

# 电商行业股票收益率预测与量化投资策略

张 余

贵州大学经济学院, 贵州 贵阳

收稿日期: 2025年7月7日; 录用日期: 2025年7月21日; 发布日期: 2025年8月27日

## 摘 要

电子商务在促进国民经济增长方面发挥着越来越重要的作用。基于机器学习算法的量化投资策略在股票价格预测方面运用越来越广泛。本文运用机器学习算法, 探索量化投资策略运用于电商概念股票收益率预测中的效果。基于2009年1月到2020年12月166只电商概念股票的4项异象因子, 本文采用监督机器学习Lasso回归算法, 构建股票收益率预测模型及投资策略。实证结果显示, 机器学习驱动的量化投资模型能够在电商概念股票中取得一定的投资收益, 但投资收益率较低; 在异象因子重要性的检验中, 股本增长率对提升模型预测能力最为重要; 在Lasso回归、岭回归和弹性网络回归三种算法的比较中, 使用弹性网络回归算法的预测模型表现最好。将机器学习引入股票板块收益分析中, 对众多股票投资者做出合理投资决策具有参考意义。

## 关键词

电子商务, 量化投资, 机器学习

# Stock Return Forecasts and Quantitative Investment Strategies for the E-Commerce Industry

Yu Zhang

School of Economics, Guizhou University, Guiyang Guizhou

Received: Jul. 7<sup>th</sup>, 2025; accepted: Jul. 21<sup>st</sup>, 2025; published: Aug. 27<sup>th</sup>, 2025

## Abstract

E-commerce is playing an increasingly important role in promoting national economic growth. Quantitative investment strategies based on machine learning algorithms are increasingly used in stock price prediction. In this paper, we use machine learning algorithms to explore the effectiveness of

文章引用: 张余. 电商行业股票收益率预测与量化投资策略[J]. 电子商务评论, 2025, 14(8): 2664-2672.

DOI: 10.12677/ecl.2025.1482826

quantitative investment strategies applied to e-commerce concept stock return forecasting. Based on the seven anomaly factors of 166 e-commerce concept stocks from January 2009 to December 2020, this paper employs supervised machine learning Lasso regression algorithm to construct stock return prediction models and investment strategies. The empirical results show that the machine learning-driven quantitative investment model is able to achieve some investment returns in e-commerce concept stocks, but the investment return is low; In the test of the importance of the dissimilarity factor, equity growth rate is the most important in enhancing the predictive power of the model; In the comparison of the three algorithms, Lasso regression, ridge regression and elastic network regression, the predictive model using the elastic network regression algorithm performed the best. Introducing machine learning into stock sector return analysis is informative for many stock investors to make sound investment decisions.

## Keywords

E-Commerce, Quantitative Investment, Machine Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

电商是企业相互之间或企业与顾客之间进行的网络化商务活动。信息网络化带来的电商行业崛起，在提高国民消费水平、开辟创造就业岗位、促进经济高质量发展等方面发挥了重要作用。不仅如此，随着数字技术广泛地渗入电商行业，更多企业跨越国门，推动跨境电商蓬勃发展，进一步促进国内经济发展。自 2015 年起，截至 2024 年，商务部先后设立了 165 个跨境电商综合试验区，旨在降低跨境贸易成本，使贸易向着更自由、更便利的方向发展，让更多企业利用跨境电商参与国际贸易。从艾瑞咨询发布的行业数据报告来看，截至 2023 年，中国跨境电商市场规模达 1.4 万亿美元，较 2017 年增长约 2 倍，实现 18.5% 的年均复合增长率，高于同期中国 GDP 年均复合增长率 6.2%。并且其预计至 2028 年，中国跨境电商市场将以 11.3% 的年均复合增长率稳定增长，达到 2.4 万亿美元。由此可见，中国跨境电商市场仍处于快速发展时期。

