

机器学习驱动的基本面量化投资研究¹

李 斌，邵新月，李玥阳

武汉大学经济与管理学院、武汉大学金融科技研究中心

2019 年 3 月 15 日

¹ [作者简介] 李斌，武汉大学经济与管理学院副教授、博士生导师；邵新月，武汉大学经济与管理学院硕士研究生；李玥阳，武汉大学经济与管理学院硕士研究生。通讯作者：李斌，手机：13971237126，电子邮箱：binli.whu@whu.edu.cn。

机器学习驱动的基本面量化投资研究

李 斌，邵新月，李玥阳

武汉大学经济与管理学院、武汉大学金融科技研究中心

[摘要] 基本面量化投资是近年来金融科技和量化投资研究的新热点。基于 1997 年 1 月至 2018 年 10 月中国 A 股市场的 96 项异象因子，本文采用预测组合算法、Lasso 回归、岭回归、弹性网络回归、偏最小二乘回归、梯度提升树、极端梯度提升树、集成神经网络和深度前馈网络等 9 种机器学习算法构建超额收益预测模型，并构建投资组合。实证结果显示，机器学习算法能够自动识别因子间的线性和非线性模式获得更好的投资绩效，而非线性机器学习算法获得的绩效提升尤为突出。基于深度前馈网络预测的多空组合能够获得 2.83% 的月度收益，而基于 OLS 回归预测构建的多空组合月度收益仅为 1.90%。即便在考虑到中国股票市场的做空限制和交易成本等问题，深度前馈网络模型构建的做多组合平均月度收益也达到 2.46%，而同期上证综合指数的月度收益仅为 0.59%，因此机器学习模型能够大幅度提升基本面量化投资策略的绩效。最后，本文对预测模型中特征的重要性进行了筛选，代表性特征显示，基于梯度提升树和深度前馈网络筛选后的特征集合能够获得超过全样本特征集合的投资绩效，交易摩擦类因子和动量因子在中国 A 股市场具有更强的预测效力。

[关键词] 基本面量化投资；机器学习；市场异象因子；深度前馈网络；

[中图分类号] F830.593 **[文献标识码]** A **[文章编号]** 1006-480X (2018)

Machine Learning-driven Quantamental Investing

Li Bin, SHAO Xinyue, LI Yueyang

Abstract: Quantamental investing is an emerging hot topic in financial technology and quantitative investments. Based on 96 anomaly factors in Chinese markets ranging from January 1997 to October 2018, we adopt forecast combination, LASSO regression, Elastic Net, Partial Least Square, Gradient Boost Tree, Extreme Gradient Boost Tree, Ensemble Neural Network and Deep Feedforward Network to build excess return forecasting model and construct portfolios. Empirical evidence shows that machine learning algorithms can automatically identify the linear and nonlinear patterns to obtain better investment performance, and nonlinear machine learning algorithms improve more. Long-short portfolios based on the forecast of Deep Feedforward Network can obtain a monthly return of 2.83%, while long-short portfolio based on traditional OLS regression can achieve a monthly return of 1.90%. Even considering the short-sell constraint and transaction costs, long short portfolio based on deep feedforward network can obtain a monthly return of 2.46%, and the return of the Shanghai Composite index's return is only 0.59%. In summary, machine learning models can significantly improve quantamental investing's performance. Finally, we select important features based on their marginal importance. Analysis on the important features shows that the feature set selected using gradient boost tree and deep feedforward network can provide a better performance than the investment performance based on all 96 features. Trading friction factors and momentum factors demonstrate a better predictive ability in the Chinese markets.

Key words: Quantamental investing; Machine Learning; Market Anomaly Factors; Deep Feedforward Network

JEL Classification:C8, G0, O1

一、引言

1. 研究动机

基本面量化投资从能够提供超额收益的市场异象因子出发构建投资组合,是目前金融科技和量化投资研究中的新热点^[1,2]。市场异象的研究已经提出了数以百计宣称能够提供超额收益的因子,但后续的样本外检验发现大部分异象因子难以持续地提供超额收益。异象因子的涌现也对现有方法提出了技术难题:其一,资产风险溢价的候选因子多达数百个^[3,4],且很多因子极为相近^[5]。而传统的组合排序和FM回归并未综合考虑各因子,也未考虑因子间的交互作用。其二,当因子维度变大时,线性和非线性的考虑使得预测函数形式的搜索复杂度急剧增加,几乎无法通过人工去指定。遗憾的是,现有研究并未提供在高维因子与预测函数形式选择方面的建议。以上两个技术难题呼唤着新研究方法的介入,前美国金融学会会长Cochrane (2011, 1061页)^[6]认为,在处理如此众多的因子时,将不得不使用“不同的研究工具”(“I suspect we will have to use different methods.”)。

机器学习和深度学习正是其中强有力的候选方法。首先,机器学习天然地适用于资产风险溢价的研究。度量资产的风险溢价本质上是一个预测问题,而机器学习和深度学习旨在自动地寻找数据中的复杂结构和模式来辅助人类进行预测^[7]。其次,针对两个技术难题,机器学习的三个特性使得其适用于该预测问题。首先,机器学习方法论提供了众多备选的预测函数形式。其次,很多机器学习方法(比如深度神经网络等)专门被设计用于逼近复杂的非线性关系。最后,参数正则化(Regularization)和模型选择(Model Selection)等技术使得在选择预测函数时亦不易过拟合(Overfitting)而导致虚假发现。

由此,本文采用机器学习方法深入研究了基本面量化投资。首先,我们收集了中国A股市场96个异象因子,基于机器学习技术建立异象因子-超额收益预测模型,并基于预测结果构建投资组合。研究结果发现,机器学习算法能够自动识别因子间的线性和非线性模式获得更好的投资绩效,而非线性机器学习算法获得的绩效提升尤为突出。基于深度前馈网络预测的多空组合能够获得2.83%的月度收益,而基于OLS回归预测构建的多空组合月度收益仅为1.90%。即便在考虑到中国股票市场的做空限制和交易成本等问题,深度前馈网络模型构建的做多组合平均月度收益也达到2.46%,而同期上证综合指数的月度收益仅为0.59%,因此机器学习模型能够大幅度提升基本面量化投资策略的绩效。

2. 文献综述

(1) 机器学习在资产定价中的应用

近年来,机器学习在资产定价中的应用逐渐增多,获得国内外广泛关注。具体而言,机器学习包含非监督式学习(Unsupervised Learning)和监督式学习(Supervised Learning)两大类。非监督式学习主要关注自动寻找规律,通常应用在图像视频识别、文本数据挖掘等领域。监督式学习主要关注点则是进行预测,对于一个给定的训练数据集 x 和结果变量集 y ,监督式机器学习通过最小化均方误差自动寻找 x 与 y 之间的映射关系,以此为基准在新数据集中对 y 值进行预测。监督式学习在计量经济学领域具有广泛的应用价值,众多研究人员对其进行了探索尝试。Rapach(2013)^[8]采用LASSO回归对全球各国股票月度收益率进行预测,发现以美国为主导的工业化国家股票回报率对非工业化国家股票回报率有显著的预测效果,并将其归因于信息的渐进扩散。Feng, Giglio, and Xiu(2017)^[9]提出了基于“套索”(LASSO)方法来衡量因子对资产定价的贡献,发现盈利性和投资因子比之前发现的数百种更具有统计上显著的解释力;后续也有学者采用了“动态套索”(Adaptive LASSO)^[10]和“动态组套索”(Adaptive Group LASSO)^[11]等方法进行因子筛选,选出更具有统计意义的因子集合。Kelly(2015)^[12]采用机器学习中的降维方法,解决了多因子可能面临的“维数爆炸问题”,并综合多个模型的预测结果构建股票投资组合,取得显著的正向收益。基于上述结论,作者采用机器学习中的自助法对1967年-2016年上百个资产定价领域因子的有效性进行了检验。Fischer and Krauss(2018)^[13]采用长短期记忆模型(Long Short-Term Memory, LSTM),利用日频收益率数据预测股票相对于其截面中值收益率的涨跌方向,而根据长短期记忆模型选出的投资组合绩效显著优于其他线性模型。王文波等(2010)^[14]利用经验模态分解(Empirical Mode Decomposition, EMD)和神经网络算法对股价的日频数据进行分解并预测股票的收盘价,发现同传统方法比,上述算法具有更高的预测效力。李斌等(2017)^[15]分别采用支持向量机、神经网络、Adaboost算法,利用股票日频交易数据构建19项技术指标,预测股价涨跌方向,研究发现通过以上机器学习算法进行的股票涨跌方向预测具有更高的准确率,而根据预测所构建的投资组合也取得了更好的投资绩效。此外,也有部分研究试图通过设计新的集成方式以获得更好的预测效果。Lewellen(2015)^[16]发现通过FM回归方法综合15个因子能够很好地预测股票超额收益。Moritz and Zimmermann(2016)^[17]设计了“基于树结构的条件组合排序方法”(Tree-based conditional portfolio sorts),发现基于该算法预测所构建的策略能够获得两倍于FM回归所构建策略的收益。Demiguel等(2019)^[18]从

投资者效用的角度预测截面收益公司特征,发现从投资者效用的角度预测截面收益公司特征,发现6个公司特征可以独立地预测平均收益。Gu et al. (2018)^[5]则检验了常见的机器学习算法在美国市场上的表现,研究发现机器学习模型可以有效地超越线性模型。

综上,与传统的线性模型相比,机器学习算法往往无需人工干预而能够自动识别金融数据背后的隐含特征。金融市场数据已被证明存在长记忆、尖峰厚尾、分形等非线性特征传统线性模型对此无能为力,而机器学习模型却能识别出金融数据中的非线性规律。此外,以上研究中模型的输入变量多为价格及技术指标或财务指标,指标的选取缺乏系统的理论支撑。而本文的输入变量则采用在过往研究中被提出的数百项异象因子,以增强模型对股票风险溢价的解释力度。

(2) 多因子策略研究

作为应用最广泛并且最重要的量化选股模型之一,国内外学者为对多因子选股模型做了大量的研究。Fama和French(1993)^[19]提出的三因子模型以及随后在此基础上提出的五因子模型成为多因子研究的典范^[20],为之后的研究提供蓝本。众多学者尝试从不同角度解释因子模型与资产收益的关系,使得基于因子模型的现代投资组合理论不断发展和进步。Ng, Engle和Rothschild (1992)^[21]从影响股票收益渠道的角度分别定义了动态和静态的因子,并区分了超额收益的动态和静态因子,研究结果表明资本市场比较像一类动态因子,规模较大的企业对生产经营因子相对更加敏感,而小型企业则对流动性因子更加敏感。Driessen, Klaassen和Melenberg (2003)^[22]分析总结了影响套期保值和资产定价结果的因子,并将这些因子应用于估计模型参数,以获得最佳的套期保值和定价结果,实证结果表明多因子模型比单因子模型在套期保值和互换等衍生品定价方面具有更好的表现。Harvey和Liu(2015)^[23]通过对现有文献中已经发现的380个异象因子的研究,认为由于数据挖掘存在的可能性,今年来新发现的大量因子尽管在具有统计意义上的显著性,但并没有切实的经济学含义。Green, Hand和Zhang (2017)^[24]利用1980-2014年美国股票市场数据构建了94项因子,在此基础上进行Fama-Macbeth回归,最终发现12个因子在此区间内对股票月度收益具有显著的预测效果。朱英伦和刘杰(2018)^[25]汇总了近几十年来在中国期刊发表的因子文章,发现在美国最早发现的接近300个因子中有29个在中国市场显示有超额收益,比例在百分之十左右。胡熠和顾明(2018)^[26]从安全性、便宜性和质量三个维度共选取8个异象因子构造综合性指标B-score,并将其应用于中国A股市场,证明了巴菲特的价值投资策略在中国市场的适用性。

随着异象因子数目集合的不断扩张,传统多因子策略面临的瓶颈也逐渐凸显。首先,随

随着因子维度的不断增大，因子间的相似性也随之提高，传统的回归模型难以识别因子间的交互作用。其次，传统FM回归中人工指定了因子间的线性关系，而此类假设与实际情形存在显著差异，此外在高维度的数据集中人工设定非线性函数形式也难以实现。机器学习算法的应用能够在很大程度上解决上述问题。一方面，机器学习算法提供了众多备选函数形式以识别特征间的复杂非线性关系，另一方面，通过加入各类限制条件（例如参数正则化）能够有效避免模型的过拟合问题。Light, Maslov, and Rytchkov (2017)^[27]采用“偏最小二乘法”（Partial Least Square, PLS）来检验公司特征对期望收益的预测能力；Kozak, Nagel, and Santosh (2017)^[28]和Kelly, Pruitt, and Su (2018)^[29]分别运用PCA方法和IPCA方法（Instrumented PCA）提取因子中的共同因素，发现基于少数主成分的模型就可以独立地预测截面收益。Gu等(2018)^[30]检验了常见的机器学习算法在美国市场上的表现，研究发现机器学习模型可以有效地超越线性模型的预测效果。

在中国股票市场上将机器学习与多因子策略结合的研究则相对缺乏，潘莉和徐建国(2011a)^[31]检验了六个因子与回报率的关系，构建了适用于A股市场的因子模型。Hsu等(2018)^[32]检验了美国研究的常见异象在中国市场上的有效性，发现中美市场中因子的有效性存在着很大区别。现有研究更多集中在市场异象（或单因子）的发现、检验和解释，比如郑振龙和汤文玉(2011)^[33]检验了波动率风险因子；潘莉和徐建国(2011b)^[34]检验了惯性与反转因子；林虎等(2013)^[35]构建了中国市场换手率波动指标，除此之外，还有特质偏度^[36]、规模^[37]、个股横截面尾部风险^[38]、低价股异象^[39]、公司质量^[40]等。在国内多因子策略研究方面，Jiang等 (2019)^[41]分别采用Fama-MacBeth回归、PCA、PLS和FC方法整合A股市场中的75个异象因子，证明以上方法能够从因子中提取出有助于预测的信息，而更多的研究集中于对单一因子的检验和解释。

据此，这里提出本文第一项研究内容：将机器学习算法与多因子策略结合构建的股价预测模型能否通过有效识别数据间的非线性关系获得更好的预测效果？

(3) 特征筛选对股价预测的价值

特征选择作为数据挖掘的一项预处理步骤，被广泛应用于股价预测的相关研究中。Atsalakis和Valavanis(2009)^[42]总结了应用智能计算方法分析股市的文献，并指出许多文献中利用特定的技术筛选重要特征。特征选择的作用主要表现在三个方面：①去除冗余特征，提高模型预测性能；②减少时间成本，提高模型效率；③揭示数据的隐含特点。Huang和Tsai(2009)^[43]利用条件属性和决策属性的相关系数来度量特征的相对重要性，从而选取7项

技术指标用以预测台湾股票指数（FITX）。Lee(2009)^[44]提出了一种基于混合特征选择方法和SVM的预测模型来预测股票市场的走势,文中使用了29项技术指标,实证结果表明SVM在优于BPNN的基础上获得了最高87.3%的预测精确率。Tsai和Hsiao(2011)^[45]以85个财务指标和宏观指标作为原始特征,设计了一种基于PCA、GA和决策树的特征选择模型,并使用人工神经网络预测股价的季度波动,最终从85个原始特征中筛选出14项代表性特征。谢合亮和胡迪(2017)^[46]在研究打分法和普通最小二乘法确定因子权重的基础上,引入Lasso和弹性网络(ElasticNet)两类前沿方法在34个因子的大数据集上进行因子筛选并确定因子权重,利用沪深300指数成份股进行回测,证明ElasticNet方法比OLS和Lasso方法更能够筛选出有效因子。宋鹏(2017)^[47]等基于粗糙集与信息熵理论,提出了一种新的特征评价函数——区间序补集条件熵,并通过风险投资项目决策与股票选择决策两个案例,验证了该算法的有效性。

基于上述分析,这里提出本文的第二项研究内容:若机器学习算法的运用能够提升预测绩效,那么将特征筛选过程引入多因子选股过程中,能否通过去除冗余特征识别在中国A股市场具有最佳预测效率的因子集合?基于机器学习算法筛选出的代表性因子集合与单因子检验显著的因子集合以及传统OLS回归筛选出的因子集合间是否存在差异?

3. 创新点

本文的主要创新点如下:



- ① 因子集合的创新。本文利用中国A股市场1997年1月-2018年10月的交易数据和相关财务报表数据构建了包含96个异象因子的数据集,作为目前中国A股市场相关研究中最大的因子集合,基于更全面的数据集上的分析预测能够获取更有效的信息。
- ② 预测模型的创新。将基本面量化投资与机器学习算法结合构建预测模型的设计在中国股票市场的研究中相对缺乏,本文使用了过往研究中常见的各种机器学习算法构建多因子预测模型,是目前中国股票市场最全面的利用机器学习算法进行的异象因子研究,实证结果证明将机器学习算法应用于因子模型中能够显著提升模型的预测效果。

4. 行文结构

本文的行文结构安排如下:第二节对本文采用的各种机器学习预测算法的相关原理、特征筛选的主要流程和模型构建方式进行简要介绍,并列示了解决第一部分的研究问题设计的主要论证框架;第三节介绍了本文所使用因子集合的构建过程,对数据集进行初步的统计性描述;第四节实证部分展示了各种机器学习算法所构建模型的预测效果,在此基础上构建

投资组合，检验模型的有效性；同时采用多种算法对原始因子集合进行特征筛选，识别在中国A股市场具有最强预测效力的因子集合；最后在结论部分对全文进行概要性总结，提出下一步工作展望。

二、机器学习方法

1. 线性预测模型

(1) 线性回归

作为学习预测模型时首选的技术之一，线性回归（Linear Regression）使用最佳的拟合直线在因变量（Y）和一个或多个自变量（X）之间建立一种线性关系，具体可以表示为：

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \\ &= \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (X_0 = 1) \\ &= \boldsymbol{\beta}^T \times \mathbf{X} \end{aligned}$$

对于给定的样本 \mathbf{X}_i ，因变量真实值为 Y_i ，预测值为 $\hat{Y}_i = \boldsymbol{\beta}^T \times \mathbf{X}_i$ 。将函数的损失函数定义为平方损失函数，即：

$$Loss(\boldsymbol{\beta}) = \frac{1}{m} \sum_{i=0}^m (\hat{Y}_i - Y_i)^2$$

其中m为有效样本数量，通过将目标损失函数最小化，即 $\argmin \sum_{i=0}^m (\hat{Y}_i - Y_i)^2$ ，即可求解方程最优拟合系数 $\boldsymbol{\beta}$ 值。线性回归具有建模快速简单，便于直观的理解和解释等优点，本文以此算法作为基准模型，对比其与各类机器学习算法的预测性能差异。

就OLS特征筛选实现方式而言，本文通过逐一去除每项因子，重新运行OLS算法，计算去除各项因子后构建多空组合收益与全样本因子（96项）的OLS算法的多空组合收益进行对比，将两者之差作为单一因子重要性的度量方式，从而进行因子筛选。

(2) 预测组合模型

预测组合模型(Forecast Combination, FC)将自变量不同的预测模型进行加权平均，构成一个整体预测模型的方法。预测组合模型在现有研究中已经被广泛的应用，如Rapach等(2009)^[8]运用预测组合模型组合了各个因子的单变量回归模型，所根据模型预测所构成的投资组合绩效优于基于OLS模型所构建的投资组合。Zhang等(2019)^[48]运用预测组合模型，组合了八个用于预测原油价格波动率的HAR模型，提升了准确率，取得了优于稀疏化方法的结果。综合已有研究，组合预测模型的重点和难点在于加权系数的确定，目前常用的加权方式有算术平均法、最优权数法、方差倒数法等。具体到模型构建方面，本文的预测组合模型由96个单一市场异象因子分别做自变量的OLS模型构成，具体构建和预测方式如下：

