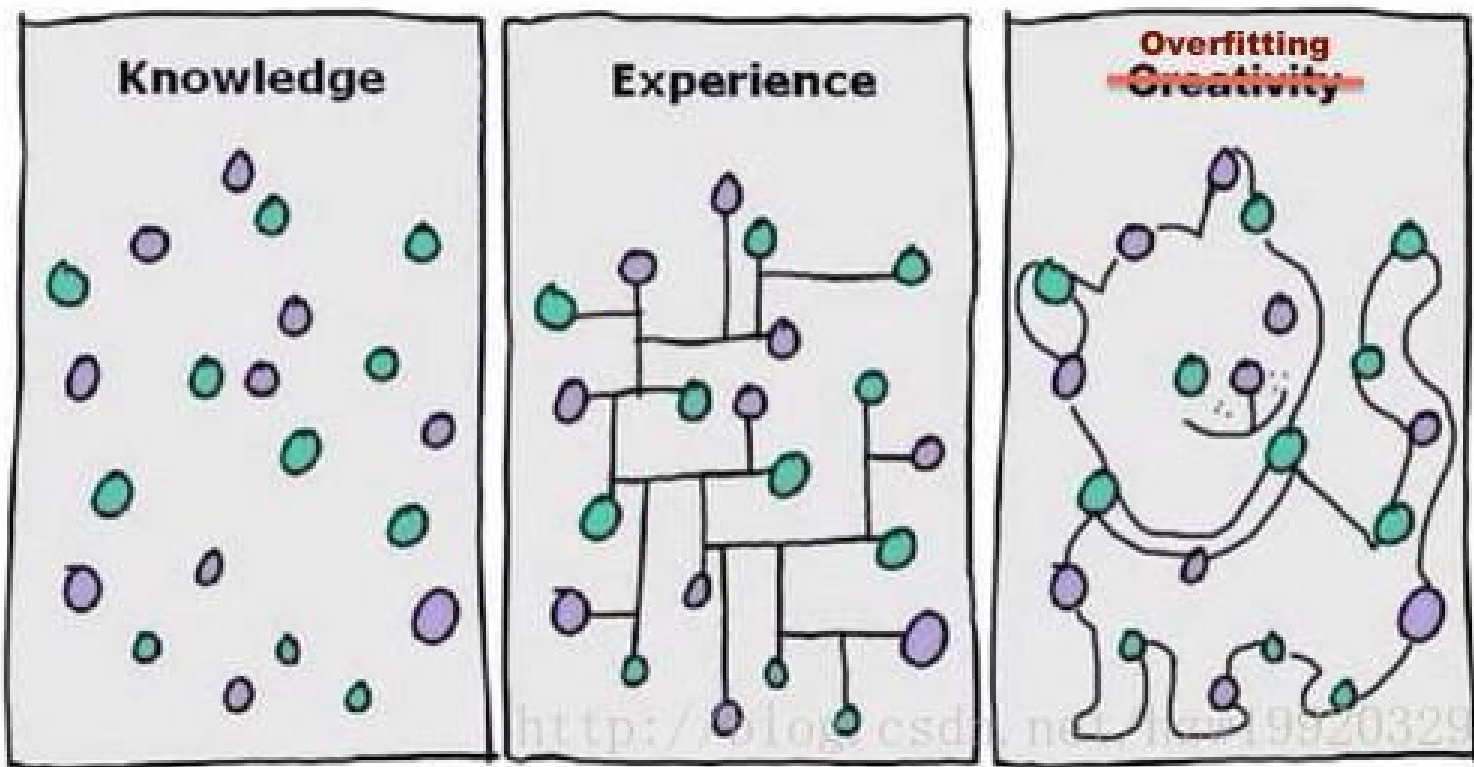
- β 显著的有股价前期高点和两个流动性指标, 有样本内预测能力。
- 样本内 R^2 很小, 所选因子对收益率的预测性能较差。
- 有样本内预测能力不保证有样本外预测能力。仅股价前期高点和流动性2016指标的 $R^2_{os} > 0$, 有样本外预测能力。

机器学习方法：带惩罚项的线性回归

➤ **OLS方法**: $\min L(\theta)$, 其中 $L(\theta) = \frac{1}{2(T-1)} \sum_{t=1}^{T-1} (R_{t+1} - f(Z_t; \theta))^2$