为了培植这一促进国民经济发展的新的增长点，早在“十一五”期间，国家就专门制定了电子商务发展“十一五”规划，强调电子商务与产业发展深度融合已经成为经济竞争新常态。“十四五”期间，习近平总书记多次就电子商务的发展做出重要指示，对发展农村电商、跨境电商、丝路电商等提出要求，明确指出电子商务大有可为。电商经济活动深入家家户户，那么上市电商企业股票是否具有投资价值，能否实现投资者资产的保值增值成为本文需要验证的一个重要问题。量化投资旨在通过数学模型和学习算法等实现投资超额收益[1]。1952 年 Markowitz 提出的均值 - 方差理论奠定了现代量化投资发展的基础[2]。威廉·夏普等人在此基础上提出著名的资本资产定价模型，即 CAPM 模型： $E(r_i) = r_f + \beta_{iM}(E(r_M) - r_f)$ 。其中， $E(r_i)$ 、 $r_f$ 、 $E(r_M)$  分别表示资产  $i$  的期望收益率、无风险收益率、市场组合  $M$  的期望收益率；系统性风险系数  $\beta_{iM} = Cov(r_i, r_M) / Var(r_M)$ ，反应了资产  $i$  的收益对市场组合  $M$  收益的敏感程度。随着时代的发展，衍生品定价模型、风险管理模型(VAR)、多因子模型等逐渐丰富量化投资理论。21 世纪以来，伴随着人工智能的进步，作为其代表性技术，机器学习被广泛运用于预测资产收益。机器学习的研究方法包括监督学习、半监督学习和无监督学习等。监督学习要求数据样本标签，例如股票收益率、股票价格升跌等。Lasso 回归作为一种监督学习线性回归算法，通过在损失函数中加入 L1 正则化项来实现系数压缩，在高维

股票特征中筛选变量,提高模型的可解释性。

因此本文基于 2009 年 1 月 1 日到 2020 年 12 月 31 日,上市电商企业股票异象因子和股票收益率数据,选用(监督)机器学习算法 Lasso 回归来预测股票收益率,并据此做出买卖股票策略,最后根据期末的累计收益率指标来评价预测模型的有效性。实证结果表明,对于电商概念股票,基于机器学习算法的量化投资模型能够为投资者带来一定的收益,但收益率不高。同时,使用不同机器学习算法构建的预测模型能为投资者带来不同的收益,在 Lasso 回归算法、岭回归算法和弹性网络回归算法中,弹性回归网络算法的表现最好。本文的研究丰富了分板块股票研究的方法,对投资者合理投资决策具有现实意义。

## 2. 文献综述

### 2.1. 电商行业发展趋势

信息技术的高速发展为网络贸易创建了开放有序的交易平台,提供了顺畅而便捷的网络环境,孕育了一大批敢于尝试的消费者,使得电子商务更具系统性和专业性。随着中国信息化的建设,国内电商行业自 20 世纪 90 年代末开始迅速发展,其如火如荼的发展态势吸引了例如 ebay 公司、亚马逊等国际投资者,就此这一新兴产业在国内生根发芽[3]。在我国加快构建“以国内大循环为主体,国内国际双循环相互促进”的新发展格局背景下,作为数字经济时代的一种贸易新业态,跨境电商发挥着不可或缺的作用。早在 2015 年,商务部便开始布局设立跨境电商综合试验区,实证表明试验区可以通过优化营商环境、促进产业结构升级进而促进外贸高质量发展[4],而跨境电商产业政策能够显著促进试点城市的出口开放度[5]。跨境电商覆盖范围广,交易品种多,速度快、效率高的特点能够高效满足全球消费者的需求,促进各国对外贸易的发展[6]。跨境电商作为电子商务的重要板块,在政策的大力支持下,其发展势头迅猛。中国许多电商企业投入重金加快构建海外布局,抢占海外市场份额的先机。例如江苏徐工电子商务股份有限公司通过构建以工程机械为核心的机电产品跨境电商平台和海外仓网络,促进了业务创新和国际竞争力。企业在跨境电商赛道中的表现反映出电商行业的高增长性和激烈的行业竞争格局,目前鲜少有文献从能够体现电商行业上述特征的异象因子的视角出发,探究何种异象因子能成为预测电商行业股票收益表现的重要因素。对于上市电商企业的研究主要集中在绩效评价方面,例如梁娅青(2024)以 48 家上市跨境电商企业为研究样本,研究证明了激进式创新和渐进式创新都对企业绩效具有促进作用[7]。常梅等(2022)基于 BP-SBM 模型,评价了中国 47 家上市电商企业在 2017~2019 年的综合绩效、生产运营绩效和资本运作绩效[8]。目前还没有研究从量化投资策略的视角,基于中国电商行业股票特征,利用机器学习模型预测结果进行投资决策,探究中国电商行业股票的投资收益状况。