① 在训练集上，分别训练96个市场异象因子所对应的单一变量最小二乘模型 $OLS_1, OLS_2, OLS_3, \dots, OLS_{96}$ ，训练模型采用的优化目标为预测值同真实值之间的均方误差。

② 在测试集上，运用上述96个市场异象因子模型分别进行收益率预测，取其预测均值（本文采用 $1/n$ 的简单算术平均法）作为最终预测结果进行输出。

作为预测组合模型基础的单变量最小二乘模型具有训练成本小，样本外预测结果不稳定的特点，而将其组合的预测组合模型以提升部分训练成本为代价，获得了模型多样性的提升，进而有利于样本外预测结果稳定性的提高。

(3) 岭回归

岭回归(ridge regression, Ridge)是一种改良的线性回归估计法，通过放弃线性回归的无偏性，以损失部分信息、降低精度为代价获得更为符合实际、更可靠的回归系数。此外，岭回归使得估计参数的波动范围变小，更加稳定。为解决线性回归内生的过拟合问题，岭回归在标准线性回归损失函数的基础上加入 L_2 范数正则化项，即：

$$Loss(\beta) = \frac{1}{m} \sum_{i=0}^m (\hat{Y}_i - Y_i)^2 + \gamma \|\beta\|_2^2 \quad (\gamma > 0)$$

与线性回归类似，通过最小化损失函数即可求解方程最优拟合系数 β 值。而同线性回归不同的 L_2 范数正则化项则提供了稀疏化参数的功能，降低不稳定自变量在回归方程中的权重，增强了模型的样本外预测能力。

(4) Lasso 回归

Lasso回归(Least absolute shrinkage and selection operator, Lasso)在损失函数中增加一个正则项 L_1 范数，即向量中各元素的绝对值之和作为正则化项，即：

$$Loss(\beta) = \frac{1}{m} \sum_{i=0}^m (\hat{Y}_i - Y_i)^2 + \gamma \|\beta\|_1 \quad (\gamma > 0)$$

相较于 L_2 范数而言， L_1 范数更易获得稀疏解，即 L_1 范数正则化求得的 β 会有更少的非零分量，因此Lasso也常被用于高维数据的特征筛选。在回归过程中，越重要的特征对应的系数绝对值越大，而与输出变量相关性越低的特征，系数就越接近于0。在金融领域，Feng, Giglio, and Xiu (2019)^[49]运用LASSO进行的因子筛选方法，成功选出了具有更高统计显著性的因子类别；Messmer and Audrino(2017)^[50]则在美国市场采用了LASSO算法的变种Adaptive LASSO从68个公司特征因子中筛选出了14个公司特征，并保有了不亚于68个公司特

征的解释能力。

(5) 弹性网络回归

弹性网络回归 (ElasticNet Regression, Elastic) 综合Lasso和岭回归两种算法, 同时使用了L1和L2正则化, 在两种算法间的权衡使得Elastic得以学习出一个只有少量非零参数的模型, 类似于Lasso算法, 同时它仍然保持了部分Ridge的正则性质, 具有较好的稳定性。已有研究中, Gu等(2018)^[30]采用了弹性网络回归模型进行了美国上市公司的收益预测, 并在其研究中证明了弹性回归网络算法具有较好的稳定性。具体而言, Elastic的损失函数可以表示为:

$$Loss(\beta) = \frac{1}{m} \sum_{i=0}^m (\hat{Y}_i - Y_i)^2 + \gamma_1 \|\beta\|_1 + \gamma_2 \|\beta\|_2^2$$

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的情况下的群体效应。因此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的情形下能够保持较好的预测效果。

(6) 偏最小二乘回归

偏最小二乘回归 (Partial Least Squares Regression, PLS) 方法在普通多元回归的基础上结合了主成分分析 (Principal components analysis, PCA) 和典型相关分析 (Canonical Correlation Analysis, CCA) 的思想, 以解决回归分析中自变量多重共线性的问题。近年来的相关研究中, PLS算法以其独特的优势获得了很高的采用率, 如Light等(2017)^[27]选用PLS来检验公司特征对期望收益的预测能力。在运用PLS进行收益预测角度则有Gu等(2018)^[30]进行的针对美国所有上市公司的研究, 研究证明了PLS算法在美国市场股票收益预测领域要好于传统的OLS算法。

考虑存在 m 个自变量 x_1, x_2, \dots, x_m 。偏最小二乘回归的基本作法是首先在自变量集合中提出第一成分 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, 如果回归方程已达到满意的精度, 则算法中止。否则继续第二对成分的提取, 直到能达到满意的精度为止。若最终对自变量集提取 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, 最终再表示为因变量与原自变量的

回归方程式，即偏最小二乘回归方程式。

2. 非线性预测模型

(1) 梯度提升树

梯度提升树(Gradient Boosting Decision Tree, GBDT) 是一种迭代的决策树算法, 该算法由多棵决策树组成, 所有树的结论累加起来做最终预测结果。已有研究中, Krauss(2017)^[51]等人运用梯度提升树算法进行标准普尔500指数成分股的运动方向预测, 并根据预测结果进行投资组合构建, 投资组合绩效结果明显好于市场投资组合。该算法的核心在于每一棵树学习之前所有树结论和的残差, 而为了消除残差, 模型在残差减少的梯度(Gradient)方向上建立一个新的模型。因此在GBDT中, 每个新的模型的建立是为了使得之前模型的残差梯度方向减少。此外, 决策树进行分支时以最小化平方误差为标准, 对每一个特征, 每一个阈值进行穷举以寻求最优的分割点。对于训练集 $\text{Train} = \{(x_1, y_1), (x_1, y_1), \dots, (x_N, y_N)\}$ 而言, 构建梯度提升树 $f_M(x)$ 的简要实现流程如下:

- ① 初始化 $f_0(x) = 0$
- ② 对于 $m = 1, 2, \dots, M$, 计算其预测残差:

$$r_{mi} = y_i - f_{m-1}(x_i) \quad i = 1, 2, \dots, N$$

- ③ 对于 N 个残差学习得到一个回归树 $T(x; \phi_m)$
- ④ 更新 $f_m(x) = f_{m-1}(x) + \text{Train}(x; \phi_m)$
- ⑤ 最终得到梯度提升树

$$f_M(x) = \sum_{m=1}^M T(x; \phi_m)$$

梯度提升树能够灵活处理包括连续值和离散值在内的各种类型的数据, 在较少的调参时间情况下, 获得相对较高的预测的准确率。同时由于梯度提升树使用的损失函数相对稳健, 算法对异常值的鲁棒性非常强。

梯度提升树也是常用于进行特征选择的算法之一, 具体而言, 某一特征 j 的重要程度由特征 j 在单棵树中重要度的平均值度量, 计算公式如下:

$$\hat{J}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{J}_j^2(T_m)$$

其中, M 是集成算法中决策树的数量。特征 j 在单棵树的重要度 $\hat{J}_j^2(T_m)$ 则通过计算按特征 j 分裂之后损失损失函数的减少值来度量。

(2) 极端梯度提升树

Boosting算法以集成弱分类器的方式实现了提高预测稳定性和预测效力的目的,成为了机器学习领域中的突出算法。而作为其中的代表性算法,由Chen等(2016)^[52]提出的极端梯度提升树(Extreme Gradient Boosting, Xgboost)也具有一般Boosting算法的特点,但其较于一般Boosting而言具有训练成本低,训练结果更为精确的优势,而究其根本,优势主要来源于作为基学习器的CART树算法。CART树作为二叉树中较为强大的一种算法,使得训练后的基学习器的集合具有更高多样性从而提升Boosting之后的预测效果。同时,CART二叉树训练简单,能够降低由于Boosting算法集成而提升的训练成本。具体模型设置如下:

- ① 基于训练集数据构建第一棵CART回归树,并计算出模型的残差
- ② 通过第一步计算出的残差训练下一棵CART回归树,再次进行残差计算。
- ③ 重复②步直到最大迭代次数。

单就上述步骤而言,XGBoost算法同GBDT具有很高的相似性,都是以前一次预测的残差作为下一步训练目标,但两者在具体实现上存在以下不同:

- ① XGBoost在拟合目标的设置上在GBDT算法基础上加入了正则化项,使模型具有更好的泛化能力。具体拟合目标函数如下,其中T为CART树的叶节点, γ 为正则化参数,w为权重系数。

$$L(\phi) = \sum_i l(\hat{y}_i - y_i) + \sum_k \Omega(f_k)$$
$$\Omega(f_k) = \gamma T + \frac{1}{2} ||w||^2$$

- ② 在计算残差过程中,XGBoost在GBDT的基础上加入了二阶导数,提升了残差估计的准确性。
- ③ 具体到每颗CART回归树的而言,在叶节点划分时,GBDT算法采用的是最小化均方差,而XGBoost算法则是最大化上述方程中的正则化项。
- ④ 此外,XGBoost还在训练内存处理上做了优化,降低了训练消耗。

正是由于细节上的不同处理使得XGBoost独立于GBDT算法之外,在机器学习算法中占有一席之地,取得了不俗的成果。

(3) 集成神经网络

神经网络集成模型(Ensemble Artificial Neural Network, EN-ANN)是一种基于人工神经网络(Artificial Neural Network, ANN)的集成学习算法,主要针对由于单一ANN模

型由于初始化问题而产生的预测结果不稳定现象。已有研究如 Gu 等人(2018)^[30]也采用了神经网络集成算法进行神经网络稳定性的优化,并证明神经网络集成算法能够很好地降低初始化多样性而导致的训练器陷入局部极小值的概率,提升预测稳定性。神经网络集成模型的核心是构建多个不同初始化状态下的简单神经网络,通过这种方式使得模型集成更多的可能性,以提升预测结果的稳定性。在训练过程中神经网络集成模型对每一个 ANN 都进行单独的训练和优化实现每一个模型的近似最优化。基于 Gu 等人(2018)^[30]的研究,本文所采用的神经网络集成模型主要实现流程如下:

- ①初始化每一个神经网络 $NN_1, NN_2, NN_3, \dots, NN_K$ 。
- ②在训练集中,对每一个神经网络模型以最小化均方误差作为优化目标进行参数拟合。
- ③选取训练集上均方误差前 50%的神经网络作为预测池构建神经网络集成模型,在预测时取预测池的平均结果作为最终结论。

神经网络集成模型在有效提升模型稳定性的同时能对由于初始化数值选取存在偏差而陷入局部最优解的训练器进行甄别,筛选出相对稳定的训练器进行预测,在很大程度上提升了预测的鲁棒性。

(4) 深度前馈网络

深度前馈网络模型(Deep Feedforward Network, DFN),又名多层感知机,是典型的深度学习算法之一。该模型的输入特征和输出变量间没有反馈链接,整体保持由上至下的纵向层级结构,每一层又由多个激活函数构成的神经元组成。作为深度学习的代表,DFN 具有很强的非线性函数的拟合能力,能够有效捕捉变量间的非线性关系。已有研究中,Krauss 等(2017)^[51]运用由 DFN 构建的深度神经网络进行了美国标普 500 指数成分股的涨跌预测,并根据预测构建了投资组合,取得了远超过市场投资组合的收益率和夏普率。基于 Krauss 的研究,本文深度前馈网络模型构建简要流程如下:

- ①确定前馈网络的层数和神经元数目以及模型重复训练次数。
- ②随机初始化每一个神经元中的参数。
- ③以前馈网络输出结果同真实值之间的均方误差作为模型损失函数,在训练数据集上计算模型损失值。
- ④将模型损失值导入训练器进行前馈网络参数的重计算,重复③步直到达到设定重复训练次数。

深度前馈模型对于非线性的金融时间序列数据来说能够很好地捕捉到其中错综复杂的

关系，进行模拟和预测，实现较好的预测效果。同时，多次训练能够降低初始化问题所带来的不稳定性，使预测结果保持较好的稳定性。

类似于 OLS 算法的因子筛选流程，本文在得出全样本因子 DFN 模型所构建多空组合月度收益率后，逐项去除单一因子进行 DFN 模型训练。在此过程中，DFN 模型参数保持不变，输入变量为剩余 95 个因子，计算基于此预测模型构建多空组合月度收益同全样本因子构建多空组合月度收益率之差，进行因子重要性排序，筛选有效因子集合。

3. 预测模型构建

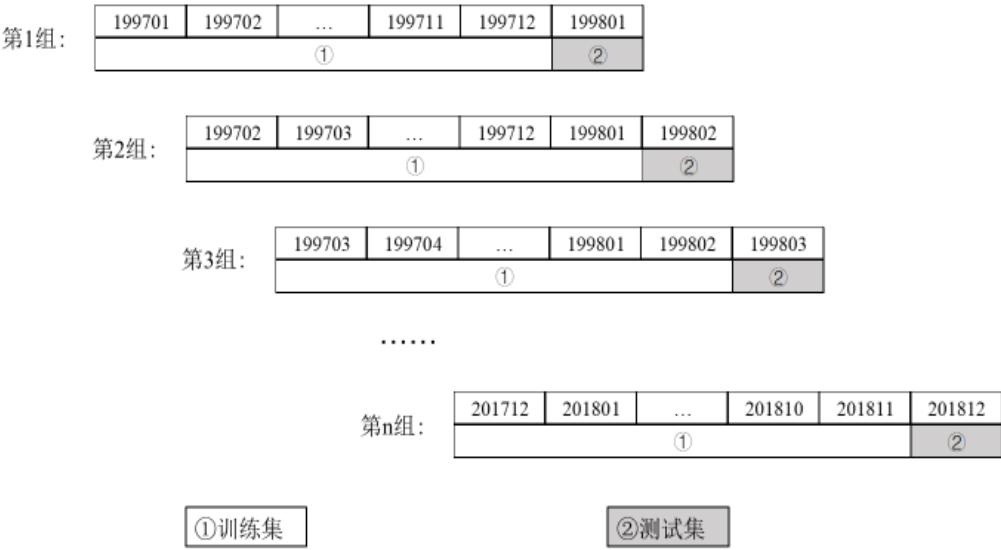


图2.1 训练测试集划分方式

资料来源：作者绘制

借鉴DeMiguel et al. (2009)^[53]，本文采用滑动窗口法划分训练、测试集合以构建预测模型，实现流程如图2.1所示，训练集的时间区间长度为12个月（为验证预测模型稳定性，文中同时检验了训练集长度分别为24个月、36个月时的预测结果），对应测试集时间长度为1个月，每次滚动完成时间窗口后移1个月，以此类推。例如，第1组训练集的时间窗口为1996年1月-1996年12月，测试集时间窗口则为1997年1月，第2组训练集的时间窗口则为1996年2月-1997年1月，测试集时间窗口则为1997年2月。相较于传统的划分方式（如留出法、交叉验证等）而言，滑动窗口划分方式在最大程度上保留了原始数据集合的时间序列特征，与现实投资过程保持一致，也在很大程度上提高了数据的利用效率。具体而言，在使用12个月的滑动窗口的划分情形下，最终得到250个训练、测试集合（即 $n = 250$ ），在每个滑动窗口内，根据训练集数据构建预测模型，用于预测测试集内股票收益率，通过做多收益率排序前10%的股票并做空收益率排序后10%的股票构建多空组合，月度换仓，最终根据投资组合

投资绩效（风险、收益）检验预测模型的有效性，具体实现流程如图2.2所示，文中所有机器学习算法均基于python的sklearn算法包和mxnet算法包实现。

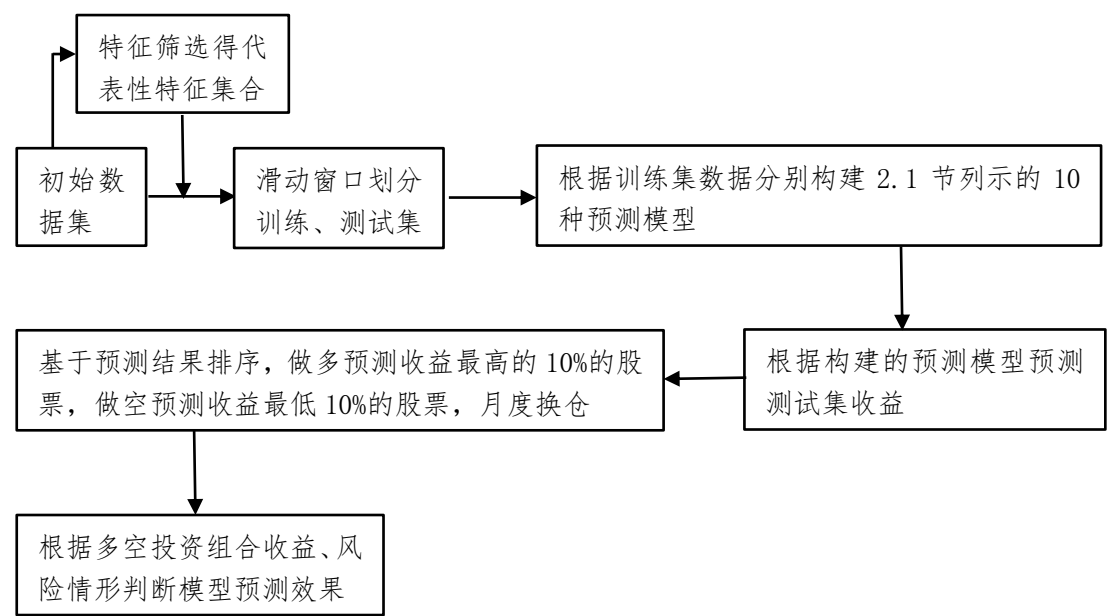


图2.2 预测模型构建流程

资料来源：作者绘制

4. 绩效衡量指标

文章实证部分分别利用第二节列出的10种算法构建预测模型，并根据模型预测结果构建多空组合，即做多模型预测收益最高的10%的股票集合，做空模型预测收益最低的10%的股票集合，投资组合月度换仓。对于某一特定算法而言，依据其预测结果构建多空组合的月度收益率和风险调节后收益（分别为Fama-French三因子调节后月度收益率、Fama-French五因子调节后月度收益率、夏普比率）将作为算法有效性的绩效衡量指标。

由于中国市场做空机制不够灵活，本文亦列出多头和空头的组合，其中多头等权重持有预测收益前10%的股票，空头等权重做空预测收益后10%的股票。

三、数据描述

借鉴Green, Hand and Zhang (2017)^[3]，本文选取了96个公司特征代理异象因子，并按照因子属性分为交易摩擦因子、动量因子、价值因子、成长因子、盈利因子、财务流动因子共六大类。由于财报数据大部分为季度公布，本文对季度数据进行了月度填充。由于上市公司财报披露时间存在延迟²，填充数据的基本原则是仅在规定的报表可用之后再进行填充。按月填充方式的示意图如图3.1所示。数据均来自CSMAR数据库。附录1展示了因子的描述性统计³。

图3.1 财务数据填充方式示例

| | | | | | | | | | | | |
|-------------|----|----|----|------------|----|----|----|-----------|-----|-----------|-----|
| 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
| t-1年三季度报表填充 | | | | t-1年年度报表填充 | | | | t年半年度报表填充 | | t年三季度报表填充 | |

本文选取的输出变量为股票月度收益率（考虑现金股利再投资），由于本文构建模型的最终目的在于预测股票下月收益，因此对于某一特定月份的输出变量而言，其对应的输入变量则为前一月份数据，例如，用于预测1997年1月股票收益率的输入特征为1996年12月的各项因子值。将t-1期的公司特征和 t 期的月度收益率进行配对后即可获得“公司-月”数据。

研究区间设定为1997年1月至2018年10月，数据频率为月度频率。自1996年12月16日起，深交所、上交所对上市的股票、基金的交易实行10%以内的涨跌幅限制，为避免这一重大交易机制的变化对研究结果的影响，我们选择从1997年1月开始。选择月度频率主要是同现有研究一致^[30, 44]。

