

# 《中国股票市场可预测性研究：基于机器学习的视角》 在线附录

李 斌，龙 真

（武汉大学经济与管理学院、武汉大学金融研究中心，武汉 430072）

作为正文内容的补充，本附录中展示了以下内容：

- 附录一展示了本文选取变量与评估指标的依据。
- 附录二展示了本文所用的 12 个预测变量的描述性统计与相关系数。
- 附录三展示了文中所采用的主要预测算法的简要说明，并比较了主要机器学习算法的优劣。
- 附录四展示了包括神经网络算法在内的完整股票市场样本外预测结果。
- 附录五展示了特征重要性及部分依赖图的计算方法，并展示了单变量与双变量部分依赖图。
- 附录六展示了投资组合的经济效益，包括存在交易成本的收益表现和不同风险厌恶系数下投资者实现的效用。
- 附录七展示了不同样本区间与超参数的稳健性检验。

# 目 录

附录一： 变量与指标的选择依据.....	1
表 A-1 经典文献和最新文献中所采用的变量以及评价指标 .....	1
附录二： 预测变量描述性统计及相关系数.....	2
表 A-2 预测变量描述性统计.....	2
表 A-3 预测变量相关系数.....	2
附录三： 主要预测模型算法简要说明.....	3
1. 历史均值预测 .....	3
2. 线性回归模型 .....	3
3. PLS 模型 .....	3
4. Ridge 回归 .....	3
5. Lasso 回归.....	3
6. Enet 回归.....	4
7. GARCH(1,1)模型 .....	4
8. MIDAS 模型 .....	4
9. BRT 模型 .....	4
10. RF 算法.....	5
11. Adaboost 算法.....	5
12. BRT 算法与其他机器学习模型优劣分析 .....	6
附录四： 股票市场样本外预测结果.....	7
表 A-4 股票市场样本外预测结果（%） .....	7
附录五： 变量重要性分析.....	8
1. 特征重要性计算方法 .....	8
2. 部分依赖图计算方法 .....	8
3. 预测变量部分依赖图 .....	9
图 A-1 提升回归树模型预测收益率的单变量部分依赖图 .....	9
图 A-2 提升回归树模型预测波动率的单变量部分依赖图 .....	10
图 A-3 提升回归树模型预测收益率和波动率的双变量部分依赖图 .....	11
附录六： 投资组合经济效益.....	12
1. 存在交易成本的经济效益 .....	12
表 A-5 存在交易成本的投资组合表现.....	12
2. 不同风险厌恶系数下的投资者效用 .....	12
表 A-6 不同风险厌恶水平下各模型平均实现效用 .....	12
附录七： 稳健性检验.....	13
1. 样本区间稳健性检验 .....	13

图 A-4 BRT、线性模型预测收益率与实际收益率折线图 .....	13
表 A-7 金融危机后(2009 年-2019 年)样本外预测结果 .....	13
<b>2. 超参数稳健性检验 .....</b>	<b>13</b>
表 A-8 不同超参数下 BRT 方法预测超额收益率结果 .....	14

## 附录一：变量与指标的选择依据

**表 A-1 经典文献和最新文献中所采用的变量以及评价指标**  
TableA-1 Variables and evaluation indicators used in classic and latest literature

名称	缩写	Welch et al. (2008)	Zhang et al. (2020)	Pettenuzzo et al. (2014)	姜富伟等 (2011)	蒋志强等 (2019)	本文
变 量							
股利支付率	D/E	√	√	√	√	√	√
股息价格比	D/P	√	√	√	√	√	√
股息率	D/Y	√	√	√	√	√	√
盈利价格比	E/P	√	√	√	√	√	√
账面市值比	B/M	√	√	√	√	√	√
股价方差	SVR	√	√	√	√	√	√
通货膨胀率	INFL	√	√	√	√	√	√
净权益增加值	NTIS	√	√	√	√		√
换手率	TO	√	√	√	√		√
M0 增长率	M0G	√	√	√	√		√
M1 增长率变动量	M1G	√	√	√	√		√
M2 增长率	M2G	√	√	√	√		√
现金收益价格比	CFP					√	
评价指标							
样本内 R <sup>2</sup>		√			√	√	√
样本外 R <sup>2</sup>		√	√	√	√	√	√
方向精度							√
经济效益			√	√	√	√	√

注：本表展示了国内外经典前沿文献和本文的变量选取与评价指标选取。

由于本文的研究重点在于从机器学习方法的视角重新审视股票市场风险溢价的预测，而非预测变量的选择，因此，本文采用的变量与评价指标与国内外相关研究的主要文献保持一致<sup>[1-5]</sup>。

## 附录二： 预测变量描述性统计

**表 A-2 预测变量描述性统计**

Table A-2 Descriptive statistics of predictive variables

	mean	std	min	25%	50%	75%	max	CV	skew	kurt
DE	-1.12	0.24	-1.79	-1.24	-1.17	-0.93	-0.51	-0.21	0.21	-0.12
DP	-4.31	0.64	-5.63	-4.87	-4.13	-3.78	-3.22	-0.15	-0.50	-1.08
DY	-4.29	0.63	-5.56	-4.84	-4.10	-3.77	-3.21	-0.15	-0.49	-1.07
EP	-3.19	0.65	-4.28	-3.85	-3.01	-2.62	-2.21	-0.21	-0.18	-1.45
BM	0.48	0.22	0.14	0.28	0.45	0.67	0.95	0.46	0.23	-1.20
SVR	0.61	0.71	0.03	0.20	0.35	0.74	4.90	1.16	2.74	9.63
INFL	0.18	0.75	-1.80	-0.21	0.10	0.68	2.60	4.19	0.11	0.41
NTIS	0.17	0.14	0.00	0.04	0.15	0.24	0.66	0.84	1.06	0.85
TO	15.31	9.03	1.72	9.14	13.19	19.34	48.60	0.59	1.21	1.50
M0G	0.01	0.07	-0.19	-0.02	0.01	0.03	0.31	6.99	1.03	4.49
M1G	0.00	0.03	-0.12	-0.02	0.00	0.02	0.10	-967.57	-0.29	0.62
M2G	0.01	0.01	-0.04	0.01	0.01	0.02	0.09	0.97	1.37	8.78

注：本表报告了 12 个预测变量的均值、标准差、最小值、最大值、25<sup>th</sup>、50<sup>th</sup>、75<sup>th</sup>分位数、变异系数、偏度和峰度。为了便于显示，表中 SVR、NTIS、M0G、M1G 和 M2G 数据为实际数据放大 1000 倍后的描述性统计。