缺点: 预测因子数量接近观测值数量, 导致**过拟合**

因子之间存在**多重共线性**, 降低拟合效果



机器学习方法:LASSO

➤ 机器学习:正则化

在目标函数中增加一项**惩罚项(度量模型复杂度)**

降低模型样本内性能, 增加样本外稳定性

➤ **LASSO**: Least absolute shrinkage and selection operator

压缩估计, 构造一个惩罚函数, 压缩一些系数, 设定为零, 保留子集收缩

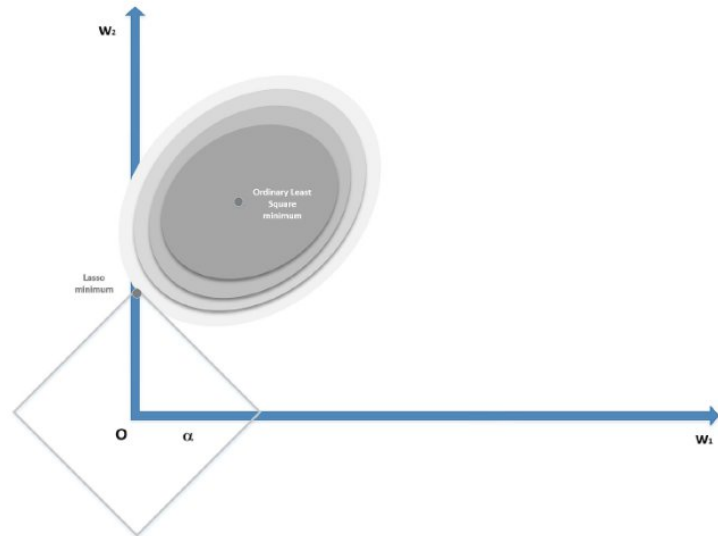
处理具有**多重共线性**数据的有偏估计

寻找最有影响力的自变量

引入**L1正则化** $\phi_1(\theta; \lambda) = \lambda \sum_{j=1}^J |\theta_j|$

构造**目标函数** $L(\theta; \lambda) = L(\theta) + \phi_1(\theta; \lambda)$

其中, θ 为系数, λ 为参数, 且为正值, 正则化强度。



机器学习方法：ElasticNet

➤ **L2正则化** $\phi_2(\theta; \lambda) = \lambda \sum_{j=1}^J \theta_j^2$

➤ **Ridge方法**引入L2正则化, 构造目标函数:

$$L(\theta; \lambda) = L(\theta) + \phi_2(\theta; \lambda)$$

➤ **ElasticNet方法**引入**L1和L2正则化**,

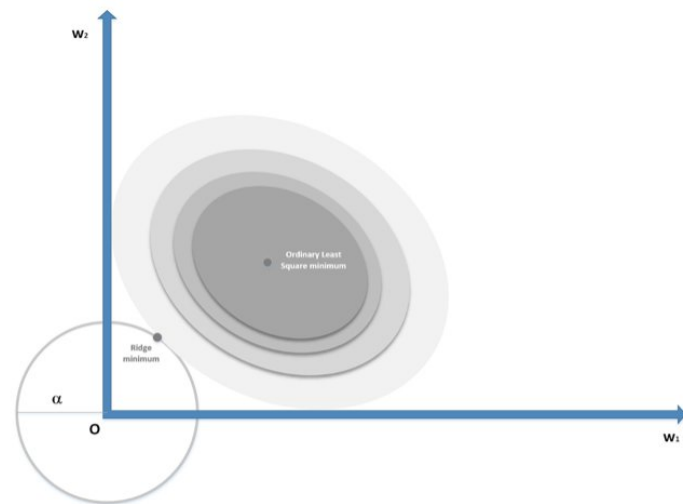
其惩罚项为:

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

目标函数为:

$$L(\theta; \cdot) = L(\theta) + \phi(\theta; \lambda, \delta)$$

其中, λ, δ 为参数, $\rho=1$ 时, 惩罚项变成L1正则化, 对应LASSO方法。



LASSO回归太过 (太多特征被稀疏为0)

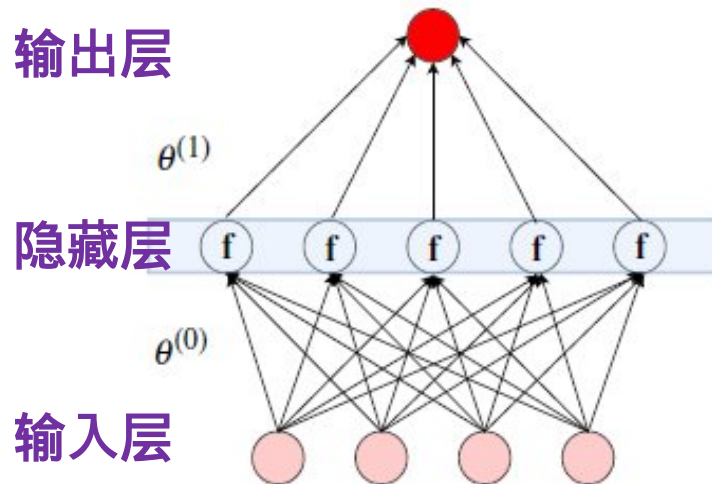
Ridge回归正则化不够 (回归系数衰减太慢)

使用ElasticNet方法回归

机器学习方法：神经网络

- **通用近似定理**：一个包含足够多隐藏层神经元的多层前馈网络，能以任意精度逼近任意预定的连续函数。
- 神经网络由**输入层**、一个或多个**隐藏层**和一个**输出层**组成。类似于生物大脑中的轴突，网络层代表一组神经元，神经元通过**激活函数**将输入信号转换成输出信号
- 神经网络存在多种**激活函数**，常用的有 **RELU**, **tanh(回归)**, **sigmoid**, **softmax(分类)**
- 目标函数用来衡量预测值与真实值间的差异，一般采用MSE。