初始股票池为 A 股市场的所有股票。由于市场中ST的股票业绩亏损，存在退市风险，为避免对股价产生额外的影响，我们从全部A股股票池内剔除ST股票；金融行业股票部分指标计量方式有别于其他上市公司，为避免异常扰动，也将金融类股票从股票池内剔除；最后，由于IPO抑价效应的存在，股票上市初年股价可能存在异常波动，本研究也剔除了股票上市第一年的数据。同时，数据中存在着一定比例的缺失值，其处理过程分为两步：①若某只股票在第t月收益数据存在缺失（通常由股票连续停牌造成，共包含3730条缺失值），则剔除该股票在月份t上的所有数据；②若某只股票的因子值存在缺失，则以0填充。

在剔除ST股票、金融股、股票上市首年数据并去除缺失值后，1997年1月-2018年10月的

² 中国证监会《上市公司信息披露管理办法》规定，上市公司年报的披露时间为每个会计年度结束之日起4个月内；季报的披露时间为每个会计年度第3个月、第9个月结束后的1个月内编制完成并披露。

³ 各个因子的构建方式以及单因子有效性检验详见《中国A股市场量化因子检验报告》。

有效样本为381062条。图3.2中展示了1997年1月-2018年10月内月度有效样本量。随着月度样本量总体随年份呈现上升态势，由1997年1月的307条有效样本增至2018年10月的3211条有效样本。

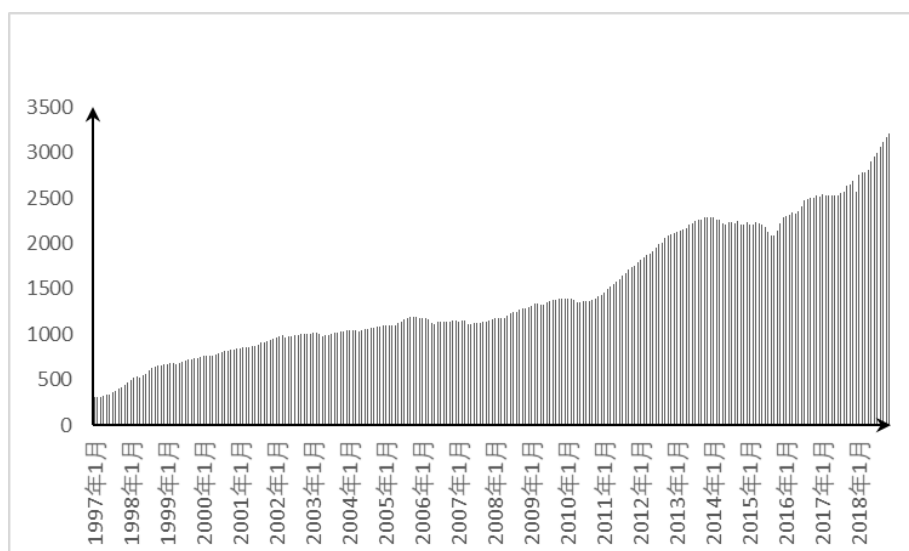


图3.2 月度有效样本量

资料来源：作者绘制

附录1的描述性统计显示，不同因子取值在数量级以及分布上均存在显著差异，为避免由于以上因素引起的预测偏差（例如数量级的差异会导致：（1）量级较大的特征在预测时占据主导地位；（2）数量级的差异会引起部分机器学习算法迭代收敛速度减慢），本文将训练集中的原始特征进行标准化，使之符合均值为0，方差为1的正态分布：

$$X_{scale} = \frac{X - Mean}{Standard\ Deviation}$$

四、实证结果

1. 机器学习算法全样本预测结果对比

(1) 12 个月滑动窗口预测结果

在实证部分，本文首先检验了以全部96项异象因子作为输入的各种模型的预测效果。表4.1展示了9种机器学习方法在12个月滑动窗口时多空组合、多头组合和空头组合的风险收益情形。作为对比，表中列出了基于线性模型预测的组合收益(OLS)和市场组合的收益(MKT，仅多头)。

结果显示，九种基于机器学习的基本面量化投资模型均能显著优于单因子模型的收益（单因子模型在12个月滑动窗口内等权重多空组合最高月度收益率仅为1.77%，对应夏普比率为0.72，由size因子取得，详见《中国A股市场量化因子检验报告》），表明机器学习增强的组合预测模式能够提升模型的投资效果。在表中列示的各种多因子模型中，线性机器学习算法（Ridge, Lasso, ElasticNet, PLS）均能够获得较基准OLS回归更好的预测结果，表明机器学习算法能够更好的识别因子间的线性关系从而提升预测绩效。同样以OLS回归为基准的FC方法，尽管所构建多空组合表现出更高的年化收益，但组合波动率相较于OLS则存在更大幅度增加，从而导致组合夏普比率（SR）较OLS方法存在小幅下降。在各种机器学习算法中，极端梯度提升算法(Xgboost)，梯度提升树(GBDT)和深度前馈网络(DFN)三种非线性算法带来的投资绩效提升最为显著，多空组合月度收益率较传统OLS回归分别能够获得0.74%，0.76%和1.06%的上升，夏普比率则分别获得0.50，0.51和0.66的提升。非线性机器学习算法整体而言表现出比线性机器学习算法（Lasso, Ridge, ElasticNet）更好的预测效果，证明因子集合间非线性关系的存在。最后，DFN算法作为深度学习的代表之一，在多因子预测中能够获得明显超过线性模型和传统机器学习算法的预测效果，表明深度学习算法能够通过更加有效地识别特征间的非线性关系，获得更好的预测效果。

由于中国股票市场缺乏有效的做空机制，单独做多预测收益前10%的股票的投资组合(long)的收益也应给予关注。表中结果显示，根据Xgboost，GBDT和DFN三种预测模型构建的单独做多投资组合能够获得相对更高的投资绩效，其中表现最好的是基于DFN构建的做多组合，较OLS而言可获得0.63%的月度收益率提升。表4.1中最后一列列示了同期上证综指的收益情形。结果显示，上证指数的平均月度收益仅为0.61%，且不显著（NW-t=1.18），各项多因子策略的做多组合均能获得明显超过市场指数的平均收益。

表4.1 基于机器学习的量化投资模型在中国市场的投资绩效。结果基于全样本12个月滑动窗口预测，样本区间为1997年1月至2018年10月。

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN | Mkt |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.31%*** (3.2094) | 2.69%*** (3.5623) | 2.41%*** (3.343) | 2.46%*** (3.3844) | 2.45%*** (3.376) | 2.49%*** (3.3978) | 2.52%*** (3.4955) | 2.69%*** (3.6471) | 2.60%*** (3.5863) | 2.94%*** (3.9346) | 0.61% (1.1757) |
| | FF3-alpha | 0.87%*** (3.6282) | 1.16%*** (4.4375) | 1.02%*** (4.2265) | 1.02%*** (4.1944) | 1.02%*** (4.1797) | 1.09%*** (4.4981) | 1.12%*** (4.9931) | 1.22%*** (5.3586) | 1.17%*** (5.1428) | 1.51%*** (5.8356) | -0.14%*** (-2.2523) |
| | FF5-alpha | 0.77%*** (3.6634) | 1.01%*** (4.5589) | 0.89%*** (4.1272) | 0.88%*** (4.1485) | 0.88%*** (4.1315) | 1.02%*** (4.7332) | 0.98%*** (4.8109) | 1.02%*** (5.2232) | 0.97%*** (4.9657) | 1.33%*** (6.2672) | -0.17%*** (-2.7593) |
| | Sharpe Ratio | 0.6997 | 0.8003 | 0.7329 | 0.7440 | 0.7424 | 0.7596 | 0.7751 | 0.8195 | 0.8000 | 0.8868 | 0.1795 |
| | 均值 | 0.41% (0.6272) | 0.24% (0.3786) | 0.34% (0.5246) | 0.30% (0.4669) | 0.31% (0.4733) | 0.12% (0.1808) | 0.26% (0.3972) | 0.05% (0.07) | -0.06% (-0.0961) | -0.02% (-0.0288) | |
| | FF3-alpha | -0.86%*** (-5.0114) | -0.84%*** (-4.1192) | -0.96%*** (-5.6995) | -1.01%*** (-5.8706) | -1.01%*** (-5.8795) | -1.13%*** (-6.2351) | -1.10%*** (-6.6674) | -1.33%*** (-7.1608) | -1.44%*** (-7.8408) | -1.38%*** (-7.6385) | |
| 空头组合 | FF5-alpha | -0.86%*** (-4.4415) | -0.82%*** (-3.9832) | -0.97%*** (-5.3131) | -1.03%*** (-5.6106) | -1.04%*** (-5.6515) | -1.17%*** (-6.1773) | -1.06%*** (-6.0904) | -1.25%*** (-6.1607) | -1.36%*** (-6.9417) | -1.31%*** (-6.6228) | |
| | Sharpe Ratio | 0.0703 | 0.0116 | 0.0453 | 0.0322 | 0.0337 | -0.0331 | 0.0173 | -0.0584 | -0.0974 | -0.0807 | |
| | 均值 | 1.90%*** (6.1017) | 2.45%*** (6.2946) | 2.07%*** (6.798) | 2.16%*** (6.9121) | 2.14%*** (6.9087) | 2.37%*** (7.2764) | 2.26%*** (7.7333) | 2.64%*** (8.6532) | 2.66%*** (8.6377) | 2.96%*** (8.7849) | |
| 多空组合 | FF3-alpha | 1.52%*** (5.3428) | 1.79%*** (5.1908) | 1.77%*** (6.1094) | 1.82%*** (6.3801) | 1.82%*** (6.3788) | 2.01%*** (6.956) | 2.01%*** (7.163) | 2.34%*** (7.8103) | 2.40%*** (7.8068) | 2.68%*** (8.0099) | |
| | FF5-alpha | 1.42%*** (5.2816) | 1.62%*** (5.0471) | 1.65%*** (6.1551) | 1.71%*** (6.531) | 1.71%*** (6.5438) | 1.98%*** (7.2691) | 1.83%*** (6.9757) | 2.06%*** (7.3683) | 2.12%*** (7.5359) | 2.43%*** (8.4045) | |
| | Sharpe Ratio | 1.4048 | 1.3817 | 1.5363 | 1.5946 | 1.5896 | 1.6581 | 1.7565 | 1.9032 | 1.9184 | 2.0618 | |

资料来源：作者计算

为证明机器学习算法相对于传统OLS回归模型的优越性，表4.2中展示了各种机器学习算法与OLS回归所构建投资组合收益率差异显著性的t检验结果，以及作为深度学习算法代表之一的DFN模型相对于其它传统机器学习算法的差异检验结果。数据显示，几乎所有机器学习算法都较OLS回归投资绩效存在明显提升（除EN-ANN外，EN-ANN与OLS多空组合差异检验t值仅为1.67），证明机器学习算法能够通过识别数据间的非线性特征有效提升预测效果。对DFN与其它模型预测效果的t检验结果显示，DFN能够显著超越Ridge、Lasso、ElasticNet和EN-ANN算法获得更高的投资绩效，证明了深度学习算法的优越性。

表4.2 算法预测构建投资组合收益差异显著性NW-t检验结果（12个月）

| OLS 与其它算法 | | | | DFN 与其它算法 | | | |
|-------------------|------------|--------|---------|-------------------|------------|--------|---------|
| | long-short | long | short | | long-short | long | short |
| FC | 2.0135 | 2.4916 | -1.0653 | FC | 1.6032 | 1.3892 | -1.4943 |
| Ridge | 2.1012 | 1.9880 | -1.7241 | Ridge | 3.6913 | 3.3440 | -2.6946 |
| Lasso | 2.4452 | 2.3466 | -1.5678 | Lasso | 3.2367 | 2.9994 | -2.3356 |
| ElasticNet | 2.3655 | 2.2769 | -1.5076 | ElasticNet | 3.2859 | 3.0435 | -2.3721 |
| PLS | 2.2178 | 1.4004 | -2.3328 | PLS | 2.2630 | 2.7979 | -0.9425 |
| EN-ANN | 1.6700 | 1.6115 | -1.2278 | EN-ANN | 4.6508 | 3.9087 | -3.3212 |
| Xgboost | 3.3697 | 2.6512 | -2.8198 | Xgboost | 1.5779 | 1.8093 | -0.6301 |
| GBDT | 3.3524 | 1.9948 | -3.7296 | GBDT | 1.5467 | 2.4639 | 0.4253 |
| DFN | 4.2040 | 3.6794 | -3.3495 | | | | |

资料来源：作者计算

(2) 3、24、36 个月滑动窗口预测结果

除采用12个月滑动窗口外，表4.4、4.5和4.6分别展示了基于3个月、24个月和36个月滑动窗口划分构建的各种模型的预测绩效，结果显示在任一滑动窗口条件下，DFN、GBDT和Xgboost三种非线性算法均优于其他线性机器学习算法和OLS的预测效果（包括多空组合和单独做多的投资组合），且深度学习算法DFN始终保持最高的投资绩效，验证了在12个月滑动窗口内所得结果的一般性。而就线性机器学习算法而言，其预测稳健性则相对较弱，如当采用36个月的滑动窗口时，FC和PLS两种算法的预测绩效表现出弱于OLS的情形，但其余机器学习算法的预测绩效基本能够保持稳定。

此外，表4.3中分别列示了在3个月、24个月和36个月滑动窗口内对OLS和其它9种算法预测结果的差异显著性检验以及DFN和其它机器学习算法的差异显著性检验结果。与上述结论相似，在各种滑动窗口内，DFN、GBDT和Xgboost三种非线性算法均表现出显著优于线性OLS回归的投资绩效（ $NW - t > 1.96$ ），同时深度学习算法DFN能够获得显著超过线性机器学习算法的预测结果。此外，在36个月滑动窗口内，与FC和PLS两种算法的预测效果相对于OLS

预测差异的t检验结果表明在较长的时间窗口内，以上两种线性算法无法有效识别因子间的复杂关系从而进行有效预测（尽管t值不显著，但数值为负数，分别为-0.85和-0.22）。

从各种滑动窗口的对比结果来看，当采用较短的训练区间（3个月滑动窗口）时的预测绩效相对较弱，而12个月、24个月和36个月的滑动窗口内的预测绩效不存在明显差异，因此下文后续检验过程中，如无特殊说明均采用12个月滑动窗口划分模式。

表4.3 算法预测构建投资组合收益差异显著性NW-t检验结果（3/24/36个月）

| OLS 与其它算法 | | | | DFN 与其它算法 | | | |
|-------------------|------------|--------|---------|-------------------|------------|--------|---------|
| | long-short | long | short | | long-short | long | short |
| 3 个月滑动窗口 | | | | | | | |
| FC | -0.4098 | 1.4039 | 2.1085 | FC | 4.2414 | 3.0284 | -4.5459 |
| Ridge | 3.0225 | 2.8706 | -1.9750 | Ridge | 4.6611 | 4.4725 | -3.0178 |
| Lasso | 2.1812 | 2.2567 | -1.2718 | Lasso | 4.5903 | 4.4683 | -3.0916 |
| ElasticNet | 2.1776 | 2.3342 | -1.1593 | ElasticNet | 4.6132 | 4.4051 | -3.2031 |
| PLS | 0.0752 | 1.1911 | 1.1516 | PLS | 5.3035 | 4.3500 | -4.5838 |
| EN-ANN | -0.3477 | 0.5203 | 1.1650 | EN-ANN | 6.2891 | 4.9401 | -5.0030 |
| Xgboost | 2.2948 | 2.2630 | -1.6575 | Xgboost | 2.5448 | 2.6833 | -1.5069 |
| GBDT | 1.7868 | 1.4313 | -1.6415 | GBDT | 3.2630 | 3.8003 | -1.6385 |
| DFN | 6.0549 | 5.4667 | -4.1950 | | | | |
| 24 个月滑动窗口 | | | | | | | |
| FC | 0.6037 | 1.3504 | 0.2240 | FC | 2.5896 | 1.7603 | -2.7793 |
| Ridge | 1.6518 | 0.9783 | -1.5945 | Ridge | 3.9976 | 3.4038 | -3.5053 |
| Lasso | 2.5097 | 2.4703 | -1.3692 | Lasso | 3.6813 | 3.0490 | -3.3845 |
| ElasticNet | 1.9132 | 1.6003 | -1.4305 | ElasticNet | 3.5470 | 2.9223 | -3.2439 |
| PLS | 0.6786 | 1.3431 | 0.0758 | PLS | 2.5176 | 1.7720 | -2.6565 |
| EN-ANN | 1.6548 | 1.4674 | -1.3811 | EN-ANN | 3.7363 | 3.0827 | -3.1868 |
| Xgboost | 2.6177 | 2.0237 | -2.3456 | Xgboost | 2.1577 | 2.0139 | -1.3767 |
| GBDT | 3.2361 | 2.2041 | -3.4011 | GBDT | 1.1725 | 1.7834 | -0.0519 |
| DFN | 4.1480 | 3.4756 | -3.6437 | | | | |
| 36 个月滑动窗口 | | | | | | | |
| FC | -0.8528 | 0.2013 | 1.5888 | FC | 3.2854 | 2.9409 | -2.9362 |
| Ridge | 1.3319 | 1.1169 | -0.7083 | Ridge | 3.7155 | 3.9244 | -2.4119 |
| Lasso | 2.3667 | 2.8630 | -0.8008 | Lasso | 3.2373 | 3.2609 | -2.2530 |
| ElasticNet | 2.2708 | 2.6915 | -0.8369 | ElasticNet | 3.2633 | 3.3157 | -2.2421 |
| PLS | -0.2180 | 0.5405 | 0.6966 | PLS | 3.2992 | 3.4849 | -2.4718 |
| EN-ANN | 1.5324 | 2.0511 | -0.5389 | EN-ANN | 3.4754 | 3.0116 | -3.0029 |
| Xgboost | 2.7422 | 2.6905 | -1.9882 | Xgboost | 1.4293 | 1.8273 | -0.6294 |
| GBDT | 3.3739 | 2.8821 | -2.9860 | GBDT | 0.7294 | 1.7140 | 0.3431 |
| DFN | 3.8316 | 4.0483 | -2.4533 | | | | |