表 A-2 展示了 12 个预测变量的描述性统计。从均值来看，股息支付率（DE）、股息价格比（DP）、股息率（DY）和盈利价格比（EP）的均值为负，其他变量均值为正，其中除了换手率（TO）的均值达到了 15.31，其他变量均值小于 1；从波动率来看，通过比较变量的变异系数，我们发现通货膨胀率（INFL）、M0 增长率（M0G）和 M1 增长率差值（M1G）波动率较大；从偏度看，收益波动率（SVR）、净权益增加值（NTIS）、换手率（TO）、M0 增长率（M0G）和 M2 增长率（M2G）偏度均大于 1，说明这些变量呈现明显正偏；从峰度看，收益波动率（SVR）、M0 增长率（M0G）和 M2 增长率（M2G）三个变量的峰度大于 3，呈现厚尾，其他变量则呈现瘦尾。

**表 A-3 预测变量相关系数**

Table A-3 Correlation of predictive variables

	DP	DY	EP	BM	SVR	INFL	NTIS	TO	M0G	M1G	M2G
DE	0.1290	0.1464	-0.2370	-0.0531	-0.0853	0.0250	0.0767	0.1060	0.0426	-0.0096	0.0729
DP		0.9911	0.9328	0.9187	-0.1253	0.0893	-0.7088	0.3314	0.0050	-0.0100	-0.0524
DY			0.9178	0.9025	-0.1495	0.0969	-0.7018	0.3803	-0.0024	-0.0010	-0.0453
EP				0.9194	-0.0918	0.0784	-0.7223	0.2862	-0.0106	-0.0063	-0.0779
BM					-0.2157	0.0440	-0.7433	0.2113	-0.0052	-0.0214	-0.1524
SVR						-0.0258	0.2045	0.3943	-0.0084	-0.0109	0.0418
INFL							-0.0144	-0.0158	-0.1869	0.0617	-0.0028
NTIS								-0.1945	0.0403	-0.0166	0.0917
TO									-0.0823	0.1240	0.0454
M0G										-0.2591	0.2631
M1G											0.2999

表 A-3 展示了 12 个预测变量之间的相关系数。可以看出股息价格比（DP）、股息率（DY）、盈利价格比（EP）和账面市值比（BM）之间的相关系数均大于 0.9，较高的相关性说明这些变量包含的信息有重复情况。此外，净权益增加值（NTIS）和股息盈利比（DE）、股息价格比（DP）、股息率（DY）、盈利价格比（EP）账面市值比（BM）的相关系数均小于-0.7，呈现较强的负相关，其他变量间的相关系数较小。

### 附录三： 主要预测模型算法简要说明

#### 1. 历史均值预测

参考 Welch and Goyal<sup>[1]</sup>，我们构建历史均值预测模型为基准模型之一，具体形式为：

$$\hat{r}_t = \frac{\sum_{i=1}^{t-1} r_i}{t-1}$$

即以前 $t-1$ 期的实际市场收益率的均值作为第 $t$ 期的预测值。

#### 2. 线性回归模型

参考 Welch and Goyal<sup>[1]</sup>，我们构建了线性模型作为基准模型之一，形式为：

$$r_t = \alpha + \beta X_{t-1} + \varepsilon_{r,t}$$

其中， $r_t$ 为 $t$ 期股市超额收益率， $X_{t-1}$ 为第 $t-1$ 期的预测因子，包含一个或多个变量， $\alpha$ 为截距项， $\beta$ 为预测因子对超额收益率的影响系数。用训练集样本对上式进行最小二乘回归，得到参数 $\hat{\alpha}$ 与 $\hat{\beta}$ ，然后代入 $t$ 期预测变量得到第 $t+1$ 期预测值 $\hat{r}_t$ ：

$$\hat{r}_t = \hat{\alpha} + \hat{\beta} X_t$$

将函数的损失函数定义为平方损失函数，

$$Loss(\beta) = \frac{1}{m} \sum_{i=1}^m (\hat{r}_t - r_t)^2$$

其中 $m$ 为有效样本数量，通过最小化目标损失函数，即可求解最优拟合系数 $\hat{\beta}$ 。

#### 3. PLS 模型

偏最小二乘回归（PLS）方法在多元回归的基础上结合了主成分分析和典型相关分析，用以解决回归分析中变量多重共线性问题。

考虑存在 $m$ 个自变量 $x_1, x_2, \dots, x_m$ 。偏最小二乘回归首先在自变量集合中提出第一主成分 $t_1$ （ $t_1$ 是 $x_1, x_2, \dots, x_m$ 的线性组合，且尽可能多地提取原自变量集中的变异信息），然后建立因变量与 $t_1$ 的回归方程，如果方程已达到满意的精度，则算法中止。否则继续第二主成分的提取，直到达到满意的精度。若最终提取了 $r$ 个成分 $t_1, t_2, \dots, t_r$ ，偏最小二乘回归将通过建立因变量与 $t_1, t_2, \dots, t_r$ 的回归式，并还原为因变量与原自变量的回归方程。

#### 4. Ridge 回归

岭回归（Ridge Regression）是一种线性方法，它在标准线性回归损失函数的基础上加入了 $L_2$ 范数正则化项，即

$$Loss(\beta) = \frac{1}{m} \sum_{i=1}^m (\hat{r}_t - r_t)^2 + \gamma \|\beta\|_2^2$$

其中 $\gamma > 0$ ，通过在损失函数中增加 $L_2$ 范数正则化项约束参数以降低模型复杂度，可以防止过拟合，并增强了模型的样本外预测能力。通过最小化损失函数即可求解方程最优拟合系数 $\hat{\beta}$ 。

#### 5. Lasso 回归

Lasso 回归（Lasso Regression）也是一种线性方法，它在标准线性回归损失函数的基础上加入了 $L_1$ 范数正则化项，即

$$Loss(\beta) = \frac{1}{m} \sum_{i=1}^m (\hat{r}_t - r_t)^2 + \gamma \|\beta\|_1$$

其中  $\gamma > 0$ , 相较于  $L_2$  范数而言,  $L_1$  范数更易获得稀疏解, 因此 Lasso 也常被用于高位数据的特征筛选, 使得与输出变量相关性较低的特征系数更接近 0, 可以有效防止过拟合, 并增强了模型的样本外预测能力. 通过最小化损失函数即可求解方程最优拟合系数  $\hat{\beta}$ .