包含五个神经元的隐藏层



- ①每个神经元从所有输入单元线性地提取信息
- ②每个神经元在将输出发送到下一层之前，将激活函数应用于其输出值。

隐藏层中的第二个神经元将输入转换为输出：

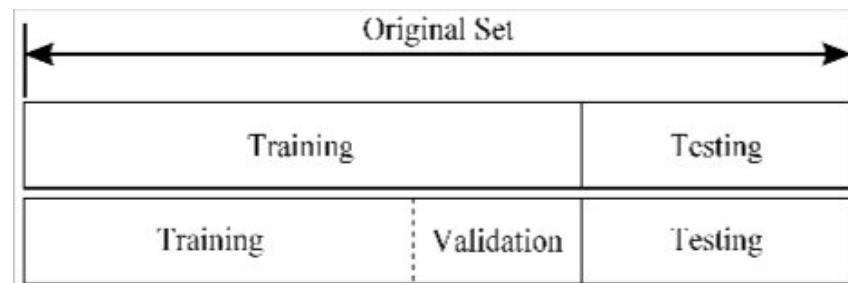
$$x_2^{(1)} = g(\theta_{2,0}^{(0)} + \sum_{j=1}^4 Z_j \theta_{2,j}^{(0)})。$$

输出预测值是每个神经元输出值的线性加总：

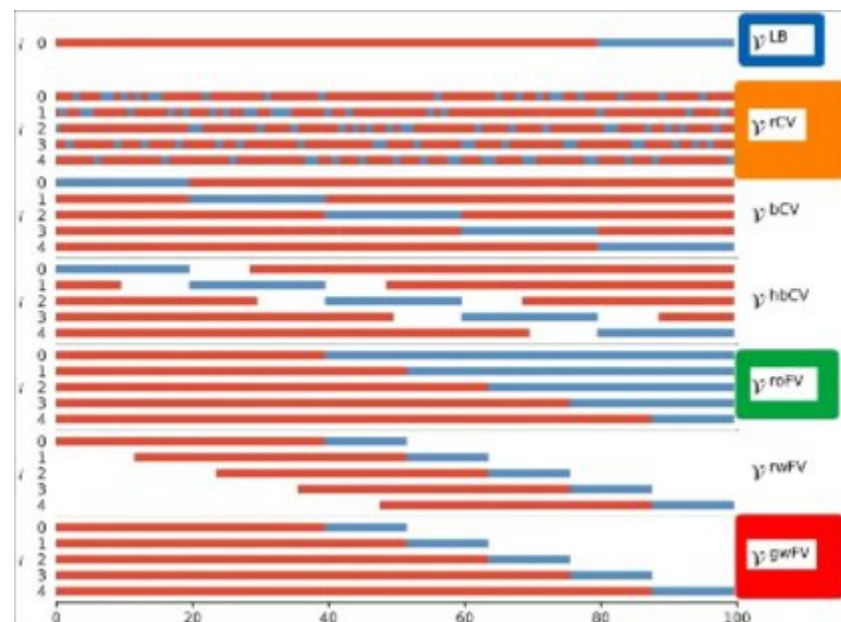
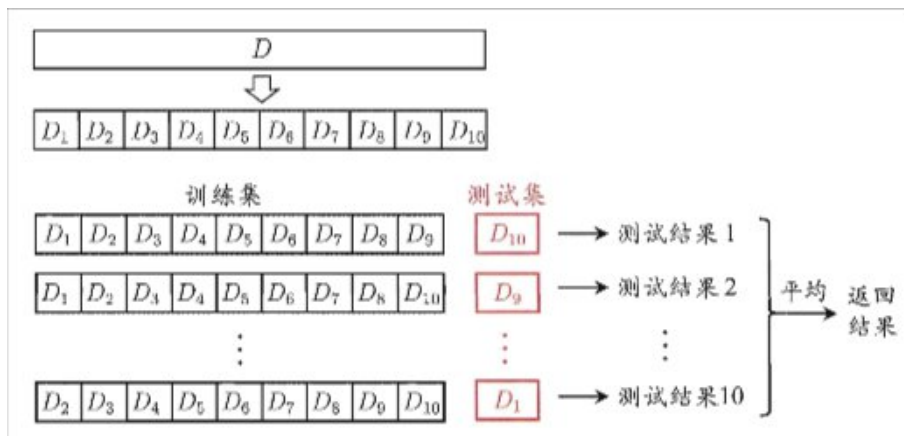
$$f(Z; \theta) = \theta_0^{(1)} + \sum_{j=1}^5 x_j^{(1)} \theta_j^{(1)}$$

预测收益率

- **训练集**: 训练机器学习模型
- **验证集**: 评估模型性能、调整超参数
- **测试集**: 最终评估模型性能
- **交叉验证**: 数据集 D 划分为 k 个大小相似的互斥子集, 共得 k 组训练/测试集, 返回 k 个测试结果的均值



$$D = D_1 \cup D_2 \cup \dots \cup D_k \quad D_i \cap D_j = \emptyset (i \neq j)$$