资料来源：作者计算

表4.4 全样本3个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN | Mkt |
|------|--------------|------------------------|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------|
| 多头组合 | 均值 | 2.04%*** (2.889) | 2.26%*** (3.1345) | 2.25%*** (3.1822) | 2.24%*** (3.1378) | 2.24%*** (3.1502) | 2.18%*** (3.0031) | 2.11%*** (3.0905) | 2.37%*** (3.259) | 2.25%*** (3.1578) | 2.71%*** (3.7392) | 0.59% (1.1588) |
| | FF3-alpha | 0.71%*** (2.9271) | 0.96%*** (3.3295) | 0.97%*** (3.8338) | 0.96%*** (3.7565) | 0.97%*** (3.7966) | 0.90%*** (3.4285) | 0.87%*** (3.9683) | 1.03%*** (4.2147) | 0.95%*** (3.7821) | 1.34%*** (5.3141) | -1.54%* (-1.8898) |
| | FF5-alpha | 0.67%*** (3.2633) | 0.77%*** (3.0791) | 0.90%*** (4.0341) | 0.88%*** (3.9395) | 0.89%*** (3.9755) | 0.81%*** (3.5212) | 0.76%*** (4.0083) | 0.87%*** (4.2175) | 0.78%*** (3.6155) | 1.21%*** (5.6751) | -1.64%** (-1.9891) |
| | Sharpe Ratio | 0.5936 | 0.6663 | 0.6681 | 0.6588 | 0.6619 | 0.6373 | 0.6471 | 0.6990 | 0.6746 | 0.8059 | 0.1630 |
| | 均值 | 0.45% (0.7066) | 0.78% (1.2037) | 0.34% (0.5318) | 0.36% (0.5687) | 0.37% (0.5816) | 0.58% (0.9067) | 0.59% (0.9527) | 0.24% (0.3824) | 0.25% (0.3966) | 0.07% (0.1201) | |
| | FF3-alpha | -0.81%*** (-4.6542) | -0.48%** (-2.1475) | -0.93%*** (-5.4958) | -0.92%*** (-5.0000) | -0.91%*** (-4.9294) | -0.69%*** (-3.5383) | -0.68%*** (-4.1094) | -1.10%*** (-5.6407) | -1.07%*** (-5.6705) | -1.18%*** (-7.0133) | |
| 空头组合 | FF5-alpha | -0.82%*** (-4.2537) | -0.31% (-1.3247) | -0.95%*** (-5.3452) | -0.91%*** (-4.8714) | -0.90%*** (-4.797) | -0.66%*** (-3.4475) | -0.64%*** (-3.6628) | -0.95%*** (-4.8125) | -0.95%*** (-4.9777) | -1.12%*** (-6.3755) | |
| | Sharpe Ratio | 0.0823 | 0.1971 | 0.0410 | 0.0485 | 0.0516 | 0.1271 | 0.1345 | 0.0065 | 0.0084 | -0.0546 | |
| 多空组合 | 均值 | 1.59%*** (5.2377) | 1.48%*** (3.4048) | 1.91%*** (6.0117) | 1.88%*** (5.5954) | 1.87%*** (5.5932) | 1.60%*** (4.4175) | 1.52%*** (5.2374) | 2.13%*** (6.2141) | 2.00%*** (5.9777) | 2.64%*** (7.854) | |
| | FF3-alpha | 1.29%*** (4.3652) | 1.22%*** (2.9538) | 1.68%*** (5.5226) | 1.65%*** (5.2202) | 1.65%*** (5.2173) | 1.36%*** (4.0178) | 1.33%*** (4.7673) | 1.90%*** (5.6755) | 1.79%*** (5.2965) | 2.30%*** (7.3851) | |
| | FF5-alpha | 1.27%*** (4.6084) | 0.86%*** (2.2525) | 1.63%*** (5.8186) | 1.57%*** (5.4149) | 1.56%*** (5.4013) | 1.25%*** (4.0687) | 1.18%*** (4.6127) | 1.60%*** (5.2517) | 1.51%*** (4.8157) | 2.11%*** (7.7722) | |
| | Sharpe Ratio | 1.0386 | 0.6567 | 1.2655 | 1.1693 | 1.1715 | 0.9176 | 1.0599 | 1.2695 | 1.2170 | 1.7385 | |
| | 均值 | 1.59%*** (5.2377) | 1.48%*** (3.4048) | 1.91%*** (6.0117) | 1.88%*** (5.5954) | 1.87%*** (5.5932) | 1.60%*** (4.4175) | 1.52%*** (5.2374) | 2.13%*** (6.2141) | 2.00%*** (5.9777) | 2.64%*** (7.854) | |
| | FF3-alpha | 1.29%*** (4.3652) | 1.22%*** (2.9538) | 1.68%*** (5.5226) | 1.65%*** (5.2202) | 1.65%*** (5.2173) | 1.36%*** (4.0178) | 1.33%*** (4.7673) | 1.90%*** (5.6755) | 1.79%*** (5.2965) | 2.30%*** (7.3851) | |

资料来源：作者计算

表4.5 全样本24个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN | Mkt |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.32%*** (3.0913) | 2.53%*** (3.2908) | 2.34%*** (3.1062) | 2.40%*** (3.196) | 2.41%*** (3.2209) | 2.53%*** (3.2888) | 2.50%*** (3.2956) | 2.61%*** (3.2922) | 2.65%*** (3.4008) | 2.87%*** (3.7756) | 0.65% (1.2029) |
| | FF3-alpha | 0.91%*** (3.8374) | 1.11%*** (4.201) | 0.92%*** (3.8844) | 0.98%*** (4.157) | 0.99%*** (4.0641) | 1.11%*** (4.194) | 1.14%*** (4.8548) | 1.16%*** (4.5531) | 1.22%*** (5.1116) | 1.48%*** (6.0143) | -1.81%*** (-2.4459) |
| | FF5-alpha | 0.75%*** (3.6288) | 0.89%*** (4.1611) | 0.76%*** (3.6945) | 0.82%*** (4.0493) | 0.82%*** (3.8764) | 0.89%*** (4.1529) | 0.98%*** (4.8021) | 0.94%*** (4.4164) | 0.99%*** (4.8845) | 1.25%*** (6.1929) | -2.17%*** (-2.8708) |
| | Sharpe Ratio | 0.6951 | 0.7489 | 0.6987 | 0.7214 | 0.7272 | 0.7482 | 0.7488 | 0.7641 | 0.7863 | 0.8690 | 0.1948 |
| | 均值 | 0.55% (0.8162) | 0.59% (0.8449) | 0.50% (0.746) | 0.50% (0.7394) | 0.47% (0.6894) | 0.57% (0.8097) | 0.36% (0.537) | 0.19% (0.2907) | 0.03% (0.0399) | 0.02% (0.0301) | |
| | FF3-alpha | -0.73%*** (-4.9679) | -0.62%*** (-2.6892) | -0.78%*** (-5.5116) | -0.77%*** (-5.2115) | -0.80%*** (-5.4093) | -0.64%*** (-2.7817) | -0.99%*** (-5.7977) | -1.17%*** (-6.2781) | -1.37%*** (-7.4357) | -1.33%*** (-7.8411) | |
| 空头组合 | FF5-alpha | -0.72%*** (-4.2912) | -0.55%*** (-2.4313) | -0.79%*** (-4.9532) | -0.76%*** (-4.5824) | -0.81%*** (-5.0103) | -0.57%*** (-2.5189) | -1.01%*** (-5.5139) | -1.14%*** (-5.6522) | -1.34%*** (-6.6108) | -1.32%*** (-7.4698) | |
| | Sharpe Ratio | 0.1249 | 0.1350 | 0.1077 | 0.1067 | 0.0945 | 0.1264 | 0.0572 | -0.0028 | -0.0614 | -0.0641 | |
| | 均值 | 1.77%*** (5.8044) | 1.94%*** (4.8885) | 1.84%*** (6.0781) | 1.90%*** (6.249) | 1.96%*** (6.34) | 1.96%*** (4.9352) | 2.14%*** (6.7965) | 2.42%*** (7.0277) | 2.62%*** (8.0467) | 2.85%*** (8.8977) | |
| 多空组合 | FF3-alpha | 1.44%*** (5.1912) | 1.53%*** (4.2581) | 1.50%*** (5.4796) | 1.55%*** (5.6053) | 1.59%*** (5.7167) | 1.55%*** (4.306) | 1.93%*** (6.7009) | 2.13%*** (6.6369) | 2.39%*** (7.6748) | 2.61%*** (8.4833) | |
| | FF5-alpha | 1.27%*** (5.1355) | 1.24%*** (3.8168) | 1.34%*** (5.5888) | 1.37%*** (5.638) | 1.43%*** (5.8813) | 1.26%*** (3.868) | 1.78%*** (6.8383) | 1.87%*** (6.8037) | 2.13%*** (7.5882) | 2.37%*** (9.1348) | |
| | Sharpe Ratio | 1.3408 | 1.1044 | 1.4023 | 1.4574 | 1.4863 | 1.1173 | 1.5646 | 1.6533 | 1.8774 | 2.1127 | |

资料来源：作者计算

表4.6 全样本36个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN | Mkt |
|------|--------------|------------------------|-----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.22%*** (2.8627) | 2.25%*** (2.8828) | 2.24%*** (2.8869) | 2.33%*** (2.9858) | 2.33%*** (2.9773) | 2.28%*** (2.9457) | 2.48%*** (3.2061) | 2.62%*** (3.2037) | 2.65%*** (3.2601) | 2.87%*** (3.6442) | 0.58% (1.0706) |
| | FF3-alpha | 0.87%*** (3.4134) | 0.96%*** (3.6419) | 0.89%*** (3.4877) | 0.97%*** (3.7767) | 0.96%*** (3.7483) | 0.89%*** (3.5012) | 1.14%*** (4.9505) | 1.24%*** (4.7354) | 1.30%*** (5.1377) | 1.56%*** (6.3254) | -1.91%** (-2.4883) |
| | FF5-alpha | 0.71%*** (3.1522) | 0.74%*** (3.1179) | 0.73%*** (3.2125) | 0.81%*** (3.5709) | 0.80%*** (3.521) | 0.73%*** (3.2067) | 1.01%*** (4.7953) | 1.03%*** (4.6655) | 1.08%*** (4.8433) | 1.35%*** (6.5599) | -2.28%*** (-2.9257) |
| | Sharpe Ratio | 0.6604 | 0.6670 | 0.6662 | 0.6934 | 0.6913 | 0.6820 | 0.7502 | 0.7690 | 0.7794 | 0.8721 | 0.1680 |
| | 均值 | 0.42% (0.6101) | 0.69% (0.9768) | 0.41% (0.5994) | 0.38% (0.5539) | 0.38% (0.5511) | 0.52% (0.7394) | 0.36% (0.5176) | 0.13% (0.1881) | -0.01% (-0.0166) | 0.04% (0.0542) | |
| | FF3-alpha | -0.82%*** (-5.0217) | -0.50%** (-2.03) | -0.83%*** (-5.0384) | -0.83%*** (-4.9386) | -0.83%*** (-4.9603) | -0.57%** (-2.4468) | -0.93%*** (-5.8163) | -1.21%*** (-7.0937) | -1.36%*** (-7.8717) | -1.25%*** (-6.8868) | |
| 空头组合 | FF5-alpha | -0.85%*** (-4.8661) | -0.48%** (-2.0769) | -0.86%*** (-4.8966) | -0.88%*** (-5.0782) | -0.89%*** (-5.0992) | -0.65%*** (-3.0032) | -0.98%*** (-5.5792) | -1.22%*** (-6.5253) | -1.37%*** (-7.1322) | -1.27%*** (-6.2888) | |
| | Sharpe Ratio | 0.0805 | 0.1757 | 0.0778 | 0.0660 | 0.0652 | 0.1147 | 0.0555 | -0.0264 | -0.0757 | -0.0571 | |
| | 均值 | 1.80%*** (5.5343) | 1.56%*** (3.9748) | 1.83%*** (5.5841) | 1.95%*** (6.0136) | 1.95%*** (6.0051) | 1.76%*** (4.8992) | 2.12%*** (7.2803) | 2.49%*** (7.1239) | 2.66%*** (7.7512) | 2.83%*** (8.5739) | |
| | FF3-alpha | 1.49%*** (4.9637) | 1.26%*** (3.5445) | 1.51%*** (5.0561) | 1.60%*** (5.3978) | 1.59%*** (5.3835) | 1.26%*** (3.6405) | 1.87%*** (7.4281) | 2.25%*** (7.0277) | 2.46%*** (7.8622) | 2.61%*** (8.1224) | |
| | FF5-alpha | 1.37%*** (4.9411) | 1.03%*** (3.0001) | 1.39%*** (5.0272) | 1.49%*** (5.4781) | 1.49%*** (5.4497) | 1.18%*** (3.7871) | 1.79%*** (7.5671) | 2.06%*** (7.3973) | 2.25%*** (7.9904) | 2.41%*** (8.5147) | |
| | Sharpe Ratio | 1.3654 | 0.8858 | 1.3783 | 1.4718 | 1.4672 | 1.1302 | 1.7349 | 1.7275 | 1.8811 | 2.1148 | |

资料来源：作者计算

(3) 集成各类算法预测结果

为了更直观的说明机器学习算法相对于传统线性模型的绩效提升，本文结合了Ridge, Lasso, ElasticNet, PLS, EN-ANN, Xgboost, GBDT和DFN共8中机器学习算法的预测结果构建预测模型。具体而言，就某一月份*t*对某支股票*i*的预测结果可以表示为：

$$R_{ensembles,t} = \frac{1}{8} \sum_{i=1}^8 R_{i,t}$$

集成预测结果即为8种机器学习算法预测收益的简单算术平均值。表4.7中展示了分别在3个月、12个月、24个月和36个月滑动窗口采用这一集成算法的预测绩效，表4.8则对4种滑动窗口该集成算法与OLS预测结果差异进行显著性检验。结果显示，在4种滑动窗口条件下，集成机器学习算法在多空组合和单独做多/空组合上均能够获得显著优于OLS算法的预测绩效。

表4.7 集成机器学习算法预测结果

| | 3 个月滑动窗口 | | | 12 个月滑动窗口 | | |
|------------------|-----------------------|-----------------------|-------------------------|------------------------|------------------------|-------------------------|
| | long-short | long | short | long-short | long | short |
| mean | 2.52% *** (6.9098) | 2.54% *** (3.4829) | -0.02% (0.0357) | 2.83% *** (8.8071) | 2.81% *** (3.7745) | -0.02% (-0.0398) |
| alpha-FF3 | 2.35% *** (6.8268) | 1.29% *** (4.7045) | -1.28% *** (-6.8509) | 30.45% *** (8.2154) | 16.34% *** (5.6929) | -16.64% *** (-7.56) |
| alpha-FF5 | 2.16% *** (6.7923) | 1.17% *** (4.7878) | -1.21% *** (-6.3149) | 27.85% *** (8.249) | 14.36% *** (5.8032) | -16.00% *** (5.8032) |
| sharp | 1.4589 | 0.7508 | -0.0736 | 2.0113 | 0.8511 | -0.0831 |
| | 24 个月滑动窗口 | | | 36 个月滑动窗口 | | |
| | long-short | long | short | long-short | long | short |
| mean | 2.74% *** (8.2013) | 2.80% *** (3.6274) | -0.06% (0.0858) | 2.75% *** (8.2576) | 2.76% *** (3.4521) | 0.01% (0.0137) |
| alpha-FF3 | 2.41% *** (7.9087) | 1.36% *** (5.5846) | -1.25% *** (-6.9761) | 2.44% *** (8.2099) | 1.38% *** (5.3946) | -1.26% *** (-7.7152) |
| alpha-FF5 | 2.15% *** (8.3251) | 1.15% *** (5.6377) | -1.20% *** (-6.3596) | 2.27% *** (8.7969) | 1.21% *** (5.549) | -1.26% *** (-7.1248) |
| sharp | 1.9706 | 0.8385 | -0.0509 | 2.0960 | 0.8229 | -0.0687 |

资料来源：作者计算

表4.8 集成机器学习算法与OLS预测结果差异NW-t检验结果

| | long-short | long | short |
|--------------------|------------|--------|---------|
| ensemble-3 | 5.3606 | 4.3689 | -4.3958 |
| ensemble-12 | 5.7558 | 4.7476 | -4.5011 |
| ensemble-24 | 5.5117 | 4.8877 | -4.5519 |
| ensemble-36 | 6.2649 | 5.5551 | -4.5354 |

资料来源：作者计算

(4) 考虑交易成本时的投资绩效

为证明本文构建的多因子模型的投资绩效具有实际操作价值,文中考虑了包含比例交易成本的情形,并将交易成本(transaction cost)分别设定为单边0.25%、0.5%、0.75%和1%,表4.9中列示了各种条件下投资组合风险收益情形。结果显示,当交易成本为0.25%时,各种多因子模型均能获得显著的超额收益,DFN算法在其中获得最好的表现,多空组合月度收益达到2.46%,经过Fama-French五因子调整后依然能够保持1.93%的月度超额收益,表明本文的多因子策略在中国A股市场能够获得显著的投资绩效。当交易成本为0.5%时,OLS算法构建的多空组合经FF-5因子调整后收益不再显著,而其它算法多空组合仍能保持显著超额收益。而当交易成本上升至1%时,所有算法经FF-5因子调整后的超额收益均不再显著。

2. 剔除市值因子后的预测结果

在单因子多空组合检验过程中,本文发现市值因子(size)构建的多空组合能够达到21.26%的平均年化收益(平均月度收益为1.77%),为了检验多因子模型是否在更大程度上受到size因子的驱动而并非是多因子带来的信息聚合效果,这里从全部96项因子集合中剔除了市值因子(size)和行业调整市值因子(size_ia),用剩余94项因子数据作为输入变量构建预测模型,表4.10中列示了各个预测模型构建投资组合的绩效表现。结果显示,在剔除市值因子后,各项预测模型的投资绩效较全样本(96项因子)而言不存在明显改变,而OLS、Ridge、GBDT和DFN四种算法的预测绩效较全样本而言反而能够获得小幅提升,在很大程度上是由于size因子与其它因子间存在较强的相关关系,在剔除之后反而消除了部分共线性影响,从而提升预测效果。此外,表中数据也证明,多因子模型的预测效果并非由size因子主导,而是多因子共同作用的结果。

表4.9 全样本12个月滑动窗口预测结果(考虑交易成本)

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|-------------------------|--------------|-----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 多空组合 transcost=0.25% | 均值 | 1.41%*** (4.5002) | 1.94%*** (5.0062) | 1.58%*** (5.1613) | 1.65%*** (5.3081) | 1.65%*** (5.2993) | 1.87%*** (5.7405) | 1.76%*** (6.0259) | 2.15%*** (7.0183) | 2.16%*** (7.0169) | 2.46%*** (7.3027) |
| | FF5-alpha | 0.92%*** (3.4243) | 1.12%*** (3.487) | 1.15%*** (4.2873) | 1.21%*** (4.6171) | 1.21%*** (4.6298) | 1.48%*** (5.4291) | 1.33%*** (5.0668) | 1.56%*** (5.5808) | 1.62%*** (5.7569) | 1.93%*** (6.674) |
| | Sharpe Ratio | 0.9901 | 1.0721 | 1.1245 | 1.1843 | 1.1789 | 1.2738 | 1.3287 | 1.5124 | 1.5274 | 1.6872 |
| | 均值 | 0.91%*** (2.8988) | 1.44%*** (3.7177) | 1.08%*** (3.5246) | 1.15%*** (3.704) | 1.15%*** (3.6899) | 1.37%*** (4.2045) | 1.26%*** (4.3185) | 1.65%*** (5.3834) | 1.66%*** (5.3961) | 1.96%*** (5.8205) |
| | FF5-alpha | 0.42% (1.5671) | 0.62%* (1.9269) | 0.65%** (2.4195) | 0.71%*** (2.7032) | 0.71%*** (2.7157) | 0.98%*** (3.5891) | 0.83%*** (3.1578) | 1.06%*** (3.7934) | 1.12%*** (3.978) | 1.43%*** (4.9435) |
| 多空组合 transcost=0.5% | Sharpe Ratio | 0.5754 | 0.7625 | 0.7127 | 0.7740 | 0.7681 | 0.8895 | 0.9010 | 1.1216 | 1.1364 | 1.3126 |
| | 均值 | 0.41% (1.2973) | 0.94%** (2.4293) | 0.58%* (1.8879) | 0.65%** (2.1) | 0.65%** (2.0805) | 0.87%*** (2.6686) | 0.76%*** (2.611) | 1.15%*** (3.7485) | 1.16%*** (3.7753) | 1.46%*** (4.3382) |
| | FF5-alpha | -0.08% (-0.2902) | 0.12% (0.3668) | 0.15% (0.5517) | 0.21% (0.7893) | 0.21% (0.8017) | 0.48%* (1.7491) | 0.33% (1.2488) | 0.56%** (2.0059) | 0.62%** (2.199) | 0.93%*** (3.213) |
| | Sharpe Ratio | 0.1606 | 0.4529 | 0.3009 | 0.3637 | 0.3574 | 0.5052 | 0.4733 | 0.7308 | 0.7454 | 0.9380 |
| | 均值 | -0.09% (-0.3041) | 0.44% (1.1409) | 0.08% (0.2512) | 0.15% (0.4959) | 0.15% (0.4712) | 0.37% (1.1327) | 0.26% (0.9036) | 0.65%** (2.1136) | 0.66%** (2.1544) | 0.96%*** (2.856) |
| 多空组合 transcost=1% | FF5-alpha | -0.58%** (-2.1475) | -0.38% (-1.1933) | -0.35% (-1.3162) | -0.29% (-1.1246) | -0.29% (-1.1124) | -0.02% (-0.0909) | -0.17% (-0.6602) | 0.06% (0.2184) | 0.12% (0.42) | 0.43% (1.4826) |
| | Sharpe Ratio | -0.2541 | 0.1433 | -0.1109 | -0.0466 | -0.0534 | 0.1209 | 0.0455 | 0.3400 | 0.3544 | 0.5634 |