## 6. Enet 回归

弹性网络回归 (ElasticNet Regression) 综合了 Lasso 和岭回归两种算法, 同时使用  $L_1$  和  $L_2$  正则化, 其损失函数可以表示为:

$$Loss(\beta) = \frac{1}{m} \sum_{i=1}^m (\hat{r}_t - r_t)^2 + \gamma_1 \|\beta\|_1 + \gamma_2 \|\beta\|_2^2$$

不同于 Lasso 将部分系数清零的做法, 弹性网络回归鼓励在高度相关变量时的群体效应. 当多个特征和另一个特征相关的情形下弹性网络往往能够取得较好的预测效果, Lasso 倾向于随机选择其中一个特征, 而弹性网络更倾向于选择两个特征. 此外, 上述回归正则化方法 (岭回归、Lasso 回归和 Elastic 回归) 往往在数据集中的变量具有高纬度以及变量间存在多重共线性时能够保持较好的预测效果.

## 7. GARCH(1,1)模型

作为波动率预测的基准模型<sup>[6]</sup>, GARCH(1, 1)在时间序列模型中引入自回归条件异方差, 定义为:

$$r_t = \mu_t + \sigma_t \epsilon_t$$

$$\hat{\sigma}_{t+1|t}^{GARCH} = \sqrt{\omega + \alpha(r_t - \mu_t)^2 + \beta\sigma_t^2} = \sqrt{\omega + \alpha\sigma_t^2\epsilon_t^2 + \beta\sigma_t^2}$$

其中  $var(\epsilon_t) = 1$ , 通过极大似然法对  $\omega$ 、 $\alpha$  和  $\beta$  进行估计, 然后代入  $t$  期数据预测  $t+1$  期的波动率.

## 8. MIDAS 模型

MIDAS 方法 (Mixed Data Sampling) 通过对过去  $D$  期日度波动率的加权加总对未来波动进行估计的线性方法<sup>[7]</sup>, 实现了用从高频数据中获取信息做出有效预测, 并且减少了参数的数量以降低过拟合风险, 并具体定义为:

$$\hat{\sigma}_{t+1|t}^{MIDAS} = 22 \sum_{d=0}^D \omega_d r_{t-d}^2$$

$$\omega_d(k_1, k_2) = \frac{\left(\frac{d}{D}\right)^{k_1-1} \left(1 - \frac{d}{D}\right)^{k_2-1}}{\sum_{i=0}^D \left(\frac{i}{D}\right)^{k_1-1} \left(1 - \frac{i}{D}\right)^{k_2-1}}$$

其中  $D$  为最大滞后期数设置为一年的近似交易天数(250 天), 参数  $k_1$  和  $k_2$  可以采用极大似然法确定. 得到  $\hat{k}_1$  和  $\hat{k}_2$  后将  $t$  期自变量代入, 计算出第  $t+1$  期波动率估计值.

## 9. BRT 模型

提升回归树是以回归树为基本模型加入提升算法, 通过迭代计算减少每次计算的残差<sup>[8]</sup>. 回归树是指选择一对参数  $(p, s)$ , 其中  $p$  为一个预测变量,  $s$  为以该变量划分子树的临界值, 将样本分成两个部分. 在每组子样本中重复这一步骤, 最终将全部样本分入  $J$  个叶节点生成一个树模型, 表示为:

$$T(x; \{S_j, c_j\}_{j=1}^J) = \sum_{j=1}^J c_j I\{x \in S_j\}$$

其中 $c_j$ 为叶节点 $S_j$ 中所有样本的因变量均值。

为了使模型得到更准确的估计，我们参考 Friedman<sup>[8]</sup>加入提升算法，即将一系列简单的回归树模型相结合，可以得到比单个模型更准确的估计，表示为：

$$f_B(x) = \sum_{b=1}^B T_b(x; \{S_{b,j}, c_{b,j}\}_{j=1}^J)$$

其中 $T_b(x; \{S_{b,j}, c_{b,j}\}_{j=1}^J)$ 是第  $b$  次迭代生成的回归树模型， $B$  表示提升迭代总次数。第  $b$  次迭代的模型因变量为 $f_{b-1}(x)$ 的残差，该模型的构建过程可以表示为：

$$\{\hat{S}_{b,j}, \hat{c}_{b,j}\}_{j=1}^J = \arg \min_{\{S_{b,j}, c_{b,j}\}_{j=1}^J} \sum_{t=0}^{T-1} [(y_{t+1} - f_{b-1}(x)) - T(x_t; \{S_{b,j}, c_{b,j}\}_{j=1}^J)]^2$$

其中 $(y_{t+1} - f_{b-1}(x))$ 即为 $f_{b-1}(x)$ 的残差，模型可以通过不断对残差建立回归树模型减少估计误差。

为了提高模型的预测能力，我们参考 Rossi<sup>[9]</sup>在传统提升回归树的基础上进行了两项改进。第一，我们加入收缩方法（Shrinkage）提高模型预测能力。Friedman<sup>[8]</sup>研究发现，收缩方法通过聚合大量子模型降低模型过拟合的风险，具体而言：

$$f_b(x) = f_{b-1}(x) + \lambda \sum_{j=1}^J c_{j,b} I\{x \in S_{j,b}\}$$

其中 $f_b(x)$ 为第  $b$  次回归树拟合值， $\lambda$ 为收缩系数， $c_{j,b}$ 为第  $b$  次子模型的第  $j$  个叶节点估计的拟合值， $S_{j,b}$  为第  $b$  次生成树模型的第  $j$  个叶节点。收缩系数 $\lambda$ 取值在 0 到 1 之间，代表了每一轮迭代的拟合值对最终估计值的影响程度。第二，为了提高结果的稳定性，降低结果的方差，避免过拟合，我们引入了袋装算法（Bagging, Bootstrap Aggregation）<sup>[10]</sup>。具体来说，在每次迭代中不放回地抽取一半训练样本，用得到的子样本训练出下一个回归树。

## 10. RF 算法

随机森林算法（RF）同样是以决策树为基分类器的集成算法。不同于提升回归树的集成对象是层数较少的决策树，RF 算法集成的是大量层数较大但相关性较小的、建立在不同样本子集上的决策树。首先，从训练集中随机抽取部分样本建立决策树，在每个节点从  $p$  个特征中随机抽取部分特征对样本进行分裂，直到生成完整的一棵决策树。最终的随机森林模型就是所有单个决策树的集成。在决策树的构建中主要有三个超参数较为重要，分别是集成决策树的数量、单个决策树的最大深度和每个节点随机选取的特征数。随机森林算法通过降低每个子树的方差和子树之间的相关性提高模型的泛化能力，具有较强的鲁棒性，在实际研究中表现较好。