交叉验证样本集划分

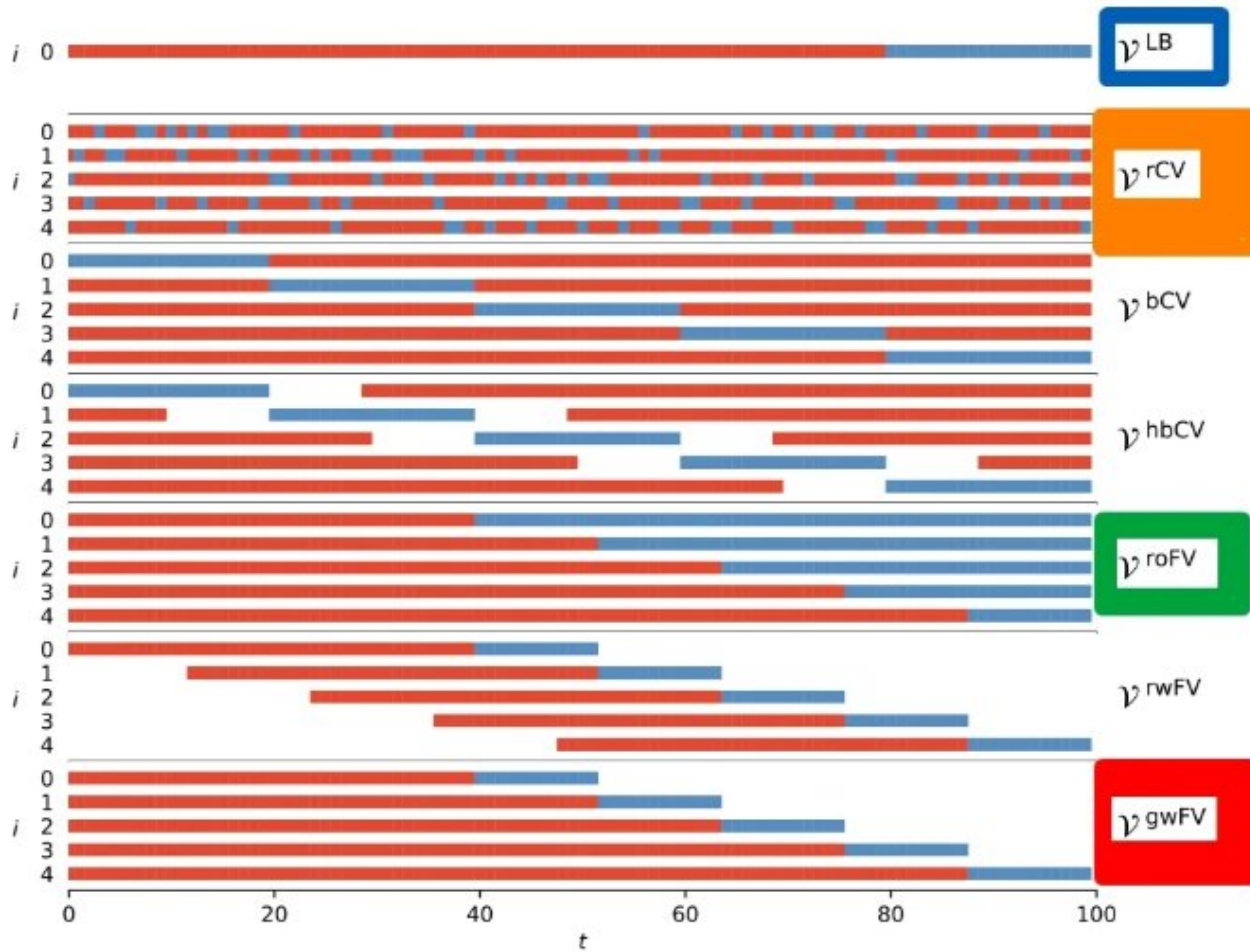


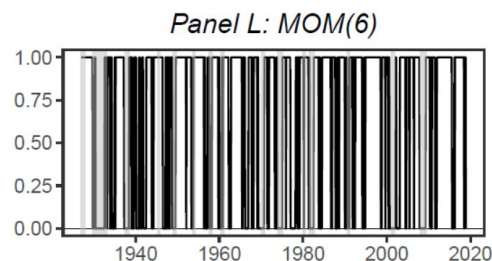
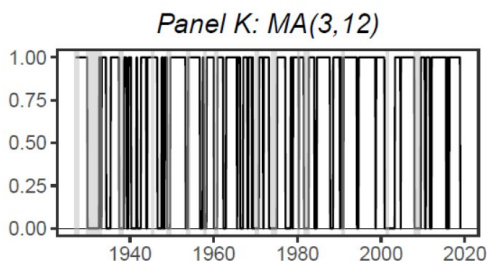
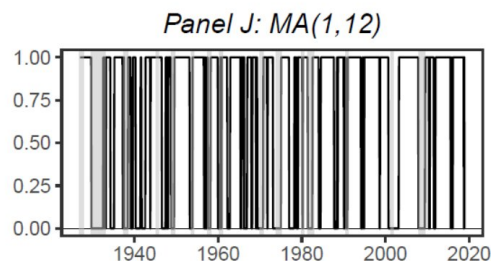
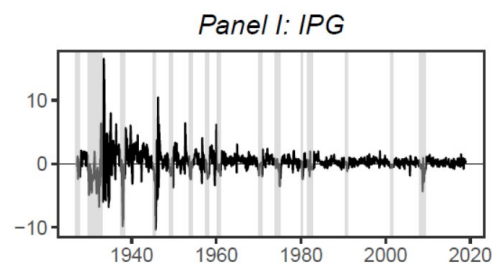
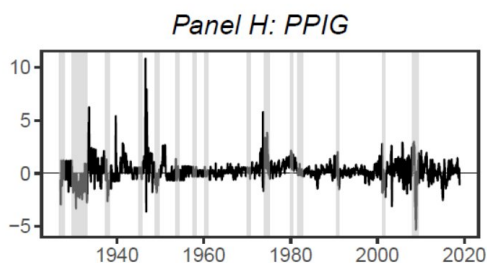
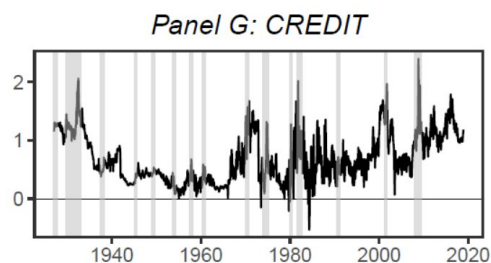
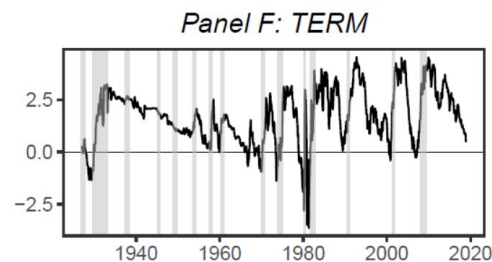
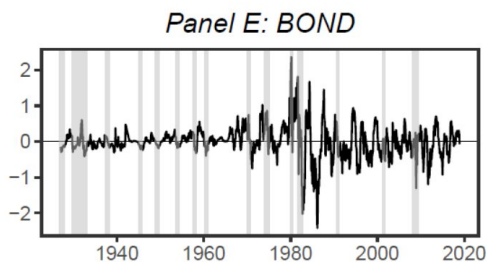
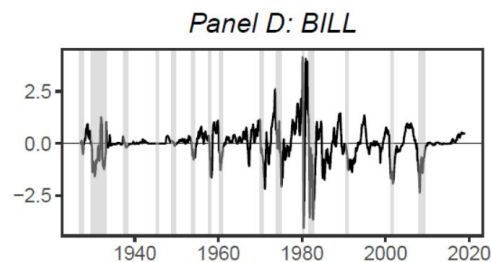
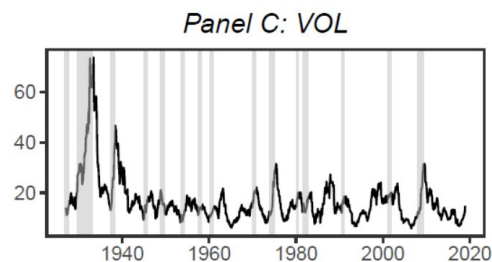
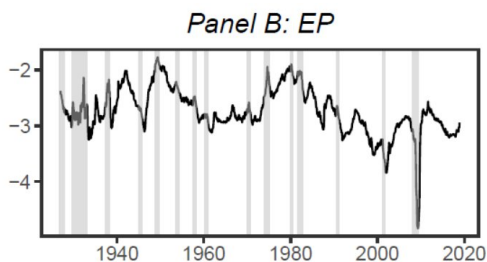
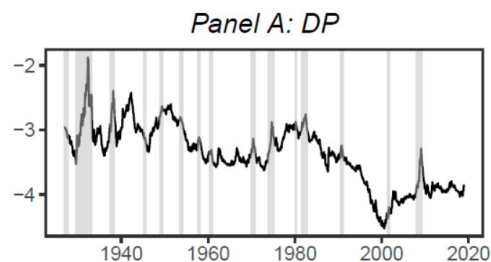
Figure 1: **Illustration of validation data splitting schemes.** We visualize different data splitting schemes \mathcal{V} by their respective validation index sets \mathcal{I}_i^v (blue) and training index sets \mathcal{I}_i^t (red), where i denotes the split index. We show the following schemes: \mathcal{V}^{LB} : last-block validation, \mathcal{V}^{rCV} : random cross-validation, \mathcal{V}^{bCV} : blocked cross-validation, $\mathcal{V}^{\text{hbCV}}$: h -blocked cross-validation, $\mathcal{V}^{\text{roFV}}$: rolling-origin forward-validation, $\mathcal{V}^{\text{rwFV}}$: rolling-window forward-validation, $\mathcal{V}^{\text{gwFV}}$: growing-window forward-validation. With the exception of \mathcal{V}^{LB} , sets are displayed for $k=5$ splits.