资料来源：作者计算

表4.10 剔除市值因子后12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.37%*** (3.2723) | 2.65%*** (3.5096) | 2.39%*** (3.3029) | 2.45%*** (3.383) | 2.45%*** (3.3851) | 2.48%*** (3.4011) | 2.53%*** (3.4938) | 2.64%*** (3.6164) | 2.61%*** (3.5559) | 2.96%*** (3.9584) |
| | FF3-alpha | 0.94%*** (3.824) | 1.13%*** (4.3524) | 0.96%*** (3.9083) | 1.00%*** (4.1243) | 1.01%*** (4.137) | 1.07%*** (4.4749) | 1.12%*** (5.155) | 1.20%*** (5.2592) | 1.18%*** (5.0348) | 1.54%*** (5.8972) |
| | FF5-alpha | 0.84%*** (3.9268) | 1.00%*** (4.5317) | 0.84%*** (3.8896) | 0.88%*** (4.1372) | 0.88%*** (4.1721) | 1.00%*** (4.6483) | 0.98%*** (5.1436) | 1.02%*** (5.1188) | 0.98%*** (4.746) | 1.37%*** (6.2823) |
| | Sharpe Ratio | 0.7147 | 0.7891 | 0.7225 | 0.7443 | 0.7450 | 0.7587 | 0.7725 | 0.8123 | 0.7995 | 0.8861 |
| | 均值 | 0.36% (0.5524) | 0.26% (0.399) | 0.30% (0.4676) | 0.29% (0.4487) | 0.28% (0.4383) | 0.12% (0.1772) | 0.25% (0.3725) | 0.00% (-0.0006) | -0.02% (-0.0257) | -0.06% (-0.0884) |
| | FF3-alpha | -0.91%*** (-5.3036) | -0.85%*** (-4.1318) | -0.97%*** (-5.8452) | -0.99%*** (-5.6182) | -0.99%*** (-5.6528) | -1.12%*** (-6.0419) | -1.15%*** (-6.9853) | -1.38%*** (-7.6613) | -1.43%*** (-7.7518) | -1.40%*** (-7.7441) |
| 空头组合 | FF5-alpha | -0.92%*** (-4.8306) | -0.84%*** (-4.0596) | -1.01%*** (-5.6336) | -1.04%*** (-5.5151) | -1.05%*** (-5.5459) | -1.17%*** (-6.0915) | -1.13%*** (-6.3064) | -1.30%*** (-6.62) | -1.33%*** (-6.8006) | -1.33%*** (-6.7728) |
| | Sharpe Ratio | 0.0523 | 0.0161 | 0.0318 | 0.0285 | 0.0260 | -0.0335 | 0.0126 | -0.0753 | -0.0806 | -0.0953 |
| 多空组合 | 均值 | 2.01%*** (6.4669) | 2.39%*** (6.2417) | 2.09%*** (6.7523) | 2.16%*** (6.8808) | 2.17%*** (6.9158) | 2.36%*** (7.3193) | 2.28%*** (8.0889) | 2.64%*** (8.9253) | 2.63%*** (8.7713) | 3.02%*** (9.0744) |
| | FF3-alpha | 1.64%*** (5.6525) | 1.77%*** (5.1531) | 1.72%*** (6.0121) | 1.78%*** (6.2155) | 1.79%*** (6.2448) | 1.98%*** (6.8625) | 2.06%*** (7.9685) | 2.37%*** (8.11) | 2.39%*** (8.0989) | 2.73%*** (8.3396) |
| | FF5-alpha | 1.56%*** (5.8037) | 1.63%*** (5.1174) | 1.64%*** (6.2416) | 1.71%*** (6.5316) | 1.72%*** (6.5669) | 1.95%*** (7.1768) | 1.90%*** (7.7613) | 2.11%*** (7.7352) | 2.10%*** (7.5997) | 2.50%*** (8.7658) |
| | Sharpe Ratio | 1.4857 | 1.3757 | 1.5505 | 1.5762 | 1.5837 | 1.6620 | 1.7729 | 1.9982 | 1.9680 | 2.1203 |
| | 均值 | 2.01%*** (6.4669) | 2.39%*** (6.2417) | 2.09%*** (6.7523) | 2.16%*** (6.8808) | 2.17%*** (6.9158) | 2.36%*** (7.3193) | 2.28%*** (8.0889) | 2.64%*** (8.9253) | 2.63%*** (8.7713) | 3.02%*** (9.0744) |

资料来源：作者计算

3. 异象因子的重要性

特征筛选作为特征工程的一项重要组成部分，能够提取特征中的重要组成部分，发现冗杂数据中关键的数据条目。具体到因子研究而言，本文试图通过采用特征筛选方法，探索和发现影响中国市场的关键性因子，从机器学习角度进一步审视中国股票截面收益的影响因素。因此，在上述分析的基础上，为了在全部96项因子集合中确定在中国市场上具有最强预测效力的有效因子集合，本文分别利用单因子检验、OLS、GBDT、Lasso和DFN5种特征筛选方式进行了有效特征筛选。

首先，对于单因子筛选来讲，筛选方式最为直接，即选出在统计学意义上显著的因子（多空组合收益t统计量大于等于1.96）作为代表性因子集合。在挑选出代表性因子集合后，本文将因子集合中显著性最高（t值最大）的因子的重要性数值记为100，其余因子的重要性数值按照最大的因子显著性进行折算得出。在此种度量方式下，本文共筛选除包含22项因子的代表性因子集合，其中包含交易摩擦因子12项，动量因子3项，成长因子3项，财务流动性因子2项，盈利因子1项，价值因子1项，具体因子名称及重要性数值见附录2。

对于OLS而言，首先根据去除单因子数据后回归模型产生的投资组合月度收益同全因子月度收益进行差运算，获得去除单一因子的收益损失。在得出收益损失后，本文将收益损失最大的因子的重要性数值记为100，剩余的因子按照收益损失最大值折算，得出整体因子重要性数值，并将因子按照该数值从大到小进行排列。在此排列下，本文选取累积收益差达到全因子模型收益率的因子集合作为本文所筛选出来的代表性因子集合。在这种度量方式下，本文共筛选出包含28项因子的代表性因子集合，其中包含成长因子9项，盈利因子8项，交易摩擦因子5项，财务流动性因子3项，动量因子2项，价值因子1项，具体因子名称以及重要性数值见附录3。

对于GBDT来讲，根据2.1.7节中列示的特征重要性的度量方式，最终获取了每项特征的重要程度(IMP)数值，将因子按照IMP数值从大至小排列，选取累计重要程度(acc-IMP)达到95%的因子集合即为本文筛选出的代表性因子集合。在这种度量方式下，本文共筛选出包含28项因子的代表性因子集合，其中包含交易摩擦因子12项，动量因子6项，价值因子1项，成长因子8项以及财务流动性因子1项，具体因子名称和重要性数值见附录4。

就Lasso而言，为了确定进行筛选的模型参数，本文首先确定了每项特征系数值相对于惩罚系数的变化情况，如图4.1所示，图形纵轴表示各项因子的系数值，横轴表示根据Lasso算法的惩罚系数 α 计算出的 $-\log(\alpha)$ 数值。图形结果显示，当 $\alpha = 0.0007$ （即 $-\log(\alpha) = 3.15$ ）

时，Lasso回归系数表现出明显的缩减态势，因此，本文将模型的惩罚系数定为 $\alpha = 0.0007$ 。Lasso回归的特征重要性度量方式即为因子系数的绝对值，本文筛选出回归中系数不为0的共32项因子作为代表性特征集合，其中包含交易摩擦因子13项，动量因子4项，价值因子7项，成长因子2项，盈利因子3项及财务流动性因子3项，具体因子名称和重要性数值见附录5。

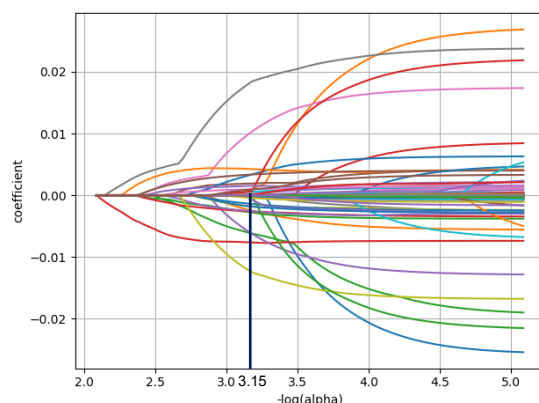


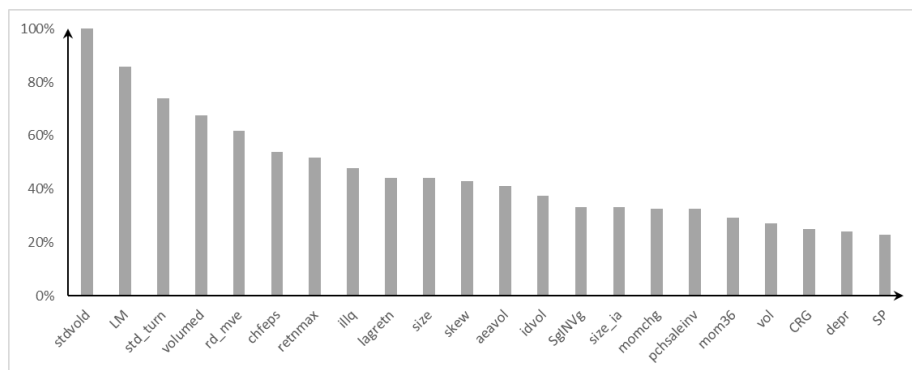
图4.1 Lasso系数衰减图形

资料来源：作者绘制

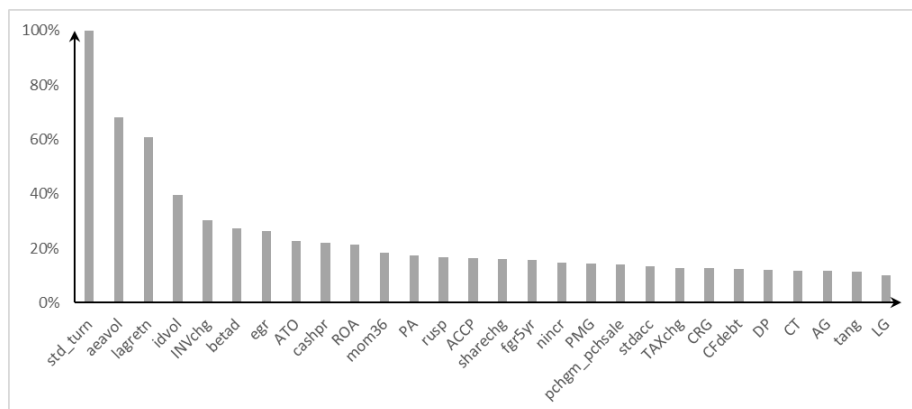
对于DFN而言，在初始的去除单一因子后的收益损失与因子重要性的计算方式上同OLS因子筛选方法中收益损失计算方式相同。但由于DFN计算出来的因子在累积收益差达到全因子月度收益附近时，其因子重要性差距较小，为了降低忽略重要因子的风险，本文在累积收益差达到全因子收益之后，计算了相邻因子重要性差距这一指标度量因子间的差异。根据这一指标，本文选取了最近的相邻因子重要性差距大于1的因子及该因子之前的所有因子作为DFN算法筛选出来的代表性因子集合。在这种度量方式下，本文共筛选出包含25项因子的代表性因子集合，其中包含成长因子10项，财务流动因子6项，交易摩擦因子5项，价值因子2项，盈利因子1项，动量因子1项，具体因子名称和重要性数值见附录6。

图4.2中列示了五种特征筛选方式筛选得到代表性特征集合及其重要程度（均以重要的因子重要性为100%进行数值转换）。单因子检验筛选出的前三位因子分别为交易额波动率(stdvold)，标准化的换手率(LM)和换手率的波动率(std_turn)，均为交易摩擦类因子；OLS算法筛选出的前三位因子为换手率的波动率(std_turn)，收益公告异常交易量(aeavol)和短期反转(lagretn)；GBDT筛选出的前三位因子分别为总波动率(retvol)，12个月动量(mom12)和折旧变动百分比(pchdepr)；Lasso筛选出的前三位因子分别为账面市值比(BM)，行业调整账面市值比(BM_ia)和总波动率(retvol)因子；DFN筛选出的前三位因子则分别为总波动率(retvol)，收益公告异常交易量(aeavol)和速动比率(QR)。其中总波动率(retvol)在上述5种算法排名前3位出现3次，换手率的波动率(std_turn)、收益公告异常交易量(aeavol)则分别出现2次，且均为交易摩擦类因子，在对各个算法筛选出的代表性因子的统计描述也显

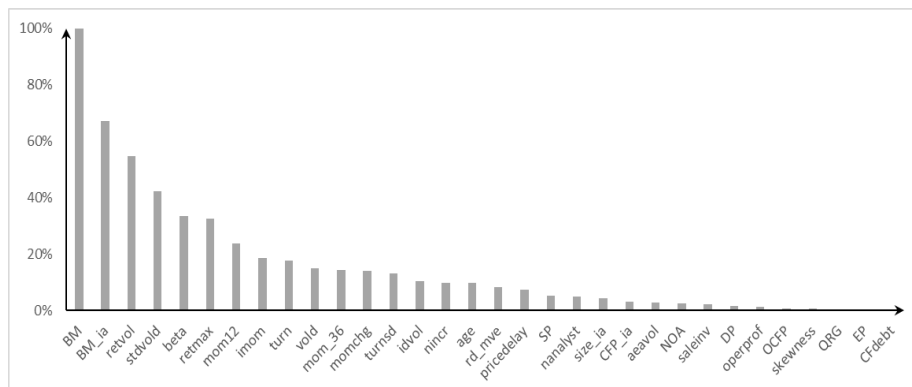
示，总体而言交易摩擦类因子在中国A股市场具有更好的预测效果。



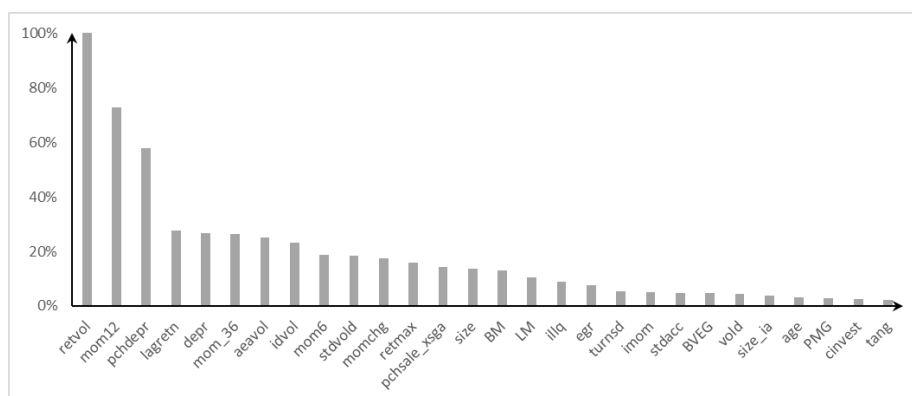
a. 单因子筛选



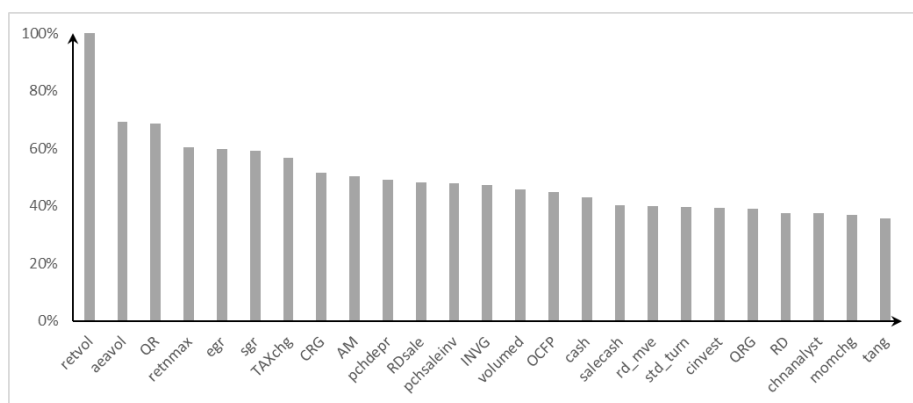
b. OLS筛选



c. Lasso筛选



d. GBDT筛选



e. DFN筛选

图4.2 各种算法筛选特征重要性对比

资料来源：作者绘制

表4.11-4.15分别列示了根据以上5种方式筛选出的代表性因子集合作为输入变量构建模型的投资绩效。结果显示，经过单因子、GBDT和DFN筛选后的特征集合构建的预测模型相较于全样本而言投资绩效能够获得小幅提升（例如全样本构建的DFN预测模型多空组合平均月度收益率为2.96%，经单因子、GBDT和DFN筛选后构建的多空组合平均月度收益则分别上升至2.99%，3.08%和3.13%），表明有效的机器学习算法能够提出冗余特征，提高预测效率和预测精确度，也显示出非线性算法(GBDT、DFN)相对于线性算法（OLS、Lasso）在筛选有效特征集合时表现出的优越性。

表4.11 单因子检验筛选特征12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.44%*** (3.3293) | 2.88%*** (4.2194) | 2.43%*** (3.3269) | 2.43%*** (3.318) | 2.46%*** (3.3511) | 2.88%*** (3.792) | 2.53%*** (3.5367) | 2.66%*** (3.5965) | 2.56%*** (3.4848) | 2.94%*** (4.0333) |
| | FF3-alpha | 0.97%*** (4.1659) | 1.32%*** (5.5115) | 0.97%*** (4.161) | 0.97%*** (4.121) | 0.99%*** (4.1113) | 1.32%*** (5.0696) | 1.14%*** (5.218) | 1.20%*** (5.155) | 1.12%*** (5.1267) | 1.50%*** (6.0714) |
| | FF5-alpha | 0.90%*** (4.2957) | 1.12%*** (4.4184) | 0.90%*** (4.2951) | 0.89%*** (4.2108) | 0.88%*** (4.1868) | 1.12%*** (5.2541) | 0.95%*** (5.0252) | 1.00%*** (5.0499) | 0.93%*** (4.9187) | 1.26%*** (5.9795) |
| | Sharpe Ratio | 0.7325 | 0.8584 | 0.7319 | 0.7308 | 0.7382 | 0.8584 | 0.7841 | 0.8084 | 0.7799 | 0.9091 |
| | 均值 | 0.21% (0.3233) | -0.13% (-0.2046) | 0.21% (0.3253) | 0.22% (0.3416) | 0.17% (0.2559) | -0.12% (-0.1922) | 0.18% (0.2837) | 0.13% (-0.2046) | 0.10% (0.1565) | -0.05% (-0.0751) |
| | FF3-alpha | -1.03%*** (-5.2477) | -1.18%*** (-5.5408) | -1.03%*** (-5.2391) | -1.02%*** (-5.1581) | -1.07%*** (-5.4972) | -1.18%*** (-5.8447) | -1.17%*** (-6.7077) | -1.18%*** (-5.5408) | -1.25%*** (-6.766) | -1.38%*** (-7.592) |
| 空头组合 | FF5-alpha | -1.07%*** (-5.1721) | -1.17%*** (-5.2273) | -1.07%*** (-5.1668) | -1.06%*** (-5.1194) | -1.11%*** (-5.5479) | -1.17%*** (-5.746) | -1.10%*** (-6.2279) | -1.17%*** (-5.2273) | -1.18%*** (-5.8283) | -1.33%*** (-7.0744) |
| | Sharpe Ratio | 0.0005 | -0.1215 | 0.0009 | 0.0047 | -0.0154 | -0.1215 | -0.0095 | -0.1215 | -0.0387 | -0.0938 |
| | 均值 | 2.23%*** (6.7642) | 3.01%*** (8.6664) | 2.22%*** (6.7526) | 2.21%*** (6.6923) | 2.29%*** (6.818) | 3.00%*** (7.6932) | 2.35%*** (8.4785) | 2.53%*** (8.2221) | 2.46%*** (8.2546) | 2.99%*** (9.4967) |
| | FF3-alpha | 1.79%*** (6.0544) | 2.29%*** (7.3042) | 1.79%*** (6.0456) | 1.78%*** (5.9591) | 1.85%*** (6.0642) | 2.29%*** (7.2) | 2.10%*** (7.6634) | 2.20%*** (7.4742) | 2.16%*** (7.4251) | 2.66%*** (8.6244) |
| | FF5-alpha | 1.75%*** (6.0645) | 2.09%*** (6.5343) | 1.75%*** (6.0611) | 1.73%*** (5.9656) | 1.78%*** (6.194) | 2.09%*** (7.2768) | 1.84%*** (7.5415) | 1.93%*** (7.0734) | 1.90%*** (7.0154) | 2.38%*** (8.6287) |
| | Sharpe Ratio | 1.5498 | 1.7688 | 1.5466 | 1.5332 | 1.5741 | 1.7688 | 1.8996 | 1.8242 | 1.8087 | 2.1962 |