## 11. Adaboost 算法

Adaboost 算法也是经典的提升算法之一，它将多个基分类器组合成为效果较好的集成模型，基分类器一般采用决策树，这一算法的核心在于对误差较大的样本的权值进行加强，将权值更新后再用来训练下一个新的弱分类器，一直迭代直到获得表现足够好的模型。在第一次迭代中，每个训练样本点被



赋予相同的权重，构建出一个决策树。在训练中，被估计得越准确的样本在下一次迭代中的权重就被降低，而未被准确估计的样本在下一次迭代中权重将被提高。然后，更新权值后的样本集被用于下一个分类器的训练，如此迭代，最终形成强分类器。这一算法不需要任何样本筛选或特征筛选，且具有较好的样本外预测能力。

## 12. BRT 算法与其他机器学习模型优劣分析

机器学习算法中可以应用于变量预测的常见的监督学习可以分为线性回归模型、决策树及其集成算法、支持向量机算法和基于神经网络的算法。本节对上述几类机器学习方法的优劣进行归纳：

线性回归算法主要包括 Lasso 回归、Ridge 回归和 Enet 回归，这些方法分别通过 L1 正则化、L2 正则化、结合 L1 和 L2 正则化实现对变量系数的收缩或筛选，解决变量共线性问题。线性的机器学习模型的优势在于训练速度快、模型相对容易理解，但难以挖掘变量之间的非线性关系。

决策树及其集成算法包括决策树、随机森林、梯度提升树等模型。单个决策树模型容易可视化、易理解，且不需要对数据进行预处理，但泛化性能较差。决策树的集成方法则对单个决策树赋予一定权重，将所有决策树结果加总获得最后的预测值，避免了单个决策树对结果产生过大影响。包括 BRT 在内的决策树集成方法对缺失值与异常值具有很好的鲁棒性，不要求对数据进行缩放，且具有出色的泛化能力，但相对线性模型而言训练时间较长。

支持向量机方法解决回归问题的原理是通过将样本数据映射到高维特征空间中，使其具有较好的线性回归特征后对样本进行线性回归，然后返回到原始空间。支持向量机可以应用于线性回归与非线性回归，它的表现在低维和高维数据上都有不错的表现，缺点是对数据预处理与超参数的选择非常敏感，而且难以对其结果做出直观解释。

神经网络算法是灵活性最大、最为复杂的机器学习算法之一，通过隐藏层个数、每层神经元个数、正则化方式等控制神经网络的具体结构和复杂度。通过神经网络模型能够获取大量数据中包含的信息，对数据集具有相当好的挖掘能力，但神经网络对数据量的要求较高，训练时间长，对数据预处理方式与参数缩放比较敏感。

#### 附录四： 股票市场样本外预测结果

**表 A-4 不同模型对中国股票市场收益率、波动率和最优风险权重的样本外预测效果（%）**

Table A-4 Various model's out-of-sample prediction of China stock market's return, volatility, and optimal allocation

	扩展窗口		滚动窗口	
	R <sup>2</sup>	Acc	R <sup>2</sup>	Acc
Panel A: 收益率预测				
Prevailing Mean	-0.57	55.13	0.18	53.85
Linear Model	0.25	50.00	-5.10	54.49
PLS	-2.75	55.77	0.13	54.49
Ridge	-1.37	50.00	0.62	55.77
Lasso	-0.62	55.13	0.41	54.49
ENet	-0.62	55.13	0.41	54.49
BRT	3.72	58.97	3.96	62.18
RF	3.17	57.69	3.82	61.54
Adaboost	3.43	55.77	2.42	58.33
NN1	-113.36	55.13	-125.47	54.49
NN2	-123.15	50.64	-93.33	50.00
NN3	-64.48	49.36	-201.06	50.64
NN4	-78.87	53.21	-118.07	52.56
NN5	-76.91	51.92	-123.26	43.59
LSTM	-21.47	54.49	-21.65	49.36
Panel B: 波动率预测				
GARCH	56.47	73.08	53.44	71.79
MIDAS	63.58	75.64	63.54	75.64
BRT	64.59	73.08	61.65	75.64
RF	63.23	74.36	59.83	75.64
Adaboost	62.09	74.36	51.47	76.92
NN1	-4211.32	56.41	-8029.14	53.85
NN2	-91.56	62.82	-884.74	55.13
NN3	-190.63	47.44	32.49	58.97
NN4	45.23	65.38	48.76	60.26
NN5	39.69	48.72	26.78	44.87
LSTM	-881.10	43.59	-736.14	47.44
Panel C: 最优风险权重预测				
Prevailing Mean	3.88	55.13	3.88	55.13
Linear + GARCH	6.15	50.00	3.81	54.49
Linear + MIDAS	6.38	50.00	3.09	54.49
BRT	6.51	58.97	6.73	62.18
RF	5.78	57.69	7.15	61.54
Adaboost	6.08	55.77	6.62	58.33
NN1	-609.31	51.92	-30.97	50.00
NN2	-8746.82	47.44	-202.55	52.56
NN3	-689.40	49.36	-2977.07	53.21
NN4	-3067.25	53.21	-297.02	53.21
NN5	-31.95	51.92	-61.41	44.23
LSTM	-97.94	51.28	-807.16	51.92

表A-4结果显示，树类算法（包括BRT、RF和Adaboost方法）在收益率预测与最优风险权重上的表现整体上优于基准、线性类算法和神经网络类算法，在波动率预测上相对MIDAS没有明显优势；BRT方法总体上表现最好；PLS等线性模型与基准的历史均值估计或线性模型估计相比表现相对较弱；而由于样本量限制，神经网络类算法在收益率预测与波动率预测中表现均极差。