机器学习+股市收益率的可预测性

D. E. Rapach, G.-F. Zhou, **Time-series and cross-sectional stock return forecasting: New machine learning methods**, in E. Jurczenko, Ed. Machine Learning for Asset Management: New Developments and Financial Applications. ISTE * Wiley 2020

- **Monthly S&P 500 excess return**
- **Log dividend-price ratio (*DP*)**. Log of the twelve-month moving sum of S&P 500 dividends minus the log of the S&P 500 price index.
- **Log earnings-price ratio (*EP*)**. Log of the twelve-month moving sum of S&P 500 earnings minus the log of the S&P 500 price index.
- **Volatility (*VOL*)**. We measure the annualized volatility for month t as $\sqrt{\frac{\pi}{2}}\sqrt{12}\hat{\sigma}_t$, where $\hat{\sigma}_t = \frac{1}{12} \sum_{s=1}^{12} |r_{t-(s-1)}|$.
- **Treasury bill yield (*BILL*)**. Three-month Treasury bill yield minus the twelve-month moving average of the three-month Treasury bill yield.
- **Treasury bond yield (*BOND*)**. Ten-year Treasury bond yield minus the twelve-month moving average of the ten-year Treasury bond yield.
- **Term spread (*TERM*)**. Difference in yields on a ten-year Treasury bond and a three-month Treasury bill.
- **Credit spread (*CREDIT*)**. Difference in yields on a AAA-rated corporate bond and a ten-year Treasury bond.
- **Inflation (*PPIG*)**. Producer price index (PPI) inflation rate.
- **Industrial production growth (*IPG*)**. Growth rate of industrial production.
- **MA(1,12) technical signal [*MA*(1,12)]**. An indicator variable that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) the twelve-month moving average of the S&P 500 price index.
- **MA(3,12) technical signal [*MA*(3,12)]**. An indicator variable that takes a value of one (zero) if the three-month moving average of the S&P 500 price index is greater than or equal to (less than) the twelve-month moving average of the S&P 500 price index.
- **Momentum technical signal [*MOM*(6)]**. An indicator variables that takes a value of one (zero) if the S&P 500 price index is greater than or equal to (less than) its value six months ago.

机器学习+股市收益率的可预测性



机器学习+股市收益率的可预测性

1. 单因子模型

$$r_t = \alpha + \beta x_{j,t-1} + \varepsilon_t, \quad \hat{r}_{t+1|t}^{(j)} = \hat{\alpha}_{1:t}^{(j)} + \hat{\beta}_{1:t}^{(j)} x_{j,t}$$

2. 多因子模型

$$r_t = \alpha + \sum_{j=1}^J \beta_j x_{j,t-1} + \varepsilon_t \quad \hat{r}_{t+1|t}^{\text{OLS}} = \hat{\alpha}_{1:t}^{\text{OLS}} + \sum_{j=1}^J \hat{\beta}_{j,1:t}^{\text{OLS}} x_{j,t}$$

3. 复合模型

平均单变量预测结果

$$\hat{r}_{t+1|t}^{\text{C}} = \frac{1}{J} \sum_{j=1}^J \hat{r}_{t+1|t}^{(j)}$$

4. 机器学习模型

多变量OLS替换为ML

$$\hat{r}_{t+1|t}^{\text{ML}} = \hat{\alpha}_{1:t}^{\text{ML}} + \sum_{j=1}^J \hat{\beta}_{j,1:t}^{\text{ML}} x_{j,t}$$

5. 复合机器学习模型

多变量因子替换为
单变量模型预测值

$$r_t = \eta + \sum_{j=1}^J \theta_j \hat{r}_{t|t-1}^{(j)} + \varepsilon_t$$

利用ML选出有效果的单变量预测值求均值

$$\hat{r}_{t+1|t}^{\text{C-ML}} = \frac{1}{|\mathcal{J}_t|} \sum_{j \in \mathcal{J}_t} \hat{r}_{t+1|t}^{(j)}$$

模型评估: 计算 R^2_{os} , MSFE-adjusted 以及 ΔU

机器学习+股市收益率的可预测性

- The initial in-sample estimation period: **1927:01 to 1946:12**
- The initial holdout out-of-sample period: **1947:01 to 1956:12**
- The out-of-sample forecast evaluation period: **1957:01 to 2018:12**

(1)	(2)	(3)	(4)	(5)	(6)
Forecast	R_{OS}^2	MSFE-adj	Forecast	R_{OS}^2	MSFE-adj

Panel A: Individual predictor variables

<i>DP</i>	-0.40%	1.81**	<i>CREDIT</i>	-0.15%	-0.16
<i>EP</i>	-1.47%	0.79	<i>PPIG</i>	-0.50%	0.18
<i>VOL</i>	0.42%	2.58***	<i>IPG</i>	-0.02%	-0.04
<i>BILL</i>	0.15%	1.63*	<i>MA(1,12)</i>	0.28%	1.38*
<i>BOND</i>	1.04%	3.37***	<i>MA(3,12)</i>	-0.15%	0.32
<i>TERM</i>	0.26%	1.59*	<i>MOM(6)</i>	-0.04%	0.71

Panel B: Multiple predictor variables

OLS multiple predictive regression	-4.33%	2.73***	ENet multiple predictive regression	0.22%	3.11***
Simple combination	1.11%	3.70***	C-ENet	2.12%	4.05***