资料来源：作者计算

表4.12 OLS筛选特征12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.50%*** (3.4129) | 2.48%*** (3.6853) | 2.51%*** (3.4166) | 2.49%*** (3.3629) | 2.48%*** (3.3603) | 2.48%*** (3.3232) | 2.39%*** (3.3378) | 2.63%*** (3.5298) | 2.52%*** (3.4371) | 2.80%*** (3.7668) |
| | FF3-alpha | 1.09%*** (4.0715) | 1.00%*** (3.8123) | 1.09%*** (4.0832) | 1.04%*** (3.7828) | 1.04%*** (3.7707) | 1.00%*** (3.5514) | 0.96%*** (4.2943) | 1.13%*** (4.8513) | 1.04%*** (4.5437) | 1.34%*** (5.0607) |
| | FF5-alpha | 0.97%*** (4.336) | 0.89%*** (3.1938) | 0.98%*** (4.3526) | 0.95%*** (4.1434) | 0.95%*** (4.1349) | 0.89%*** (3.8293) | 0.76%*** (3.9196) | 0.97%*** (4.7456) | 0.82%*** (4.0874) | 1.19%*** (5.3023) |
| | Sharpe Ratio | 0.7617 | 0.7400 | 0.7626 | 0.7498 | 0.7491 | 0.7398 | 0.7341 | 0.7900 | 0.7654 | 0.8471 |
| | 均值 | 0.18% (0.2752) | 0.23% (0.3722) | 0.18% (0.2773) | 0.15% (0.2208) | 0.15% (0.222) | 0.22% (0.329) | 0.58% (0.8456) | 0.20% (0.3093) | 0.26% (0.3994) | 0.13% (0.1965) |
| | FF3-alpha | -1.07%*** (-5.2743) | -0.96%*** (-4.8656) | -1.07%*** (-5.2696) | -1.10%*** (-5.2142) | -1.10%*** (-5.2133) | -0.97%*** (-4.1067) | -0.81%*** (-4.8467) | -1.23%*** (-6.4251) | -1.16%*** (-6.4987) | -1.21%*** (-6.9803) |
| 空头组合 | FF5-alpha | -1.15%*** (-6.0028) | -1.04%*** (-4.961) | -1.15%*** (-5.9977) | -1.19%*** (-5.9559) | -1.18%*** (-5.9543) | -1.05%*** (-4.9448) | -0.76%*** (-4.0634) | -1.17%*** (-5.4767) | -1.09%*** (-5.5255) | -1.22%*** (-6.8469) |
| | Sharpe Ratio | -0.0109 | 0.0071 | -0.0104 | -0.0235 | -0.0232 | 0.0021 | 0.1265 | -0.0023 | 0.0182 | -0.0281 |
| | 均值 | 2.32%*** (6.1942) | 2.25%*** (6.5772) | 2.33%*** (6.2011) | 2.34%*** (6.0269) | 2.33%*** (6.0135) | 2.26%*** (5.7652) | 1.81%*** (6.4057) | 2.43%*** (7.7317) | 2.26%*** (7.1626) | 2.67%*** (8.3702) |
| | FF3-alpha | 1.95%*** (5.3179) | 1.75%*** (5.0774) | 1.95%*** (5.324) | 1.93%*** (5.0121) | 1.93%*** (4.998) | 1.76%*** (4.2009) | 1.55%*** (5.8204) | 2.15%*** (7.1902) | 1.98%*** (6.7097) | 2.35%*** (7.6499) |
| 多空组合 | FF5-alpha | 1.92%*** (6.2351) | 1.72%*** (4.7738) | 1.92%*** (6.2468) | 1.92%*** (6.0113) | 1.92%*** (5.9986) | 1.73%*** (5.043) | 1.31%*** (5.2692) | 1.92%*** (6.7659) | 1.70%*** (6.0058) | 2.19%*** (8.4242) |
| | Sharpe Ratio | 1.4723 | 1.3080 | 1.4738 | 1.4299 | 1.4267 | 1.3173 | 1.4273 | 1.7169 | 1.5879 | 1.9639 |

资料来源：作者计算

表4.13 GBDT筛选特征12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.43%*** (3.3071) | 2.67%*** (3.4878) | 2.48%*** (3.3402) | 2.46%*** (3.3089) | 2.46%*** (3.3148) | 2.68%*** (3.4958) | 2.55%*** (3.5306) | 2.74%*** (3.7139) | 2.68%*** (3.6596) | 2.93%*** (3.9157) |
| | FF3-alpha | 0.97%*** (4.1606) | 1.12%*** (4.5082) | 1.00%*** (4.2304) | 0.96%*** (3.9268) | 0.97%*** (4.0391) | 1.12%*** (4.5259) | 1.17%*** (5.5329) | 1.29%*** (5.6579) | 1.24%*** (5.7652) | 1.50%*** (6.2568) |
| | FF5-alpha | 0.89%*** (4.2497) | 0.97%*** (4.6369) | 0.90%*** (4.3014) | 0.86%*** (4.0115) | 0.88%*** (4.143) | 0.98%*** (4.6579) | 1.04%*** (5.4673) | 1.15%*** (5.8554) | 1.11%*** (5.8827) | 1.35%*** (6.5966) |
| | Sharpe Ratio | 0.7262 | 0.7894 | 0.7373 | 0.7293 | 0.7299 | 0.7914 | 0.7863 | 0.8368 | 0.8158 | 0.8831 |
| | 均值 | 0.36% (0.553) | 0.10% (0.1608) | 0.32% (0.4902) | 0.23% (0.3591) | 0.33% (0.5124) | 0.10% (0.151) | 0.18% (0.2815) | 0.05% (0.0737) | -0.03% (-0.0473) | -0.15% (-0.228) |
| | FF3-alpha | -0.93%*** (-5.1217) | -0.99%*** (-5.0466) | -0.96%*** (-5.2463) | -1.03%*** (-5.6413) | -0.96%*** (-5.1633) | -0.99%*** (-5.0668) | -1.18%*** (-7.1614) | -1.30%*** (-6.872) | -1.41%*** (-7.66) | -1.50%*** (-7.8492) |
| 空头组合 | FF5-alpha | -0.95%*** (-5.0508) | -1.01%*** (-4.9737) | -0.98%*** (-5.232) | -1.05%*** (-5.5689) | -0.98%*** (-5.213) | -1.01%*** (-4.9932) | -1.15%*** (-6.7467) | -1.26%*** (-6.0544) | -1.35%*** (-6.7611) | -1.44%*** (-7.2185) |
| | Sharpe Ratio | 0.0523 | -0.0388 | 0.0380 | 0.0077 | 0.0431 | -0.0410 | -0.0094 | -0.0578 | -0.0861 | -0.1288 |
| | 均值 | 2.07%*** (6.6927) | 2.57%*** (6.7245) | 2.16%*** (6.9586) | 2.23%*** (6.8657) | 2.13%*** (6.663) | 2.58%*** (6.7485) | 2.37%*** (8.8171) | 2.71%*** (8.7887) | 2.71%*** (9.1122) | 3.08%*** (9.6389) |
| | FF3-alpha | 1.69%*** (5.922) | 1.90%*** (6.0081) | 1.75%*** (6.0556) | 1.78%*** (5.9695) | 1.72%*** (5.8098) | 1.91%*** (6.0313) | 2.14%*** (8.1401) | 2.39%*** (7.8455) | 2.44%*** (8.3532) | 2.79%*** (8.9773) |
| | FF5-alpha | 1.63%*** (5.9027) | 1.77%*** (6.1906) | 1.67%*** (6.1437) | 1.70%*** (6.1516) | 1.65%*** (5.9784) | 1.78%*** (6.216) | 1.98%*** (8.0917) | 2.20%*** (7.8748) | 2.24%*** (8.2481) | 2.58%*** (9.0554) |
| | Sharpe Ratio | 1.4708 | 1.5035 | 1.5402 | 1.5407 | 1.4718 | 1.5107 | 1.9642 | 1.9735 | 2.0498 | 2.2728 |

资料来源：作者计算

表4.14 Lasso筛选特征12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.28%*** (3.2059) | 2.36%*** (3.1769) | 2.28%*** (3.2047) | 2.28%*** (3.2091) | 2.28%*** (3.2076) | 2.39%*** (3.2228) | 2.44%*** (3.3755) | 2.64%*** (3.5557) | 2.66%*** (3.6238) | 2.81%*** (3.8385) |
| | FF3-alpha | 0.88%*** (4.0025) | 0.86%*** (3.2894) | 0.88%*** (4.0032) | 0.88%*** (4.0217) | 0.88%*** (4.0157) | 0.87%*** (3.3775) | 1.03%*** (4.6763) | 1.16%*** (4.8509) | 1.21%*** (5.4618) | 1.36%*** (5.5943) |
| | FF5-alpha | 0.80%*** (3.9874) | 0.74%*** (3.301) | 0.80%*** (3.9981) | 0.80%*** (4.0057) | 0.80%*** (4.0105) | 0.74%*** (3.3016) | 0.86%*** (4.4074) | 0.97%*** (4.9014) | 1.01%*** (5.3591) | 1.23%*** (5.9758) |
| | Sharpe Ratio | 0.6925 | 0.6999 | 0.6925 | 0.6931 | 0.6928 | 0.7094 | 0.7460 | 0.8021 | 0.8137 | 0.8535 |
| | 均值 | 0.30% (0.4628) | 0.16% (0.2485) | 0.30% (0.4628) | 0.31% (0.468) | 0.30% (0.4665) | 0.14% (0.2225) | 0.34% (0.5264) | 0.26% (0.3952) | 0.18% (0.2864) | 0.06% (0.0846) |
| | FF3-alpha | -0.98%*** (-5.3332) | -0.96%*** (-4.7323) | -0.98%*** (-5.3271) | -0.97%*** (-5.3117) | -0.97%*** (-5.3163) | -0.97%*** (-4.7664) | -1.05%*** (-6.6563) | -1.15%*** (-6.2036) | -1.21%*** (-7.229) | -1.30%*** (-8.3609) |
| 空头组合 | FF5-alpha | -1.01%*** (-5.3332) | -0.96%*** (-4.674) | -1.01%*** (-5.3165) | -1.01%*** (-5.3182) | -1.01%*** (-5.3211) | -0.96%*** (-4.6747) | -1.00%*** (-5.8103) | -1.01%*** (-5.0613) | -1.10%*** (-6.0489) | -1.24%*** (-7.5466) |
| | Sharpe Ratio | 0.0324 | -0.0184 | 0.0324 | 0.0336 | 0.0333 | -0.0244 | 0.0459 | 0.0171 | -0.0096 | -0.0552 |
| | 均值 | 1.98%*** (7.3064) | 2.20%*** (6.2119) | 1.98%*** (7.3045) | 1.97%*** (7.3279) | 1.98%*** (7.3179) | 2.25%*** (6.3224) | 2.10%*** (7.6555) | 2.38%*** (7.5306) | 2.48%*** (8.5624) | 2.75%*** (9.5871) |
| | FF3-alpha | 1.65%*** (6.3182) | 1.61%*** (4.9373) | 1.65%*** (6.3032) | 1.64%*** (6.3292) | 1.64%*** (6.322) | 1.63%*** (5.0362) | 1.87%*** (7.3417) | 2.11%*** (6.7312) | 2.21%*** (7.6305) | 2.45%*** (8.5726) |
| | FF5-alpha | 1.60%*** (6.341) | 1.49%*** (4.8586) | 1.60%*** (6.3222) | 1.60%*** (6.3467) | 1.60%*** (6.347) | 1.49%*** (4.8635) | 1.65%*** (6.8155) | 1.77%*** (6.349) | 1.91%*** (7.2269) | 2.26%*** (9.0573) |
| | Sharpe Ratio | 1.6309 | 1.3065 | 1.6309 | 1.6336 | 1.6325 | 1.3378 | 1.6609 | 1.6633 | 1.8653 | 2.1346 |

资料来源：作者计算

表4.15 DFN筛选特征12个月滑动窗口预测结果

| | | OLS | FC | Ridge | Lasso | ElasticNet | PLS | EN-ANN | Xgboost | GBDT | DFN |
|------|--------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 多头组合 | 均值 | 2.43%*** (3.3035) | 2.38%*** (3.5905) | 2.43%*** (3.3044) | 2.44%*** (3.2987) | 2.44%*** (3.3031) | 2.39%*** (3.2366) | 2.43%*** (3.4178) | 2.78%*** (3.7533) | 2.74%*** (3.7007) | 3.02%*** (4.0275) |
| | FF3-alpha | 1.02%*** (4.1019) | 0.95%*** (3.8574) | 1.02%*** (4.1025) | 1.01%*** (4.0562) | 1.01%*** (4.0702) | 0.95%*** (3.5975) | 1.07%*** (4.8011) | 1.63%*** (5.7616) | 1.36%*** (5.6722) | 1.62%*** (6.1447) |
| | FF5-alpha | 0.92%*** (4.1952) | 0.80%*** (3.0217) | 0.92%*** (4.1955) | 0.91%*** (4.1308) | 0.91%*** (4.1513) | 0.80%*** (3.6314) | 0.92%*** (4.5163) | 1.39%*** (5.3296) | 1.17%*** (5.3119) | 1.44%*** (6.6569) |
| | Sharpe Ratio | 0.7222 | 0.7183 | 0.7224 | 0.7210 | 0.7221 | 0.7188 | 0.7540 | 2.0260 | 0.8382 | 0.9068 |
| | 均值 | 0.00% (0.7368) | 0.07% (-0.0119) | -0.01% (-0.011) | 0.01% (0.0159) | 0.01% (0.016) | 0.07% (0.1121) | 0.47% (0.7368) | 0.09% (0.1408) | 0.04% (0.0585) | -0.11% (-0.1704) |
| | FF3-alpha | -0.90%*** (-5.9861) | -1.27%*** (-6.9742) | -1.27%*** (-6.9683) | -1.26%*** (-6.7842) | -1.26%*** (-6.7884) | -1.16%*** (-5.9242) | -0.90%*** (-5.9861) | -1.32%*** (-7.7223) | -1.37%*** (-8.1689) | -1.47%*** (-9.3644) |
| 空头组合 | FF5-alpha | -0.82%*** (-4.8559) | -1.30%*** (-6.8152) | -1.30%*** (-6.8135) | -1.28%*** (-6.7909) | -1.28%*** (-6.7918) | -1.13%*** (-5.6671) | -0.82%*** (-4.8559) | -1.25%*** (-6.6159) | -1.28%*** (-6.9333) | -1.47%*** (-8.767) |
| | Sharpe Ratio | 0.0927 | -0.0788 | -0.0786 | -0.0730 | -0.0730 | -0.0502 | 0.0927 | -0.0410 | -0.0606 | -0.1176 |
| 多空组合 | 均值 | 2.43%*** (7.5366) | 2.31%*** (7.5589) | 2.44%*** (7.5303) | 2.43%*** (7.447) | 2.43%*** (7.4513) | 2.32%*** (6.6523) | 1.96%*** (6.9511) | 2.69%*** (8.9813) | 2.70%*** (9.2481) | 3.13%*** (10.0672) |
| | FF3-alpha | 2.07%*** (6.8181) | 1.90%*** (6.1379) | 2.08%*** (6.8097) | 2.06%*** (6.7311) | 2.06%*** (6.7366) | 1.91%*** (5.7538) | 1.75%*** (6.3792) | 2.48%*** (8.3492) | 2.52%*** (8.5782) | 2.87%*** (9.4673) |
| | FF5-alpha | 2.01%*** (7.0397) | 1.72%*** (5.4169) | 2.01%*** (7.0361) | 1.98%*** (6.9901) | 1.98%*** (6.9964) | 1.72%*** (5.6995) | 1.53%*** (5.8377) | 2.20%*** (7.6762) | 2.24%*** (8.0126) | 2.70%*** (10.4025) |
| | Sharpe Ratio | 1.4021 | 1.5077 | 1.7012 | 1.6736 | 1.6754 | 1.5083 | 1.5415 | 2.0260 | 2.0648 | 2.3230 |
| | 均值 | 2.43%*** (7.5366) | 2.31%*** (7.5589) | 2.44%*** (7.5303) | 2.43%*** (7.447) | 2.43%*** (7.4513) | 2.32%*** (6.6523) | 1.96%*** (6.9511) | 2.69%*** (8.9813) | 2.70%*** (9.2481) | 3.13%*** (10.0672) |
| | FF3-alpha | 2.07%*** (6.8181) | 1.90%*** (6.1379) | 2.08%*** (6.8097) | 2.06%*** (6.7311) | 2.06%*** (6.7366) | 1.91%*** (5.7538) | 1.75%*** (6.3792) | 2.48%*** (8.3492) | 2.52%*** (8.5782) | 2.87%*** (9.4673) |

资料来源：作者计算

五、结论

本文利用1997年1月-2018年10月中国A股市场数据构建了包含96项异象因子的数据集，基于预测组合算法、Lasso回归、岭回归、弹性网络回归、偏最小二乘回归、梯度提升树、极端梯度提升树、集成神经网络和深度前馈网络9种机器学习算法构建回归模型，将其与因子策略结合构建基本面量化投资策略，对比了传统OLS回归与机器学习算法构建组合投资绩效差异，并从中原始因子集合中筛选出中国A股市场的代表性因子集合。

第一，各个因子间存在传统OLS回归难以识别的复杂非线性关联，而机器学习算法能够自动识别因子间的隐含特征获得更好的预测效果。实证结果显示，在12个月滑动窗口下，将各种机器学习算法预测结果进行简单加权构建多空组合的平均月度收益为2.83%，而同期OLS回归构建多空组合月度收益仅为1.90%。统计检验结果也表明机器学习算法能够获得显著超过OLS的超额收益。此外，本文构建的投资策略即使在考虑交易成本和中国股票市场做空限制的条件下也能够获得显著的超额收益，可指导未来的投资决策。

第二，非线性机器学习算法(GBDT、Xgboost、DFN)在多因子策略中的绩效表现总体优于线性机器学习算法(PLS、Lasso、Ridge、ElasticNet、FC)。作为深度学习算法的代表，深度前馈网络算法(DFN)在各个时间窗口和因子集合内构建的预测模型均获得了最高的投资绩效，显示了深度学习算法相对于传统机器学习算法的优越性。

第三，采用5种特征选择方式（单因子检验、OLS、GBDT、Lasso和DFN）从96项原始因子集合中筛选出因子，结果显示经过单因子检验、GBDT和DFN筛选后的因子集合构建的预测模型投资绩效能够小幅超越原始因子集合，说明有效的特征筛选方式能够通过去除冗余特征提高算法效率和预测精度，也表现出非线性算法在特征筛选过程中的优越性。同时对筛选出的代表性因子集合的分析显示，在中国A股市场，以总波动率、换手率的波动率和收益公告异常交易量为代表的交易摩擦类因子对股票收益具有很强的预测效力，而基于股票基本面特征构建的异象因子对收益的预测绩效则相对较弱。