## 附录五：变量重要性分析

### 1. 特征重要性计算方法

特征重要性分析通过计算每个变量作为回归树内部节点时划分样本累计减少的误差平方和来判断变量重要性. 一个变量对模型越重要, 则对样本越有区分度, 也就可以通过以此变量为节点划分样本, 使得样本误差平方和快速降低. 具体而言, 对单个回归树  $T$ , 我们定义变量  $k$  的重要性指标为:

$$L_k^2(T) = \sum_{t=1}^{J-1} l_t^2 I(v(t) = k)$$

其中  $l_t^2$  表示在节点  $t$  分支前后的样本误差平方和,  $v(t)$  表示节点  $t$  用于分支的变量, 则  $L_k^2(T)$  表示总共  $J-1$  个内部节点中, 所有按变量  $k$  划分子树的节点在划分前后误差减少值的总和. 提升回归树是多个回归树的加总, 所以为了度量变量对整个提升回归树模型的重要性, 我们将所有单个回归树中的变量重要性做平均, 具体表示为:

$$L_k^2(T) = \sum_{t=1}^{J-1} l_t^2 I(v(t) = k)$$

$L_k^2$  即为对  $M$  个回归树做等权平均则得到变量  $k$  的变量重要性. 为了对比不同变量之间的相对重要性, 对所有变量重要性做标准化处理, 使得所有变量的相对重要性之和为 100%, 即变量的相对特征重要性为:

$$\bar{L}_k^2 = L_k^2 / \sum_{i=1}^K L_i^2$$

### 2. 部分依赖图计算方法

部分依赖图 (PDP) 通过图线体现变量对被预测变量的边际效应, 图线变化越明显说明变量对预测模型的重要性越大, 此外, 图线的形状可以直观体现变量与被预测变量之间的关系是否为线性. 具体而言, 在用训练集得到一个模型后, 我们固定其他变量的取值, 改变所有样本中变量  $k$  的取值为所有样本的  $x_k$  最小值得到一组新的样本, 用已训练的模型对新的样本进行预测, 对新的预测结果取平均值, 取不同的  $x_k$  值重复此步骤得到部分依赖曲线  $f_k(x_k)$ . 这一过程可以表示为:

$$\bar{f}_k(x_k) = \frac{1}{N} \sum_{i=1}^N f(x_{i,k}, X_{-k})$$

其中  $x_{i,k}$  表示变量  $k$  的全部取值集合  $\{x_{1,k}, x_{2,k}, \dots, x_{N,k}\}$ ,  $f(x_{i,k}, X_{-k})$  表示在固定除了变量  $k$  外的其他变量值, 不断改变变量  $k$  取值得到的一系列预测值,  $\bar{f}_k(X_k)$  表示对一组新样本进行简单平均, 得到变量  $k$  在整个区间的总体部分依赖曲线. 为了同时研究两个变量与被预测变量的边际效应可以将上述过程推广到二维空间, 只需将取值集合变为两个变量值的成对组合即可.