### 2.2. 股票价格预测模型

股票作为人们资产配置的一项重要选择,为了实现资产的保值增值,能否准确预测股票价格的波动从而获取超额收益成为人们关注的焦点。Tim Bollerslev 在 ARCH 模型的基础上提出经典的预测时间序列数据的 GARCH 模型,假设时间序列的条件方差依赖于过去的方差和残差平方和,能够有效地捕捉时间序列中的波动聚集性和波动持久性。陈千里等(2002)在对上证指数的波动性实证研究中就采用了 GARCH 类模型[9]。李胜利(2002)运用 VS-GARCH 模型发现上证综指在空头期存在着明显的波动不对称的反转现象[10]。而近年来,机器学习被广泛运用于股票价格预测模型之中。闫政旭等(2021)在传统的随机森林模型中引入 Pearson 系数,显著提升了股票价格的预测效果[11]。包振山等(2020)基于神经网络的机器学习算法,加入遗传算法改进调参模型,提升了股票价格涨跌的预测效果[12]。张晨希等(2006)基于支持向量机对上市公司股票走势进行预测,预测精度明显优于使用传统神经网络的测试结果[13]。李斌等(2019)基于 96 项异象因子,系统地探究了 12 种机器学习算法在预测股票收益上的实证绩效和不同异象因子在

模型中的预测能力[14]。任君等(2019)将支持向量机、神经网络分别与 Lasso 方法相结合,通过这两种算法对沪深 300 中所有的成分股进行涨跌预测进而构建投资组合[15]。李辉等(2022)首先使用随机森林方法处理股票特征冗余的问题,再使用神经网络方法对股票价格进行预测,实证表明该组合模型可以提高股票价格预测的准确性[16]。张鹏等(2024)仍然将随机森林和神经网络这两种机器学习算法结合起来预测股票收盘价,构建多目标投资组合[17]。从现有文献来看,使用创新型的例如基于机器学习算法构建资产收益预测模型的研究多于传统模型,并且使用前者模型所得到的预测效果更好。

### 3. 研究设计

#### 3.1. 模型总体设计

随着互联网技术的不断进步和消费者购物习惯的逐渐改变,电商行业市场规模持续增大,电商企业的营收和利润增长速度较快。例如,拼多多使用创新性的社交电商模式,其用户数量和交易规模迅速增加,股价自上市以来随着业绩的增长不断提高。与传统企业相比,电商企业通常无需大规模的店铺、厂房和生产设备,而是通过与供应商和第三方制造商合作,直接采购成品。在精准的市场需求预测和灵活的供应链管理下,电商企业无需投入大量资金用于库存,保证了资产的高流动性。然而,电商企业对先进的技术平台依赖程度高,例如电商平台系统、大数据分析工具、人工智能推荐系统等。在激烈的竞争环境下,若想不被瓜分市场份额,企业需要不断大量投入资金进行技术研发和创新,以保持竞争优势。对于上市企业而言,他们可能会通过增发股票等方式频繁获取融资,以支撑技术研发和抢占市场份额的巨额支出。电商行业新进入者不断涌现,竞争格局尚未完全稳定,电商企业时刻面临着被市场淘汰的风险。对于风险偏好的投资者而言,他们更青睐具有较高股价波动性的电商企业股票,这为他们提供了更多的交易机会。李斌等(2019)基于 A 股市场数据,通过目前最全的 96 个异象因子库探究股票截面收益的决定因素,根据因子属性将其分为交易摩擦因子、动量因子、价值因子、成长因子、盈利因子、财务流动因子六大类别,并使用 13 种算法筛选异象因子重要性的方式,被算法选中为重要因子超过 5 次的异象因子有 13 项,证明其在中国股票市场中的强预测能力[14]。本文考虑到电商行业所具有的高增长性、轻资产性、技术驱动性和高风险性的特点,因此选取 13 项重要因子中能够分别体现上述特征的股东权益变化(egr)、非流动性风险(illq\_risk)、股本增长率(sharechg)、换手率的波动率(turn\_std)这 4 项异象因子,基于监督式学习模型 Lasso 回归,对电商行业股票月度收益率进行预测。