机器学习+中国股市收益率的可预测性

研究对象:沪深A股市场2100多只股票

数据:日度交易数据来自Wind数据库

季度财务数据, 无风险收益率和市场收益率来自CSMAR数据库

时间:2000-2018, 72个季度

因子:基本面 104, 技术因子 38

英文变量名	中文字段名	JLRXJHL	净利润现金含量
TRZB(SSZB)HBL	投入资本(实收资本)回报率	JYHDCSDXJLLJEBJYHDJ	经营活动产生的现金流量净额比经营活动净收益
TRZB(FZYSYZQYZJ)HBL	投入资本(负债与所有者权益总计)回报率	ZZCXJHSL	总资产现金回收率
ZBHBL	资本回报率	JYHDCSDXJLLJEBYYSR	经营活动产生的现金流量净额比营业收入
XSOLRBTRZB	息税前利润比投入资本	XSSPTGLWSDDXJBYYSR	销售商品提供劳务收到的现金比营业收入
JZCSYL	净资产收益率	XJBL	现金比率
ZCSYL	资产收益率	XJLZCBHZCHBLZC	现金流资产比和资产回报率之差
YYWLRBLRZE	营业外利润比利润总额	XSSJL	销售税金率
XSJLL	销售净利率	JLRL	净利润率
YYLRL	营业利润率	GLFYBYYZSR	管理费用比营业总收入
YYLRZCL	营业利润增长率	OYHBL	权益回报率
ZZCBCL	总资产报酬率	XSCBL	销售成本率
YYLRBYYZSR	营业利润比营业总收入	YYSRZS	营业收入指数
JLRBYYZSR	净利润比营业总收入	YYZCL	营业增长率
XSMML	销售毛利率	XSGLFYZS	销售管理费用指数
GDOYZZL	股东权益周转率	XSGLFYZCL	销售管理费用增长率
CBFYLRL	成本费用利润率	JYHDCSDXJLLJEZS	经营活动产生的现金流量净额指数
ZYXJLBYYSR	自由现金流比营业收入	JYHDCSDXJLLJEZCL	经营活动产生的现金流量净额增长率
TRZB(SSZB)XJHBL	投入资本(实收资本)现金回报率	LRZEZS	利润总额指数
TRZB(FZYSYZQYZJ)	投入资本(负债与所有者权益总计)现金回报率	LRZEZCL	利润总额增长率
XJHBL			

机器学习+中国股市收益率的可预测性

GSYMGSSYZDJLRZS	归属于母公司所有者的净利润指数	YSZKZS	应收账款指数
GSYMGSSYZDJLRZCL	归属于母公司所有者的净利润增长率	YXJZZWL	有形净值债务率
CZHDSCDXJLLJEZCL	筹资活动产生的现金流量净额增长率	LDZCZZL	流动资产周转率
JLRZCL	净利润增长率	LDBL	流动比率
YYFYBYYZSR	营业费用比营业总收入	CQFZBYYZJ	长期负债比营运资金
JZWJSBTRZB(SSZB)	净债务减少比投入资本(实收资本)	JYHDCSDXJLLJEBJFZ	经营活动产生的现金流量净额比净负债
JZWJSBTRZB(FZYSYZQYZJ)	净债务减少比投入资本(负债与所有者权益总计)	JYHDCSDXJLLJEBZFZ	经营活动产生的现金流量净额比总负债
FLDZCBL	非流动资产比率	XJLDFZB	现金流动负债比
SOPJZBZCBTRZB(SSZB)	三期平均资本支出比投入资本(实收资本)	CSDBL	超速动比率
SQPJZBZCBTRZB(FZYSYZQYZJ)	三期平均资本支出比投入资本(负债与所有者权益总计)	SDBL	速动比率
COBL	产权比率	5NPJSYSZB	5年平均收益市值比
COJKBZCZJ	长期借款比资产总计	JZBDJSYBLRZE	价值变动净收益比利润总额
CWFBYBYZSR	财务费用比营业总收入	LRZEBYYSR_TTM	利润总额比营业收入(TTM)
GDOYYGDZCBL	股东权益与固定资产比率	SYSZB_TTM	收益市值比(TTM)
ZWZZCB	债务总资产比	JYHDJSYBLRZE_TTM	经营活动净收益比利润总额(TTM)
GDOYBL	股东权益比率	JYHDJSYBLRZE	经营活动净收益比利润总额
GDZCBL	固定资产比率	YYLRBLRZE_TTM	营业利润比利润总额(TTM)
ZZWBTRZB(SSZB)	总债务比投入资本(实收资本)	YYWSZJEBLRZE_TTM	营业外收支净额比利润总额(TTM)
ZZWBTRZB(FZYSYZQYZJ)	总债务比投入资本(负债与所有者权益总计)	JLRZZL_TTM	净利润增长率(TTM)
ZYXJLBCOZW	自由现金流比长期债务	JXJLLZZL_TTM	净现金流量增长率(TTM)
ZBZCBGDZCJE	资本支出比固定资产净额	JZCZZL	净资产增长率
JYXXJLBZBZC	经营性现金流比资本支出	LRZEZZL_TTM	利润总额增长率(TTM)
JYXXJLBZBZCYLXZH	经营性现金流比资本支出与利息之和	GSMGSGDDJLRZZL_TTM	归属母公司股东的净利润增长率(TTM)
CHBH	存货变化	ZZCZZL	总资产增长率
YNCHBHJYSZKZZL	一年存货变化加应收账款周转率	TZHDCSDXJLLJEZZL_TTM	投资活动产生的现金流量净额增长率(TTM)
CHZZL	存货周转率	MLVZZL_TTM	毛利率增长率(TTM)
CHZZTS	存货周转天数	CZHDSCDXJLLJEZZL_TTM	筹资活动产生的现金流量净额增长率(TTM)
YSZKZZL	应收账款周转率	JYHDCSDXJLLJEZZL_TTM	经营活动产生的现金流量净额增长率(TTM)
YSZKZZTS	应收账款周转天数	YYLRZZL_TTM	营业利润增长率(TTM)
YYZO	营业周期	YYSRZZL_TTM	营业收入增长率(TTM)
ZZCZZL	总资产周转率		
COFZBZZC	长期负债比总资产		
GDZCZZL	固定资产周转率		
ZCFZL	资产负债率		
CWGGZS	财务杠杆指数		