### 3. 预测变量部分依赖图

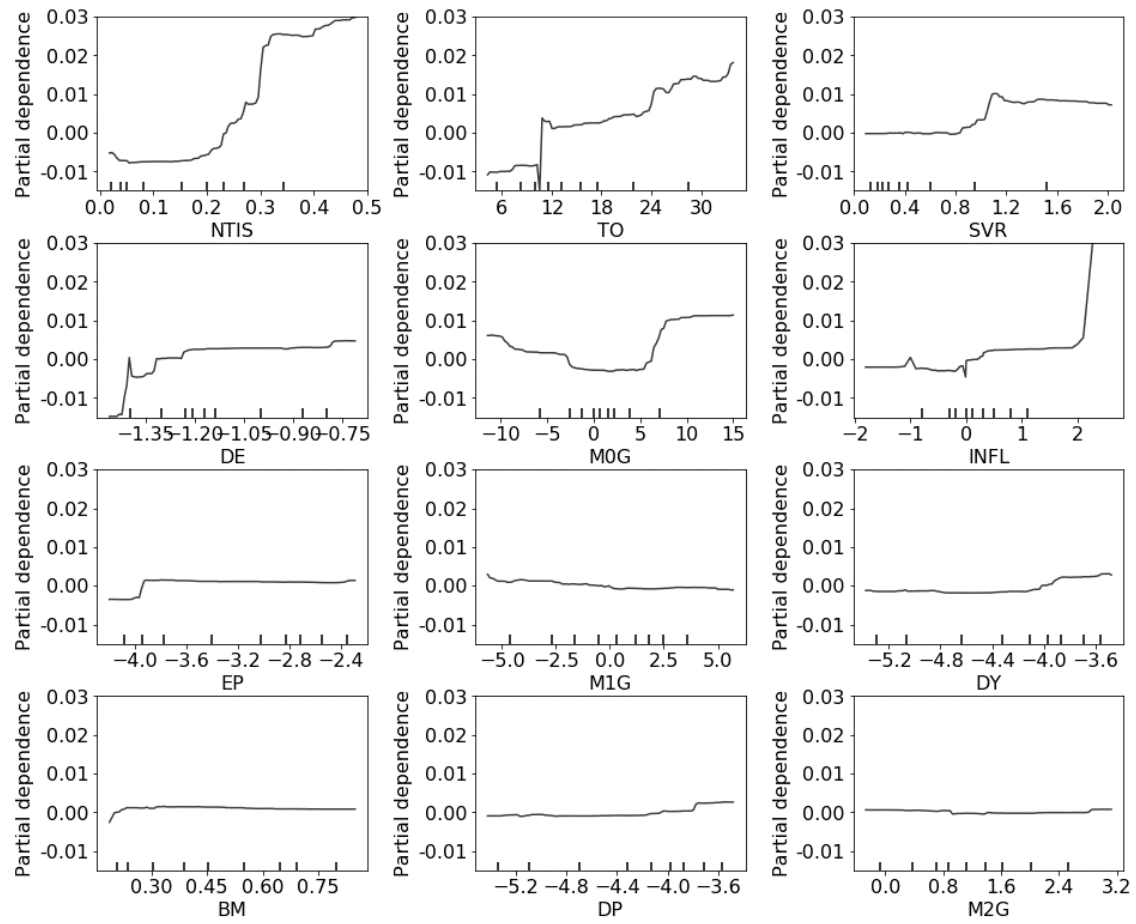


图 A-1 提升回归树模型预测收益率的单变量部分依赖图

Fig. A-1 Univariate partial dependence plot of BRT prediction of return

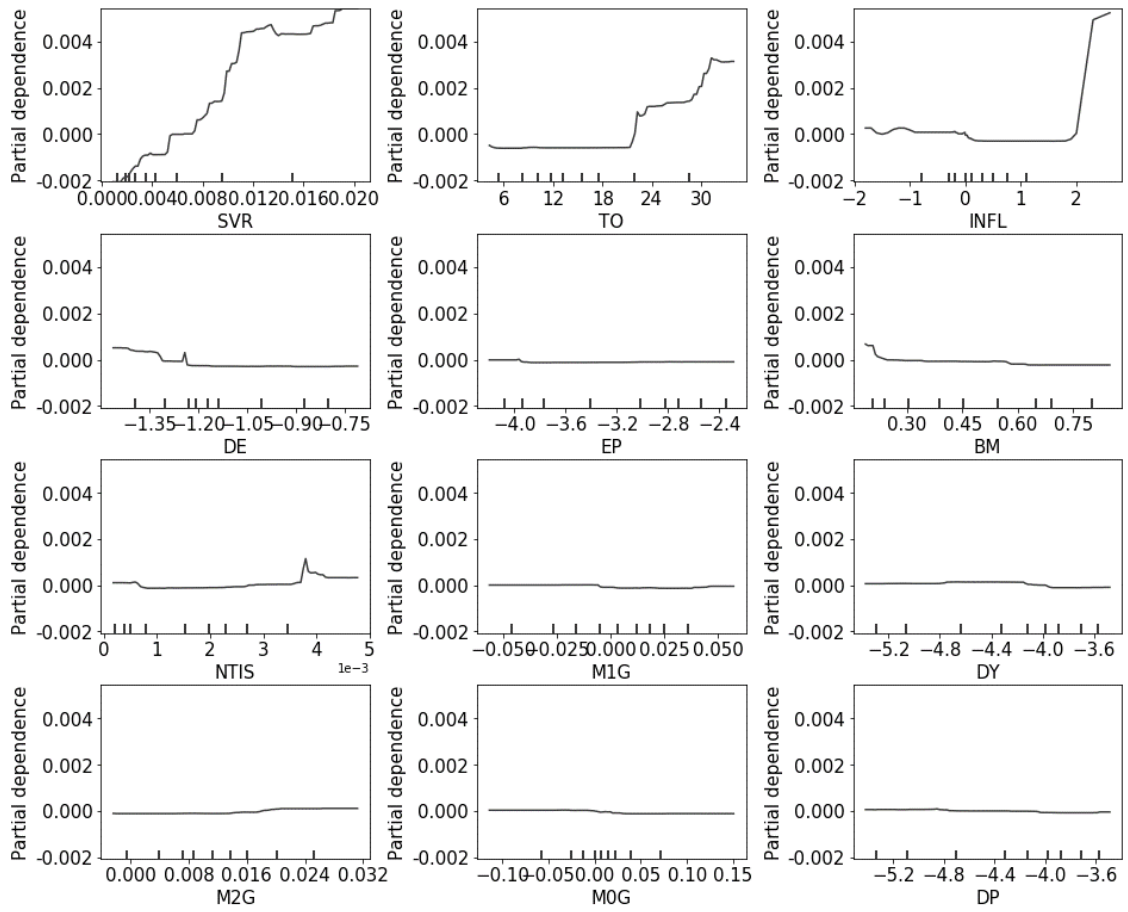
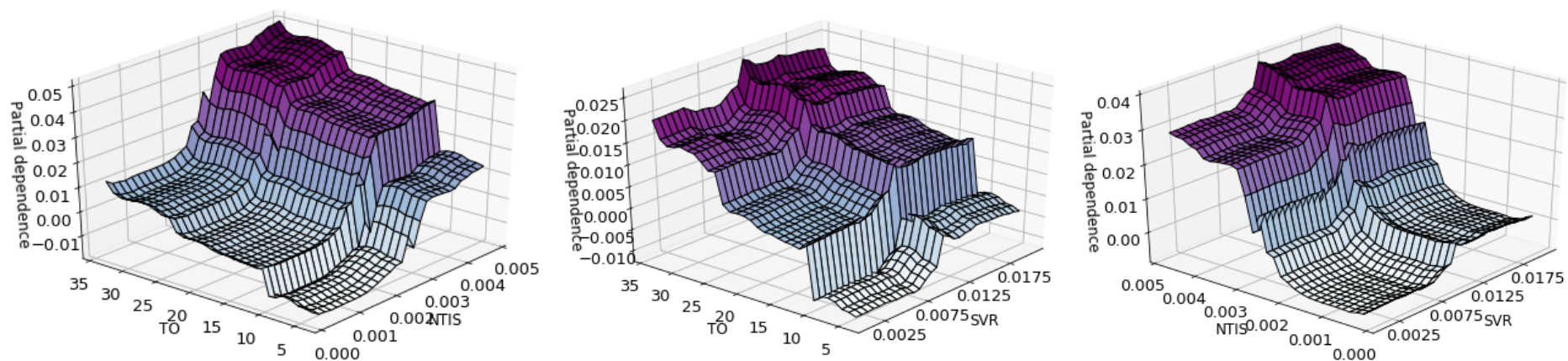
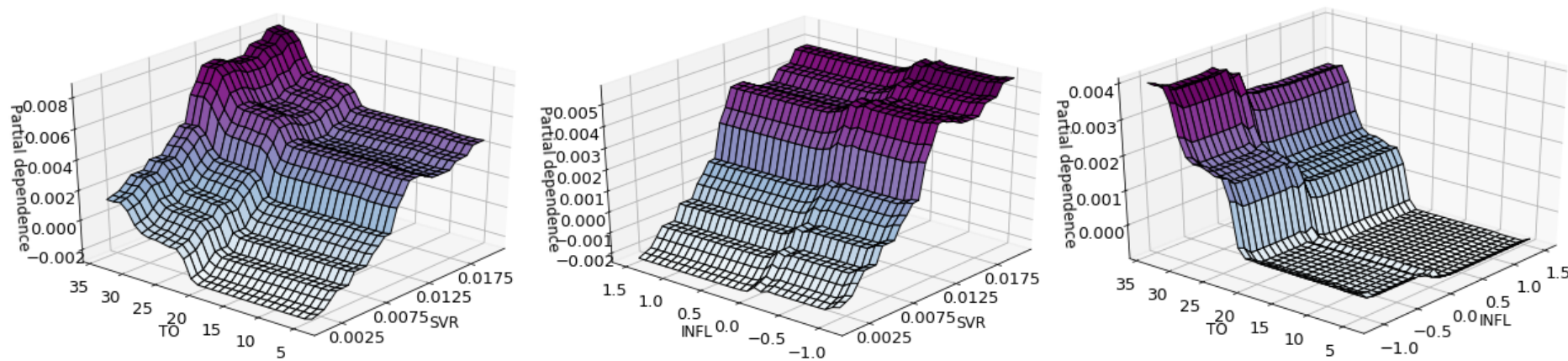