本文借鉴 DeMiguel (2009)的研究,选取滑动窗口法划分训练集和测试集[18]。具体而言,依次使用样本的第 1~12 个月异象因子-月收益率数据作为训练样本,得到 Lasso 回归模型的拟合参数,并结合第 12 个月异象因子的观测值对第 13 个月收益率进行预测。在第 13 个月月初,根据模型收益率预测值决定投资策略,若预测该月收益率为正,则买入股票;否则,卖空股票。到了第 13 个月月末,若预测正确,则记该月收益率为本月真实收益率的绝对值;否则,记为本文真实收益率的绝对值的负值。在第 14 个月月初,重复上述步骤,直到研究时间段期末。最后通过计算累计收益的方式评价此策略的有效性和这 4 项异象因子的预测能力。

本文采取网格搜索交叉验证的方式选择 Lasso 回归模型参数。在预先设定的参数空间内穷举所有参数组合,对每个参数组合采取交叉验证的方式测算评估指标,最终选择使负均方误差达到最大值的 alpha 为最优参数,并用于测试集预测收益率。为了提高算法效率,本文将在第一个滑动窗口调参得到的最优参数用于此后不同时期模型之中。为了评价这 4 项异象因子在预测电商行业股价模型中的重要性,本文采取逐个去掉因子的方式重新计算期末累计收益。若去掉某项因子所导致的累计收益下降越大,则该项因子越重要。文章所使用的 Lasso 回归算法、异象因子的计算和数据的处理等都通过 Python 语言实现。

### 3.2. 数据来源与样本选取

本文选取同花顺电子商务概念板块的成分股作为研究样本，剔除在 2009 年 1 月 1 日到 2020 年 12 月 31 日期间出现 ST、ST\* 标志的股票，同时考虑到 IPO 抑价效应带来的影响，本文仅选取在 2018 年及以前上市的电商企业，最终获得 166 家上市电商企业在 2009 年到 2020 年的月度观测数据。文章选取的异象因子输入变量有股东权益变化(egr)、非流动性风险(illq\_risk)、股本增长率(sharechg)、换手率的波动率(turn\_std)，具体计算方式如表 1 所示。使用 Lasso 回归的输出变量为股票月收益率，通过计算  $t$  月最后一个交易日的收盘价除以  $t-1$  月最后一个交易日的收盘价再减 1 所得。所有数据均通过 Python 从 Tushare 平台获取。将输入变量在月度上滞后一期与输出变量对齐，得到“上市电商企业-月”数据。同时，不同变量的取值在数量级上的差异显著，因此本文采用“ $(X - \bar{X})/\sigma_X$ ”的方式对变量进行标准化。其中， $\bar{X}$ 、 $\sigma_X$  分别为变量  $X$  的均值和标准差。