第四，本文已经证明深度学习的代表性算法深度前馈网络能够获得相较于传统算法更高的投资绩效，但随着机器学习算法在投资领域研究的不断深入，更有效的深度学习算法被不断应用以提高数据的表示能力及模型的泛化能力，因此本文下一阶段的研究重点之一即为探究深度学习算法在多因子量化策略中的应用。此外，本文的因子集合仅包含96项异象因子，但截至目前相关研究中提出的因子数目已多达400余项，更广泛的因子集合能够通过聚合更

多信息提高预测精度，基于此，本文在下一步研究中也将在当前因子集合，以期构建更有效的量化投资策略。

参考文献

- [1] 张然, 汪荣飞. 基本面量化投资: 运用财务分析和量化策略获取超额收益[M]. 北京大学出版社, 2017.
- [2] Lee C M, So E C. Alphanomics: The informational underpinnings of market efficiency[J]. Foundations and Trends® in Accounting, 2015, 9(2–3): 59–258.
- [3] Green J, Hand J R M, Zhang X F. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns[J]. The Review of Financial Studies, 2017, 30(12): 4389–4436.
- [4] Hou K, Xue C, Zhang L. Replicating Anomalies[J]. The Review of Financial Studies, 2019, forthcoming.
- [5] Gu S, Kelly B T, Xiu D. Empirical Asset Pricing via Machine Learning[J]. Chicago Booth Research Paper No. 18-04, 2018.
- [6] Cochrane J H. Presidential address: Discount rates[J]. The Journal of finance, 2011, 66(4): 1047–1108.
- [7] 黄乃静, 于明哲. 机器学习对经济学研究的影响研究进展[J]. 经济学动态, 2018(7): 115–129.
- [8] Rapach D E, Strauss J K, Zhou G. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy[J]. The Review of Financial Studies, 2010, 23(2): 821–862.
- [9] Feng G, Giglio S, Xiu D. Taming the Factor Zoo: A Test of New Factors[J]. Fama-Miller Working Paper, Chicago Booth Research Paper No. 17-04, 2017.
- [10] Messmer M, Audrino F. The (Adaptive) Lasso in the Zoo - Firm Characteristic Selection in the Cross-Section of Expected Returns[J]. SSRN, 2017.
- [11] Freyberger J, Neuhierl A, Weber M. Dissecting Characteristics Nonparametrically[J]. Fama-Miller Working Paper, Chicago Booth Research Paper No. 17-32, 2018.
- [12] Kelly B T, Pruitt S, Su Y. Characteristics Are Covariances: A Unified Model of Risk and Return[J]. SSRN, 2018.
- [13] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2018, 270(2): 654–669.
- [14] 王文波, 费浦生, 羿旭明. 基于 EMD 与神经网络的中国股票市场预测[J]. 系统工程理论与实践, 2010, 30(06): 1027–1033.
- [15] 李斌, 林彦, 唐闻轩. ML-TEA: 一套基于机器学习和技术分析量化投资算法[J]. 系统工程理论与实践, 2017, 37(5): 1089–1100.
- [16] Lewellen J. The Cross-section of Expected Stock Returns[J]. Critical Finance Review, 2015, 4(2): 1–44.
- [17] Moritz B, Zimmermann T. Tree-based conditional portfolio sorts: The relation between past and future stock returns[J]. SSRN, 2016.
- [18] Demiguel V, Martin-Utrera A, Nogales F J, Uppal R. A Transaction-Cost Perspective on the Multitude of Firm Characteristics[J]. SSRN, 2019.
- [19] Fama E F, French K R. Fama & French - Common risk factors in the returns on stocks and bounds.pdf[M]. 1993, 33.
- [20] Fama E F, French K R. A five-factor asset pricing model[J]. Journal of Financial Economics, 2015, 116(1): 1–22.

- [21] Ng V, Engle R F, Rothschild M. A multi-dynamic-factor model for stock returns[J]. *Journal of Econometrics*, 1992, 52(1): 245–266.
- [22] Driessen J, Klaassen P, Melenberg B. The Performance of Multi-Factor Term Structure Models for Pricing and Hedging Caps and Swaptions[J]. *Journal of Financial and Quantitative Analysis*, 2003, 38(3): 635–672.
- [23] Harvey C R, Liu Y, Zhu H. ... and the Cross-Section of Expected Returns[J]. *The Review of Financial Studies*, 2015, 29(1): 5–68.
- [24] Green J, Hand J R M, Zhang X F. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns[J]. *The Review of Financial Studies*, 2017, 30(12): 4389–4436.
- [25] 朱英伦, 刘杰. 中国股票市场因子研究综述[J]. *现代管理科学*, 2018, No.304(07): 44–46.
- [26] 胡熠, 顾明. 巴菲特的阿尔法:来自中国股票市场的实证研究[J]. *管理世界*, 2018.
- [27] Light N, Maslov D, Rytchkov O. Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach[J]. *The Review of Financial Studies*, 2017, 30(4): 1339–1381.
- [28] Kozak S, Nagel S, Santosh S. Shrinking the Cross Section[J]. 2017.
- [29] Kelly B, Pruitt S, Su Y. Characteristics Are Covariances: A Unified Model of Risk and Return[J]. *National Bureau of Economic Research Working Paper Series*, 2018, No. 24540.
- [30] Gu S, Kelly B, Xiu D. Empirical Asset Pricing via Machine Learning[J]. *National Bureau of Economic Research Working Paper Series*, 2018, No. 25398.
- [31] 潘莉, 徐建国. A 股市场的风险与特征因子[J]. *金融研究*, 2011(10): 140–154.
- [32] Hsu J C, Viswanathan V, Wang C, Wool P. Anomalies in Chinese A-Shares[J]. *Social Science Electronic Publishing*, 2018.
- [33] 郑振龙, 汤文玉. 波动率风险及风险价格——来自中国 A 股市场的证据[J]. *金融研究*, 2011(4): 143–157.
- [34] 潘莉, 徐建国. A 股个股回报率的惯性与反转[J]. *金融研究*, 2011(1): 149–166.
- [35] 林虎, 孙博, 刘力. 换手率波动、转售期权与股票横截面收益率[J]. *金融研究*, 2013(12): 181–193.
- [36] 郑振龙, 王磊, 王路跖. 特质偏度是否被定价?[J]. *管理科学学报*, 2013, 16(5): 1–12.
- [37] 韦倩. 1996-2009 年上海 A 股市场规模效应研究: 基于收益的可预测性[J]. *管理工程学报*, 2014(01): 74-80+93.
- [38] 陈国进, 许秀, 赵向琴. 罕见灾难风险和股市收益--基于我国个股横截面尾部风险的实证研究[J]. *系统工程理论与实践*, 2015, 35(9): 2186–2199.
- [39] 罗进辉, 向元高, 金思静. 中国资本市场低价股的溢价之谜[J]. *金融研究*, 2017, 01: 191–206.
- [40] 李斌, 冯佳捷. 中国股票市场的质量因子研究[J]. *管理评论*, 2018, forthcoming.
- [41] Jiang F, Tang G, Zhou G. Firm Characteristics and Chinese Stocks[J]. *Journal of Management Science and Engineering*, 2019, 3(4): 259–284.
- [42] Atsalakis G S, Valavanis K P. Surveying stock market forecasting techniques – Part II: Soft computing methods[J]. *Expert Systems with Applications*, 2009, 36(3, Part 2): 5932–5941.
- [43] Huang C-L, Tsai C-Y. A hybrid SOFM-SVR with a filter-based feature selection for stock

- market forecasting[J]. *Expert Systems with Applications*, 2009, 36(2, Part 1): 1529–1539.
- [44] Lee M-C. Using support vector machine with a hybrid feature selection method to the stock trend prediction[J]. *Expert Systems with Applications*, 2009, 36(8): 10896–10904.
- [45] Tsai C F, Hsiao Y C. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches[J]. *Decision Support Systems*, 2011, 50(1): 258–269.
- [46] 谢合亮, 胡迪. 多因子量化模型在投资组合中的应用——基于 LASSO 与 Elastic Net 的比较研究[J]. *统计与信息论坛*, 2017, 32(10): 36–42.
- [47] 宋鹏, 梁吉业, 钱宇华, 李常洪. 区间数分级决策的特征选择方法研究[J]. *中国管理科学*, 2017, 25(7): 141–152.
- [48] Zhang Y, Wei Y, Zhang Y, Jin D. Forecasting oil price volatility: Forecast combination versus shrinkage method[J]. *Energy Economics*, 2019, 80: 423–433.
- [49] Feng G, Giglio S, Xiu D. Taming the Factor Zoo: A Test of New Factors[J]. Fama-Miller Working Paper; Chicago Booth Research Paper No. 17-04., 2019.
- [50] Messmer M, Audrino F. The (Adaptive) Lasso in the Zoo - Firm Characteristic Selection in the Cross-Section of Expected Returns ([J]. 2017.
- [51] Krauss C, Do X A, Huck N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500[J]. *European Journal of Operational Research*, 2017, 259(2): 689–702.
- [52] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016: 785–794.
- [53] DeMiguel V, Garlappi L, Uppal R. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?[J]. *The Review of Financial Studies*, 2009, 22(5): 1915–1953.

附录 1 因子描述性统计

| | | mean | std | min | 0.25 | median | 0.75 | max |
|------------|-----------|----------|----------|-----------|-----------|-----------|-----------|----------|
| A.交易摩擦因子 | | | | | | | | |
| size | A 股流通市值 | 4.27E+06 | 1.54E+07 | 1.92E+05 | 1.18E+07 | 2.02E+06 | 3.81E+06 | 5.82E+08 |
| size_ia | 行业调整市值 | 0.00 | 1.51E+07 | -2.02E+07 | -2.61E+06 | -1.66E+06 | -2.10E+04 | 5.62E+08 |
| beta | 系统性风险 | 0.96 | 0.27 | -0.04 | 0.82 | 0.97 | 1.10 | 3.04 |
| betasq | 系统性风险的平方 | 1.05 | 2.18 | 0.03 | 0.68 | 0.95 | 1.22 | 48.33 |
| betad | 下行风险 | 0.96 | 0.32 | -0.21 | 0.79 | 0.96 | 1.13 | 3.59 |
| idvol | 异质波动率 | 1.46 | 0.17 | 0.98 | 1.34 | 1.45 | 1.57 | 1.91 |
| vol | 总波动率 | 2.80 | 1.90 | 0.72 | 2.19 | 2.62 | 3.14 | 41.02 |
| idskew | 特定偏态 | -0.95 | -0.95 | -12.84 | -1.16 | -0.77 | -0.39 | 2.77 |
| skew | 总偏态 | 0.34 | 1.23 | -2.16 | -0.06 | 0.16 | 0.39 | 12.28 |
| coskew | 共同偏态 | -0.44 | 5.25 | -16.66 | -2.88 | -0.70 | 1.56 | 60.99 |
| turn | 交易换手率 | 2.47 | 1.39 | 0.23 | 1.56 | 2.17 | 3.03 | 10.75 |
| std_turn | 换手率的波动率 | 1.40 | 1.57 | 0.06 | 0.58 | 0.96 | 1.64 | 18.39 |
| volumed | 交易额 | 7.84E+07 | 9.41E+07 | 6.05E+06 | 3.17E+07 | 5.10E+07 | 8.83E+07 | 1.40E+09 |
| std_dvol | 交易额的波动率 | 3.97E+07 | 6.06E+07 | 1.24E+06 | 1.23E+07 | 2.26E+07 | 4.41E+07 | 1.08E+09 |
| retnmax | 最大日收益率 | 0.06 | 0.11 | 0.01 | 0.04 | 0.05 | 0.07 | 2.40 |
| illq | 非流动性风险 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.19 |
| LM | 标准化的换手率 | 21.00 | 0.00 | 21.00 | 21.00 | 21.00 | 21.00 | 21.09 |
| sharechg | 股本的增长率 | 0.30 | 0.60 | -0.26 | 0.00 | 0.05 | 0.35 | 8.77 |
| age | 公司年龄 | 7.38 | 4.26 | 0.30 | 3.69 | 7.10 | 10.99 | 16.54 |
| aeavol | 收益公告异常交易量 | 0.34 | 7.29 | -0.91 | -0.40 | -0.15 | 0.29 | 295.50 |
| pricedelay | 价格延迟 | 0.80 | 0.18 | 0.09 | 0.70 | 0.85 | 0.94 | 1.00 |

| | | | | | | | | |
|----------------|-----------|----------|----------|-----------|---------|----------|----------|----------|
| B.动量因子 | | | | | | | | |
| mom12 | 12 个月动量 | 1.38 | 0.91 | 0.38 | 0.97 | 1.18 | 1.48 | 13.67 |
| mom6 | 6 个月动量 | 1.15E+03 | 1.01E+03 | 0.47 | 9.99 | 1.13E+03 | 2.27E+03 | 2.29E+03 |
| mom36 | 36 个月动量 | 1.89 | 1.67 | 0.32 | 1.12 | 1.48 | 2.05 | 23.34 |
| momchg | 动量变化 | -0.11 | 0.74 | -9.81 | -0.27 | -0.02 | 0.19 | 2.51 |
| imom | 特定动量 | 0.07 | 2.93 | -25.96 | -1.4 | 0.08 | 1.54 | 26.16 |
| lagretn | 短期反转 | 1.15E+03 | 1.01E+03 | 0.65 | 10.01 | 1.13E+03 | 2.27E+03 | 2.28E+03 |
| C.价值因子 | | | | | | | | |
| BM | 公司账面市值比 | 360.56 | 248.51 | -1.60E+03 | 226.32 | 332.64 | 469.26 | 2.40E+03 |
| BM_ia | 行业调整账面市值比 | 0.00 | 239.87 | -1.94E+03 | -125.09 | -20.86 | 107.18 | 2.05E+03 |
| AM | 总资产市值比 | 2.04E+03 | 2.48E+03 | 74.13 | 931.64 | 1.45E+03 | 2.29E+03 | 4.91E+04 |
| LEV | 总负债市值比 | 1.10E+03 | 1.78E+03 | 6.55 | 320.74 | 637.06 | 1236.76 | 3.71E+04 |
| EP | 收益价格比 | 37.24 | 140.66 | -2.04E+03 | 14.64 | 37.53 | 68.91 | 1.37E+03 |
| CFP | 现金流价格比 | 13.39 | 197.96 | -1.63E+03 | -52.93 | -3.07 | 49.9 | 1.98E+03 |
| CFP_ia | 行业调整 CFP | 0.00 | 196.29 | -1.63E+03 | -69.26 | -15.86 | 41.24 | 1.95E+03 |
| OCFP | 营业现金流价格比 | 49.41 | 208.38 | -1.90E+03 | -12.51 | 31.83 | 93.1 | 2.96E+03 |
| DP | 股利价格比 | 8.29 | 23.26 | -1.59 | 0.04 | 1.13 | 7.62 | 401.34 |
| SP | 营业收入价格比 | 1.41E+03 | 2.93E+03 | -8.49 | 376.44 | 757.38 | 1.51E+03 | 6.08E+04 |
| D.成长因子 | | | | | | | | |
| AG | 总资产增长率 | 1.01 | 30.32 | -0.86 | 0.00 | 0.10 | 0.25 | 1218.82 |
| LG | 总负债增长率 | 24.28 | 1.16E+03 | -30.65 | -0.04 | 0.13 | 0.39 | 5.68E+04 |
| BVEG | 净资产增长率 | -1.33 | 94.54 | -4.21E+03 | 0.01 | 0.07 | 0.18 | 392.64 |
| SG | 营业收入增长率 | 3.88 | 153.78 | -191.01 | -0.05 | 0.12 | 0.32 | 6.86E+03 |
| PMG | 营业利润增长率 | 1.63 | 102.56 | -1.16E+03 | -0.47 | -0.01 | 0.39 | 2576.8 |
| INVG | 存货增长率 | 18.4 | 796.1 | -1.41 | -0.08 | 0.11 | 0.38 | 3.63E+04 |
| INVchg | 存货变化 | 0.02 | 0.09 | -0.68 | -0.01 | 0.01 | 0.04 | 0.93 |

| | | | | | | | | |
|---------------------|---------------|-----------|----------|-----------|-----------|-----------|----------|----------|
| SgINVg | 营业收入与存货增长率的差 | -14.64 | 801.38 | -3.42E+04 | -0.27 | -0.01 | 0.23 | 3348.05 |
| TAXchg | 税收增长率 | 8.43 | 496.61 | -3.50E+03 | -0.59 | -0.01 | 0.60 | 1.88E+04 |
| ACC | 应记项目 | -0.17 | 5.72 | -251.44 | -0.05 | 0.02 | 0.07 | 43.77 |
| absacc | 应计项目绝对值 | 0.33 | 5.69 | 0.00 | 0.03 | 0.06 | 0.13 | 257.91 |
| stdacc | 应计项目波动率 | 0.23 | 4.05 | 0.00 | 0.02 | 0.04 | 0.09 | 181.88 |
| ACCP | 应计项目变化 | -9.48E+03 | 3.78E+05 | -1.45E+07 | -0.80 | 0.23 | 1.84 | 1.83E+05 |
| cinvest | 公司投资 | 27.83 | 977.98 | -96.31 | 0.31 | 0.62 | 1.17 | 3.62E+04 |
| depr | 折旧率 | 1.71 | 67.55 | -0.08 | 0.05 | 0.06 | 0.08 | 2.97E+03 |
| pchdepr | 折旧变动百分比 | 1.94 | 28.23 | -36.87 | 1.09 | 1.15 | 1.25 | 1.21E+03 |
| egr | 股东权益变化 | 1.41 | 14.76 | -203.83 | 1.01 | 1.07 | 1.18 | 441.38 |
| fgr5yr | 预期 5 年每股收益增长 | 24.55 | 85.52 | -3.36 | 1.71 | 7.01 | 23.34 | 3046.45 |
| grCAPX | 资本支出变化 | 27.81 | 1.02E+03 | -1.36 | 0.59 | 1.05 | 1.90 | 4.77E+04 |
| pchcapx_ia | 行业调整资本支出变动 | 0.00 | 996.84 | -644.31 | -3.36 | -2.52 | -1.04 | 4.65E+04 |
| grltnoa | 长期净经营资产增长 | 1.16 | 14.15 | -238.84 | 0.96 | 1.04 | 1.14 | 437.19 |
| invest | 资本支出和存货变化 | 35.51 | 1.19E+03 | -2.34 | 1.67 | 2.27 | 3.29 | 5.40E+04 |
| pchsale_invt | 销售增长减存货增长 | -2.04 | 329.87 | -7.44E+03 | -0.29 | 0.02 | 0.30 | 6.85E+03 |
| pchsale_rect | 销售增长减应收账款增长 | -6.39 | 564.01 | -2.23E+04 | -0.28 | 0.06 | 0.42 | 6.76E+03 |
| pchsale_xsga | 销售增长减 SG&A 增长 | 3.3 | 156.87 | -321.89 | -0.45 | -0.21 | 0.02 | 6.86E+03 |
| realestate | 不动产持有量 | 1.94E+09 | 1.21E+10 | -4.84E+06 | 1.50E+08 | 3.56E+08 | 9.23E+08 | 3.60E+11 |
| sgr | 销售增长 | 4.88 | 153.59 | -190.02 | 0.96 | 1.13 | 1.33 | 6864.07 |
| NOA | 净经营资产 | 3.43E+09 | 1.95E+10 | -7.18E+09 | 5.11E+08 | 1.05E+09 | 2.27E+09 | 7.32E+11 |
| chnanalyst | 分析师人数变化 | 0.12 | 7.09 | -60.67 | -0.62 | 0.00 | 0.64 | 65.33 |
| nanalyst | 涉及股票的分析师人数 | 2.31 | 6.1 | 0.00 | 0.00 | 0.00 | 1.34 | 76.45 |
| hire | 雇员增长率 | 3154.36 | 1.40E+04 | 20.00 | 689.22 | 1.36E+03 | 2.64E+03 | 4.05E+05 |
| chempia | 行业调整后雇员人数变动 | 0.00 | 1.38E+04 | -13951.72 | -2.06E+03 | -1.18E+03 | -11.44 | 3.91E+05 |
| RD | 研发成本 | 1.80E+08 | 1.04E+09 | -2.22E+07 | 2.93E+07 | 5.66E+07 | 1.21E+08 | 3.39E+10 |