图 A-2 提升回归树模型预测波动率的单变量部分依赖图

Fig. A-2 Univariate partial dependence plot of BRT prediction of volatility



(a) 预测收益率



(b) 预测波动率

图 A-3 提升回归树模型预测收益率和波动率的双变量部分依赖图  
Fig. A-3 Bivariate partial dependence plot of BRT prediction of return and volatility

附录六：投资组合经济效益

1. 存在交易成本的经济效益

表 A-5 存在交易成本的投资组合表现

Table A-5 Portfolio performance with transaction costs

	Mean	Sharpe	CAPM $\alpha$	Mean	Sharpe	CAPM $\alpha$
	低交易成本（0.1%）			高交易成本（0.5%）		
Panel A: BRT						
依据超额收益率	1.16%*	0.1310	0.54%**	1.11%	0.1233	0.48%*
	(1.67)		(2.01)	(1.58)		(1.79)
最优风险资产权重	1.03%	0.1252	0.46%*	0.93%	0.1094	0.36%
	(1.59)		(1.69)	(1.42)		(1.31)
Panel B: Linear + GARCH						
依据超额收益率	0.63%	0.0613	0.05%	0.52%	0.0450	-0.06%
	(0.99)		(0.17)	(0.80)		(-0.19)
最优风险资产权重	0.82%	0.1057	0.32%	0.69%	0.0821	0.19%
	(1.46)		(1.11)	(1.21)		(0.64)
Panel D: Linear + MIDAS						
依据超额收益率	0.63%	0.0613	0.05%	0.52%	0.0450	-0.06%
	(0.99)		(0.17)	(0.80)		(-0.19)
最优风险资产权重	0.72%	0.0938	0.24%	0.59%	0.0691	0.11%
	(1.31)		(0.87)	(1.05)		(0.38)

注：此表展示了存在交易费用与不同投资限制条件下，扩展窗口 BRT 模型预测的收益率表现。\*、\*\*、\*\*\*分别表示在 10%、5%和 1%水平下显著。

现实中的交易存在费用，所以我们比较了存在双边交易成本的情况下 BRT 算法的表现。表 A-5 展示了在低交易成本（0.1%）和高交易成本（0.5%）条件下，BRT 模型与两种对比模型的收益率表现<sup>[11]</sup>。对比表 A-5 中各模型的收益率表现，可以发现交易成本的存在对 BRT 模型相较于其他模型的优势影响不大，BRT 方法依然得到了更高的收益率、夏普比率和 CAPM  $\alpha$ 。

2. 不同风险厌恶系数下的投资者效用

我们还度量了不同策略的投资组合对均值-方差效用投资者的实现效用：

$$U_p(\cdot) = \frac{1}{T} \sum_{t=1}^T [r_{p,t} - \frac{1}{2} \gamma \hat{w}_t^2 \sigma_t^2] \quad (1)$$

其中  $\sigma_t^2$  表示 t 月的市场日收益率平方和， $[r_{p,t} - \frac{1}{2} \gamma \hat{w}_t^2 \sigma_t^2]$  表示 t 期投资者实现的效用，对全部预测区间效用取平均值得到平均实现效用为  $U_p$ 。风险资产权重估计值波动过于剧烈可能影响了策略表现，Ter Horst 等<sup>[12]</sup>和 Maenhout<sup>[13]</sup>将这个问题定义为估计的不确定性。为了解决这个问题，设定风险厌恶系数的两倍作为“伪风险厌恶系数”，用伪风险厌恶系数确定风险资产权重，以降低权重估计值的波动<sup>[14]</sup>。表 A-6 中我们设定了不同的风险厌恶系数，并且考虑了估计的不确定性，计算不同模型相应策略所得收益率序列的实现效用。

表 A-6 不同风险厌恶水平下各模型平均实现效用

Table A-6 Average realized utility of models under different risk aversions

	$\gamma=2$	$\gamma=4$	$\gamma=6$	$\gamma=8$	$\gamma=10$	$\gamma=2$	$\gamma=4$	$\gamma=6$	$\gamma=8$	$\gamma=10$
	未考虑估计不确定性					考虑估计不确定性				
Panel A: BRT										
依据超额收益率	0.41	0.31	0.27	0.26	0.25	0.41	0.31	0.27	0.26	0.25
最优风险资产权重	0.61	0.36	0.31	0.35	0.30	0.69	0.41	0.34	0.35	0.29
Panel B: Linear + GARCH										

依据超额收益率	-0.57	-0.18	-0.05	0.01	0.05	-0.57	-0.18	-0.05	0.01	0.05
最优风险资产权重	0.39	0.00	0.07	0.22	0.25	0.80	0.30	0.27	0.30	0.27
Panel C: Linear + MIDAS										
依据超额收益率	-0.57	-0.18	-0.05	0.01	0.05	-0.57	-0.18	-0.05	0.01	0.05
最优风险资产权重	0.30	0.01	0.07	0.22	0.25	0.91	0.25	0.24	0.28	0.26
Panel D: Passive Strategy										
市场投资组合	0.10	-0.54	-1.17	-1.80	-2.43					

注：此表展示了考虑估计不确定性与不同风险厌恶系数条件下，不同模型相应策略实现的效用水平，表中数值均为百分数。

## 附录七：稳健性检验

### 1. 样本区间稳健性检验

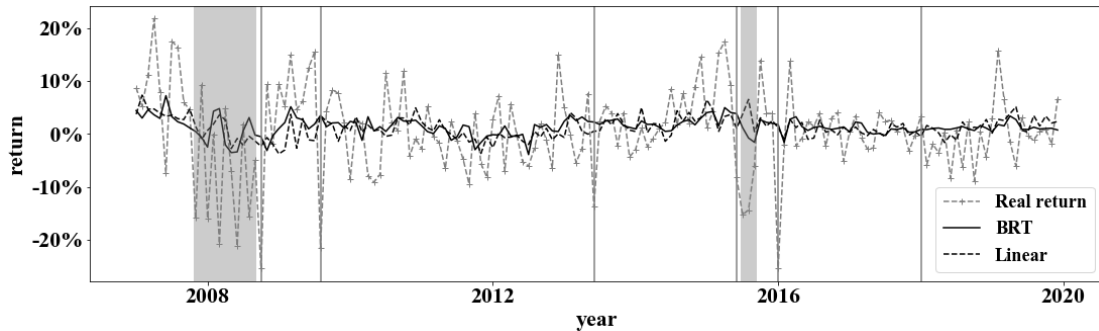


图 A-4 BRT、线性模型预测收益率与实际收益率折线图

Fig. A-4 BRT、Linear Model Prediction and Real Return Plot

注：此图展示了真实收益率序列、线性模型和 BRT 模型预测的收益率时间序列，图中的灰色阴影表示了 2008 年金融危机、2015 年股灾和 2016 年初股灾。

表 A-7 金融危机后(2009 年-2019 年)样本外预测结果

Table A-7 Out-of-sample prediction results after the financial crisis (2009-2019)