机器学习+中国股市收益率的可预测性

数据库变量名	中文字段名		
		60rarbrqqyyzb	60日ARBR人气意愿指标
60rpjhs1	60日平均换手率	90rarbrqqyyzb	90日ARBR人气意愿指标
90rpjhs1	90日平均换手率	120rarbrqqyyzb	120日ARBR人气意愿指标
120rpjhs1	120日平均换手率	60rcrn1zb	60日CR能量指标
60rhslDSPj	60日换手率对数平均	90rcrn1zb	90日CR能量指标
90rhslDSPj	90日换手率对数平均	120rcrn1zb	120日CR能量指标
120rhslDSPj	120日换手率对数平均	dtld	多头力道
gqygyjylbgq3gydpjjyl	过去一个月交易量比过去3个月的平均交易量	ktld	空头力道
jylbdl	交易量波动率	scnlzb	市场能量指标
cjegx	成交额惯性	scqrzb	市场强弱指标
cjlzd	成交量震荡	msx	梅斯线
12vlisv	12月累计收益	adtszb	艾达透视指标
30r120rhb	30日120日回报方差比率	14rizb	14日均幅指标
60rarrqzb	60日AR人气指标	60rydix	60日移动均线
90rarrqzb	90日AR人气指标	90rydix	90日移动均线
120rarrqzb	120日AR人气指标	120rydix	120日移动均线
60rbryyzb	60日BR意愿指标	60rzsdydix	60日指数移动均线
90rbryyzb	90日BR意愿指标	90rzsdydix	90日指数移动均线
120rbryyzb	120日BR意愿指标	120rzsdydix	120日指数移动均线
		macd	平滑异同移动平均线

机器学习+中国股市收益率的可预测性

1. 验证因子数据完整性, 剔除数据严重缺失的因子, 最终保留了139个因子;
2. 调用因子数据, 划分训练集与测试集;
3. 设置机器学习方法的超参数(根据经验手动设置或者数据自适应);
4. 使用训练集训练模型参数;
5. 将模型用于测试集, 得出预测值;
6. 计算 R_{os}^2 , MSFE-adjusted 统计量及Utility gain。

机器学习+中国股市收益率的可预测性

LASSO示例结果:

股票代码	R^2_{os}	MSFE-adjusted	P值	Utility gain
000721	0.2338	2.3681	0.0328	0.1128

因子结果:

股票代码	投入资本回报率	总资产周转率	60日平均换手率	过去一个月交易量比过去3个月的平均交易量	成交量震荡	多头力道
	-0.0041	0.0675	0.0037	0.0011	0.0020	-0.0290
000721	市场能量指标	艾达透视指标	120日移动均线	60日指数移动均线	90日指数移动均线	120日指数移动均线
	0.0429	-0.0585	-0.0323	-0.0007	-0.0026	-0.0060

机器学习+中国股市收益率的可预测性

机器学习方法	$R_{os}^2 > 0$	R_{os}^{2***}	R_{os}^{2**}	R_{os}^{2*}	Utility gain > 0	Utility gain > 0 且 R_{os}^2 显著
OLS	7	0	0	0	427	0
LASSO	255	0	14	34	358	16
Ridge	12	0	1	2	324	1
ElasticNet	180	0	9	30	319	12
PCA	227	0	6	12	277	5
PLSR	46	0	2	5	302	1
SVR	69	0	3	11	270	2
AdaBoost	450	3	19	59	567	31
GBR	252	2	19	36	342	19
RF	334	1	10	33	295	15
NN-3	212	2	15	36	315	13

PCA: 主成分分析; PLSR: 偏最小二乘回归; SVR: 支持向量回归机; AdaBoost;

GBRT: 梯度提升回归树; RF: 随机森林; NN-3: 人工神经网络(3层隐藏层)