| | | | | | | | | |
|-------------------|--------------|----------|----------|-----------|-----------|----------|----------|----------|
| rd_mve | 研发支出市值比 | 61.83 | 93.01 | -30.75 | 21.71 | 38.51 | 68.56 | 1587.52 |
| RDsale | 研发成本收入比 | 3.83 | 139.17 | -172.81 | 0.04 | 0.06 | 0.1 | 4981.25 |
| E.盈利因子 | | | | | | | | |
| ROE | 净资产收益率 | 2.06E+04 | 3.59E+05 | -4.37E+05 | 0.02 | 0.06 | 0.1 | 1.17E+07 |
| ROA | 总资产收益率 | 1.97E+04 | 3.56E+05 | -1.13E+06 | 0.01 | 0.03 | 0.06 | 1.08E+07 |
| CT | 资本换手率 | 3.77E+09 | 2.84E+10 | -2.54E+06 | 3.14E+08 | 7.48E+08 | 1.94E+09 | 9.75E+11 |
| PA | 利润资产比率 | 2.24E+04 | 5.13E+05 | -7.77E+05 | 0.02 | 0.04 | 0.07 | 1.74E+07 |
| cashpr | 现金生产力 | -33.97 | 396.03 | -12783.3 | -14.14 | -7.81 | -4.67 | 40.5 |
| cash | 现金净资产比 | 8.94E+08 | 3.31E+09 | 9.72E+04 | 1.19E+08 | 2.78E+08 | 6.34E+08 | 8.13E+10 |
| operprof | 营业利润率 | 0.36 | 1.82 | -4.57 | 0.02 | 0.12 | 0.29 | 38.36 |
| pchgmsale | 毛利率变动%-销售变动% | -3.73 | 141.42 | -4.95E+03 | -0.42 | -0.15 | 0.05 | 470.26 |
| ATO | 总资产周转率 | 1.15 | 23.24 | -541.56 | 0.39 | 0.66 | 1.12 | 531.2 |
| chfeps | 预期每股收益的变化 | 0.02 | 0.36 | -3.87 | 0.00 | 0.00 | 0.00 | 5.18 |
| nincr | 收入增加期数 | 1.15 | 1.87 | 0.00 | 0.00 | 0.18 | 1.47 | 6.01 |
| roic | 投入资本回报 | 0.48 | 23.98 | -55.92 | 0.02 | 0.05 | 0.09 | 1.11E+03 |
| rusp | 意外收入 | 66.6 | 3.28E+03 | -57531.02 | -8.73 | 64.25 | 212.68 | 2.00E+04 |
| sfe | 收益预测 | 3.04E+09 | 5.46E+11 | -1.28E+13 | -6.89E+08 | 0.00 | 1.32E+09 | 1.27E+13 |
| F.财务流动性因子 | | | | | | | | |
| CR | 流动比率 | 2.37 | 6.82 | -2.38 | 1.04 | 1.48 | 2.34 | 250.49 |
| QR | 速动比率 | 1.89 | 6.31 | -1.20 | 0.69 | 1.06 | 1.80 | 229.63 |
| CFdebt | 现金流负债比 | 0.16 | 0.47 | -4.98 | 0.02 | 0.07 | 0.19 | 8.19 |
| salecash | 营业收入现金比 | 14.42 | 175.16 | -0.82 | 1.68 | 3.44 | 6.8 | 4.22E+03 |
| saleinv | 营业收入存货比 | 108.97 | 3.28E+03 | -7.43 | 1.93 | 3.64 | 6.84 | 1.46E+05 |
| CRG | 流动比率增长率 | 0.16 | 3.83 | -3.36 | -0.2 | -0.04 | 0.12 | 124.09 |
| QRG | 速动比率增长率 | 0.14 | 2.86 | -2.43 | -0.25 | -0.06 | 0.15 | 100.29 |
| pchsaleinv | 营业收入存货比增长率 | 5.50E+03 | 2.52E+05 | -1.75E+03 | -0.21 | 0.00 | 0.24 | 1.16E+07 |

| | | | | | | | | |
|----------------|-----------|----------|----------|----------|----------|----------|----------|----------|
| salerec | 营业收入应收账款比 | 1.34E+03 | 6.08E+04 | -31.61 | 1.99 | 3.86 | 9.45 | 2.94E+06 |
| tang | 偿债能力/总资产 | 1.93E+09 | 7.90E+09 | 2.67E+06 | 3.09E+08 | 6.13E+08 | 1.32E+09 | 2.00E+11 |

资料来源：作者计算

附录 2 单因子特征筛选结果

| | factor | 因子名称 | 因子类别 | 年化收益率 | t-statistics | 因子打分 |
|----|------------|--------------|---------|--------|--------------|---------|
| 1 | std_dvol | 交易额的波动率 | 交易摩擦因子 | 18.18% | 8.7984 | 100.00% |
| 2 | LM | 标准化的换手率 | 交易摩擦因子 | 17.28% | 4.1900 | 86.00% |
| 3 | std_turn | 换手率的波动率 | 交易摩擦因子 | 11.62% | 6.5113 | 74.01% |
| 4 | volumed | 交易额 | 交易摩擦因子 | 11.65% | 5.9414 | 67.53% |
| 5 | rd_mve | 研发支出市值比 | 成长因子 | 13.93% | 5.4255 | 61.66% |
| 6 | chfeps | 预期每股收益的变化 | 盈利因子 | 8.44% | 4.7377 | 53.85% |
| 7 | retnmax | 最大日收益率 | 交易摩擦因子 | 5.99% | 4.5582 | 51.81% |
| 8 | illq | 非流动性风险 | 交易摩擦因子 | 9.21% | 4.2123 | 47.88% |
| 9 | lagretn | 短期反转 | 动量因子 | 15.17% | 3.8841 | 44.15% |
| 10 | size | A 股流通市值 | 交易摩擦因子 | 21.26% | 3.8701 | 43.99% |
| 11 | skew | 总偏态 | 交易摩擦因子 | 4.54% | 3.7740 | 42.89% |
| 12 | aeavol | 收益公告异常交易量 | 交易摩擦因子 | 6.67% | 3.6105 | 41.04% |
| 13 | idvol | 异质波动率 | 交易摩擦因子 | 4.74% | 3.2926 | 37.42% |
| 14 | SgINVg | 营业收入与存货增长率的差 | 成长因子 | 3.63% | 2.9240 | 33.23% |
| 15 | size_ia | 行业调整市值 | 交易摩擦因子 | 12.13% | 2.9110 | 33.09% |
| 16 | momchg | 动量变化 | 动量因子 | 9.01% | 2.8739 | 32.66% |
| 17 | pchsaleinv | 营业收入存货比增长率 | 财务流动性因子 | 3.70% | 2.8627 | 32.54% |
| 18 | mom36 | 36 个月动量 | 动量因子 | 9.07% | 2.5630 | 29.13% |
| 19 | vol | 总波动率 | 交易摩擦因子 | 5.00% | 2.3778 | 27.03% |
| 20 | CRG | 流动比率增长率 | 财务流动性因子 | 2.91% | 2.1851 | 24.83% |
| 21 | depr | 折旧率 | 成长因子 | 8.09% | 2.1071 | 23.95% |
| 22 | SP | 营业收入价格比 | 价值因子 | 5.23% | 2.0128 | 22.88% |

资料来源：作者计算

附录 3 OLS 特征筛选结果

| | factor | 因子名称 | 因子类别 | 平均月度收益差值 | 因子打分 |
|----|---------------|--------------|---------|----------|--------|
| 1 | std_turn | 换手率的波动率 | 交易摩擦因子 | 0.30% | 100% |
| 2 | aeavol | 收益公告异常交易量 | 交易摩擦因子 | 0.20% | 67.97% |
| 3 | lagretn | 短期反转 | 动量因子 | 0.18% | 60.75% |
| 4 | idvol | 异质波动率 | 交易摩擦因子 | 0.12% | 39.54% |
| 5 | INVchg | 存货变化 | 成长因子 | 0.09% | 30.29% |
| 6 | betad | 下行风险 | 交易摩擦因子 | 0.08% | 27.47% |
| 7 | egr | 股东权益变化 | 成长因子 | 0.08% | 26.29% |
| 8 | ATO | 总资产周转率 | 盈利因子 | 0.07% | 22.67% |
| 9 | cashpr | 现金生产力 | 盈利因子 | 0.07% | 22.12% |
| 10 | ROA | 总资产收益率 | 盈利因子 | 0.06% | 21.47% |
| 11 | mom36 | 36 个月动量 | 动量因子 | 0.06% | 18.42% |
| 12 | PA | 利润资产比率 | 盈利因子 | 0.05% | 17.48% |
| 13 | rusp | 意外收入 | 盈利因子 | 0.05% | 16.78% |
| 14 | ACCP | 应计项目变化 | 成长因子 | 0.05% | 16.49% |
| 15 | sharechg | 股本的增长率 | 交易摩擦因子 | 0.05% | 15.97% |
| 16 | fgr5yr | 预期 5 年每股收益增长 | 成长因子 | 0.05% | 15.79% |
| 17 | nincr | 收入增加期数 | 盈利因子 | 0.04% | 14.83% |
| 18 | PMG | 营业利润增长率 | 成长因子 | 0.04% | 14.34% |
| 19 | pchgm_pchsale | 毛利率变动%-销售变动% | 盈利因子 | 0.04% | 14.05% |
| 20 | stdacc | 应计项目波动率 | 成长因子 | 0.04% | 13.35% |
| 21 | TAXchg | 税收增长率 | 成长因子 | 0.04% | 12.86% |
| 22 | CRG | 流动比率增长率 | 财务流动性因子 | 0.04% | 12.85% |
| 23 | CFdebt | 现金流负债比 | 财务流动性因子 | 0.04% | 12.45% |
| 24 | DP | 股利价格比 | 价值因子 | 0.04% | 12.04% |
| 25 | CT | 资本换手率 | 盈利因子 | 0.04% | 11.71% |
| 26 | AG | 总资产增长率 | 成长因子 | 0.04% | 11.70% |
| 27 | tang | 偿债能力/总资产 | 财务流动性因子 | 0.03% | 11.51% |
| 28 | LG | 总负债增长率 | 成长因子 | 0.03% | 10.24% |

资料来源：作者计算

附录 4 GBDT 特征筛选结果

| | factor | 因子名称 | 因子类别 | IMP | 因子打分 |
|----|-----------------|---------------|---------|--------|---------|
| 1 | retvol | 总波动率 | 交易摩擦因子 | 0.1765 | 100.00% |
| 2 | mom12 | 12 个月动量 | 动量因子 | 0.1286 | 72.87% |
| 3 | pchdepr | 折旧变动百分比 | 成长因子 | 0.1019 | 57.69% |
| 4 | lagretn | 短期反转 | 动量因子 | 0.0490 | 27.74% |
| 5 | depr | 折旧率 | 成长因子 | 0.0473 | 26.77% |
| 6 | mom_36 | 36 个月动量 | 动量因子 | 0.0463 | 26.21% |
| 7 | aeavol | 收益公告异常交易量 | 交易摩擦因子 | 0.0442 | 25.06% |
| 8 | idvol | 异质波动率 | 交易摩擦因子 | 0.0407 | 23.06% |
| 9 | mom6 | 6 个月动量 | 动量因子 | 0.0333 | 18.84% |
| 10 | stdvold | 交易额的波动率 | 交易摩擦因子 | 0.0323 | 18.30% |
| 11 | momchg | 动量变化 | 动量因子 | 0.0306 | 17.33% |
| 12 | retmax | 最大日收益率 | 交易摩擦因子 | 0.0283 | 16.03% |
| 13 | pchsale_pchxsga | 销售增长减 SG&A 增长 | 成长因子 | 0.0254 | 14.41% |
| 14 | size | A 股流通市值 | 交易摩擦因子 | 0.0243 | 13.79% |
| 15 | BM | 公司账面市值比 | 价值因子 | 0.0229 | 12.99% |
| 16 | LM | 标准化的换手率 | 交易摩擦因子 | 0.0187 | 10.60% |
| 17 | illq | 非流动性风险 | 交易摩擦因子 | 0.0156 | 8.84% |
| 18 | egr | 股东权益变化 | 成长因子 | 0.0134 | 7.58% |
| 19 | turnsd | 换手率的波动率 | 交易摩擦因子 | 0.0095 | 5.37% |
| 20 | imom | 特定动量 | 动量因子 | 0.0088 | 4.97% |
| 21 | stdacc | 应计项目波动率 | 成长因子 | 0.0086 | 4.86% |
| 22 | BVEG | 净资产增长率 | 成长因子 | 0.0082 | 4.62% |
| 23 | vold | 交易额 | 交易摩擦因子 | 0.0079 | 4.49% |
| 24 | size_ia | 行业调整市值 | 交易摩擦因子 | 0.0068 | 3.83% |
| 25 | age | 公司年龄 | 交易摩擦因子 | 0.0058 | 3.27% |
| 26 | PMG | 营业利润增长率 | 成长因子 | 0.0047 | 2.68% |
| 27 | cinvest | 公司投资 | 成长因子 | 0.0046 | 2.63% |
| 28 | tang | 偿债能力/总资产 | 财务流动性因子 | 0.0041 | 2.31% |

资料来源：作者计算

附录5 Lasso 特征筛选结果

| | factor | 因子名称 | 因子类别 | coef | 因子打分 |
|----|------------|------------|---------|---------|---------|
| 1 | BM | 公司账面市值比 | 价值因子 | 0.0180 | 100.00% |
| 2 | BM_ia | 行业调整账面市值比 | 价值因子 | -0.0121 | 67.17% |
| 3 | retvol | 总波动率 | 交易摩擦因子 | 0.0099 | 54.80% |
| 4 | stdvold | 交易额的波动率 | 交易摩擦因子 | -0.0076 | 42.18% |
| 5 | beta | 系统性风险 | 交易摩擦因子 | -0.0060 | 33.52% |
| 6 | retmax | 最大日收益率 | 交易摩擦因子 | -0.0058 | 32.41% |
| 7 | mom12 | 12 个月动量 | 动量因子 | 0.0043 | 23.69% |
| 8 | imom | 特定动量 | 动量因子 | 0.0033 | 18.51% |
| 9 | turn | 交易换手率 | 交易摩擦因子 | 0.0032 | 17.83% |
| 10 | vold | 交易额 | 交易摩擦因子 | -0.0027 | 15.00% |
| 11 | mom_36 | 36 个月动量 | 动量因子 | -0.0026 | 14.40% |
| 12 | momchg | 动量变化 | 动量因子 | -0.0025 | 14.05% |
| 13 | turnsd | 换手率的波动率 | 交易摩擦因子 | -0.0024 | 13.06% |
| 14 | idvol | 异质波动率 | 交易摩擦因子 | 0.0019 | 10.47% |
| 15 | nincr | 收入增加期数 | 盈利因子 | -0.0018 | 9.79% |
| 16 | age | 公司年龄 | 交易摩擦因子 | -0.0018 | 9.78% |
| 17 | rd_mve | 现金生产力 | 盈利因子 | 0.0015 | 8.36% |
| 18 | pricedelay | 价格延迟 | 交易摩擦因子 | -0.0013 | 7.41% |
| 19 | SP | 营业收入价格比 | 价值因子 | 0.0010 | 5.35% |
| 20 | nanalyst | 涉及股票的分析师人数 | 成长因子 | 0.0009 | 4.99% |
| 21 | size_ia | 行业调整市值 | 交易摩擦因子 | 0.0008 | 4.20% |
| 22 | CFP_ia | 行业调整 CFP | 价值因子 | 0.0006 | 3.12% |
| 23 | aeavol | 收益公告异常交易量 | 交易摩擦因子 | 0.0005 | 2.97% |
| 24 | NOA | 净经营资产 | 成长因子 | -0.0005 | 2.49% |
| 25 | saleinv | 营业收入存货比 | 财务流动性因子 | -0.0004 | 2.11% |
| 26 | DP | 股利价格比 | 价值因子 | -0.0003 | 1.61% |
| 27 | operprof | 营业利润率 | 盈利因子 | -0.0003 | 1.44% |
| 28 | OCFP | 营业现金流价格比 | 价值因子 | 0.0002 | 0.86% |
| 29 | skewness | 总偏态 | 交易摩擦因子 | 0.0001 | 0.81% |
| 30 | QRG | 速动比率增长率 | 财务流动性因子 | -0.0001 | 0.51% |
| 31 | EP | 收益价格比 | 价值因子 | -0.0001 | 0.48% |
| 32 | CFdebt | 现金流负债比 | 财务流动性因子 | 0.0000 | 0.15% |

资料来源：作者计算

附录6 DFN 特征筛选结果

| | factor | 因子名称 | 因子类别 | 平均月度收益差值 | 因子打分 |
|----|------------|------------|---------|----------|---------|
| 1 | vol | 总波动率 | 交易摩擦因子 | 0.39% | 100.00% |
| 2 | aeavol | 收益公告异常交易量 | 交易摩擦因子 | 0.27% | 69.16% |
| 3 | QR | 速动比率 | 财务流动性因子 | 0.27% | 68.49% |
| 4 | retnmax | 最大日收益率 | 交易摩擦因子 | 0.23% | 60.38% |
| 5 | egr | 股东权益变化 | 成长因子 | 0.23% | 59.82% |
| 6 | sgr | 销售增长 | 成长因子 | 0.23% | 59.00% |
| 7 | TAXchg | 税收增长率 | 成长因子 | 0.22% | 56.84% |
| 8 | CRG | 流动比率增长率 | 财务流动性因子 | 0.20% | 51.50% |
| 9 | AM | 总资产市值比 | 价值因子 | 0.20% | 50.43% |
| 10 | pchdepr | 折旧变动百分比 | 成长因子 | 0.19% | 49.06% |
| 11 | RDsale | 研发成本收入比 | 成长因子 | 0.19% | 48.01% |
| 12 | pchsaleinv | 营业收入存货比增长率 | 财务流动性因子 | 0.19% | 47.73% |
| 13 | INVG | 存货增长率 | 成长因子 | 0.18% | 47.36% |
| 14 | volumed | 交易额 | 交易摩擦因子 | 0.18% | 45.83% |
| 15 | OCFP | 营业现金流价格比 | 价值因子 | 0.17% | 44.91% |
| 16 | cash | 现金净资产比 | 盈利因子 | 0.17% | 42.84% |
| 17 | salecash | 营业收入现金比 | 财务流动性因子 | 0.16% | 40.25% |
| 18 | rd_mve | 研发支出市值比 | 成长因子 | 0.16% | 39.87% |
| 19 | std_turn | 换手率的波动率 | 交易摩擦因子 | 0.15% | 39.70% |
| 20 | cinvest | 公司投资 | 成长因子 | 0.15% | 39.21% |
| 21 | QRG | 速动比率增长率 | 财务流动性因子 | 0.15% | 38.85% |
| 22 | RD | 研发成本 | 成长因子 | 0.15% | 37.54% |
| 23 | chnanalyst | 分析师人数变化 | 成长因子 | 0.15% | 37.47% |
| 24 | momchg | 动量变化 | 动量因子 | 0.14% | 36.97% |
| 25 | tang | 偿债能力/总资产 | 财务流动性因子 | 0.14% | 35.55% |

资料来源：作者计算