	扩展窗口		滚动窗口	
	R <sup>2</sup> (%)	Acc (%)	R <sup>2</sup> (%)	Acc (%)
Panel A: 超额收益率预测				
BRT	3.07	57.58	2.88	60.61
Prevailing Mean	0.38	55.30	0.20	53.79
Linear Model	-7.26	46.97	-21.45	52.27
Panel B: 波动率预测				
BRT	54.92	74.24	51.69	71.21
MIDAS	53.85	69.70	53.85	69.70
GARCH	49.57	72.73	49.57	72.73
Panel C: 最优风险资产权重预测				
BRT	6.00	57.58	6.28	60.61
Linear + GARCH	4.89	46.97	2.22	52.37
Linear + MIDAS	1.97	46.97	1.50	52.37

注：此表报告了将金融危机后的时期，即 2009 年至 2019 年，作为样本外预测时期时，不同模型对市场收益率、波动率与最优风险投资权重的样本外预测表现，包括各种模型的样本外 R<sup>2</sup> 与方向预测精度 (Acc)。表中数据均表示百分数。

从表 A-7 的数据可以看出，将样本外预测区间改为金融危机后时期，提升回归树模型仍然整体优于其他模型，说明提升回归树模型的预测能力较为稳健。

### 2. 超参数稳健性检验

为了检验提升回归树方法的预测效果是否对模型的超参数敏感，我们还设置了不同的迭代次数 (Iteration) 和不同的学习率 (Learning Rate)，计算了相应预测模型预测超额收益率的样本外 R<sup>2</sup> 和方向精度。



**表 A-8 不同超参数下 BRT 方法预测超额收益率结果**

Table A-8 Results of excess return prediction of BRT under different iteration times

(Iterations, Learning Rate)	扩展窗口		滚动窗口	
	R <sup>2</sup> (%)	Acc (%)	R <sup>2</sup> (%)	Acc (%)
(100, 0.01)	2.50	55.13	3.18	59.62
(500, 0.01)	2.51	58.33	-1.97	57.05
(5000, 0.001)	2.10	57.05	0.48	57.05
(10000, 0.0001)	3.14	55.77	3.96	62.18

从表 A-8 可以看出，设置不同的迭代次数和学习率，不论采用扩展窗口估计还是滚动窗口估计，样本外 R<sup>2</sup> 均超过了 3%，方向精度也都超过了 56%，说明提升回归树方法的预测能力对迭代次数设置是较为稳健的。改变学习率参数，在扩展窗口和滚动窗口情形下，BRT 模型的样本外 R<sup>2</sup> 与方向精度仍远高于历史均值估计和线性方法。<sup>[15]</sup>

#### 参考文献:

- [1] Welch I, Goyal A. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction[J]. Review of Financial Studies, 2008, 21 (4):1455-1508.
- [2] Zhang H, He Q, Jacobsen B, et al. Forecasting Stock Returns with Model Uncertainty and Parameter Instability[J]. Journal of Applied Econometrics, 2020, 35 (5):629-644.
- [3] Pettenuzzo D, Timmermann A, Valkanov R. Forecasting Stock Returns Under Economic Constraints[J]. Journal of Financial Economics, 2014, 114 (3):517-553.
- [4] 姜富伟, 涂俊, 周国富, 等. 中国股票市场可预测性的实证研究[J]. 金融研究, 2011, (9):107-121.  
Jiang F, Tu J, Zhou G, et al. An Empirical Study On the Predictability of China's Stock Market[J]. Journal of Financial Research, 2011, (9):107-121.(in Chinese)
- [5] 蒋志强, 田婧雯, 周炜星. 中国股票市场收益率的可预测性研究[J]. 管理科学学报, 2019, 22 (04):92-109.  
Jiang Z, Tian J, Zhou W. Return Predictability in the Chinese Stock Markets[J]. Journal of Management Sciences in China, 2019, 22 (04):92-109.(in Chinese)
- [6] Hansen P R, Lunde A. A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1, 1)?[J]. Journal of Applied Econometrics, 2005, 20 (7):873-889.
- [7] Ghysels E, Santa-Clara P, Valkanov R. Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies[J]. Journal of Econometrics, 2006, 131 (1-2):59-95.
- [8] Friedman J H. Greedy Function Approximation: A Gradient Boosting Machine[J]. Annals of Statistics, 2001, 29 (5):1189-1232.
- [9] Rossi A G. Predicting Stock Market Returns with Machine Learning[J]. Working Paper, 2018.
- [10] Breiman L. Bagging Predictors[J]. Machine Learning, 1996, 24 (2):123-140.
- [11] Rapach D E, Ringgenberg M C, Zhou G. Short Interest and Aggregate Stock Returns[J]. Journal of Financial Economics, 2016, 121 (1):46-65.
- [12] Ter Horst J, de Roon F, Werker B. Incorporating Estimation Risk in Portfolio Choice[J]. Working Paper, 2000.
- [13] Maenhout P. Robust Portfolio Rules and Asset Pricing[J]. Review of Financial Studies, 2004, 17 (4):951-983.
- [14] Marquering W, Verbeek M. The Economic Value of Predicting Stock Index Returns and Volatility[J]. Journal of financial and Quantitative Analysis, 2001, 39 (2):407-427.
- [15] 徐浩峰, 朱松. 机构投资者与股市泡沫的形成[J]. 中国管理科学, 2012, 20 (4):18-26.  
Xu H, Zhu S. Does the Trade From Institutional Investors Induce Stock Bubble[J]. Chinese Journal of Management Science, 2012, 20 (4):18-26.(in Chinese)