**Table 1.** Description of the anomaly factor

**表 1.** 异象因子说明

因子名称	因子符号	计算方式
股东权益变化	egr	股东权益的同比变化[19]
非流动性风险	illq_risk	$t$ 月及其前 11 个月月内交易日日收益率的绝对值除以日交易金额，再取平均值，最后乘以 $10^6$ [20]
股本增长率	sharechg	$t$ 月末流通股本与前 12 个月月末流通股本的比值减 1 [21]
换手率的波动率	turn_std	$t$ 月内交易日换手率的标准差，要求月交易日至少有 10 天[22]

## 4. 实证分析

### 4.1. 策略评价指标

#### 4.1.1. 累计收益率(Cumulative\_Return)

累计收益率用于衡量在特定时间段内投资获得的总收益情况，它能够反应投资从初始价值到投资期末价值的总百分比变化。通过累计收益率可以比较不同股票在研究时间段内，通过模型预测进行的投资决策而实现的收益状况，以评估哪些股票的表现更好。计算公式为：

$$\text{Cumulative\_Return} = \frac{V_f - V_i}{V_i} \quad (1)$$

$V_f$ 、 $V_i$  分别表示股票期末价格和期初价格，本文使用根据模型预测收益率进行的买卖股票决策而获得的月收益率累加求得。从计算结果来看，累计收益率大于 0 的股票数量为 93 只，超过平均水平。但累计收益率均值只有 3.94%，最大值有 70.31%，最小值为亏损 54.15%，说明模型预测能力在股票之间存在明显的异质性，使用各只股票的累计收益率绘制的折线图也能够反映该问题(如图 1 所示)。若将模型单独使用在单只股票上风险较大，扩大股票范围能够减小投资风险获得一定收益。

#### 4.1.2. 胜率(Profit\_Ratio)

本文参考王子豪(2024)的研究，使用胜率作为投资策略评价指标之一[23]。胜率表示根据模型收益率预测结果进行的投资决策中，月收益率大于 0 的投资次数所占的比例，能够反应投资策略盈利能力的强弱。计算公式为：

$$\text{Profit\_Ratio} = \frac{N_{\text{profit}}}{N} \quad (2)$$

$N_{\text{profit}}$ 、 $N$  分别表示月收益率大于 0 的投资次数和总投资次数。从使用 Python 计算的结果来看, 166 只电商企业股票在 2010~2020 年投资胜率大于 1/2 的有 80 只, 略大于平均水平。该指标视角表明, 基于 Lasso 回归模型的投资策略盈利能力一般。

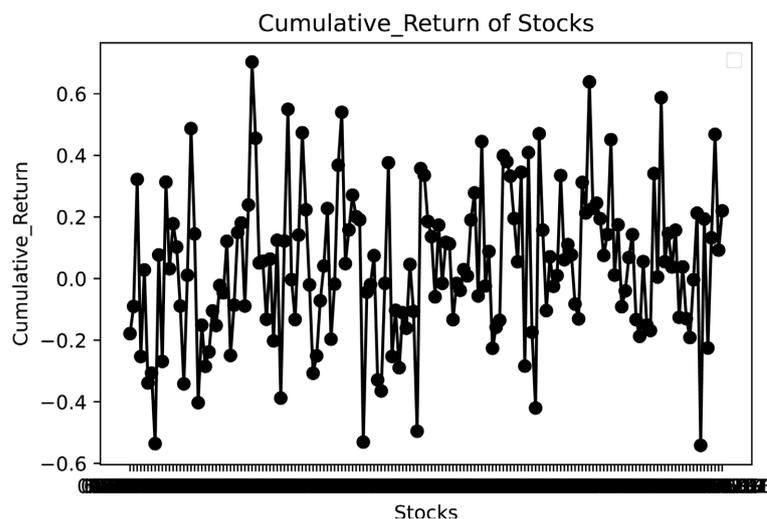


Figure 1. Cumulative Return of stocks  
图 1. 股票的累计收益率

#### 4.1.3. 年化收益率(Annualized\_Return)

年化收益率从年度视角来衡量投资收益水平, 其将投资收益的标准化到一年的时间跨度上, 有利于比较不同投资的收益情况。将年化收益率与其他指标, 例如标准差、夏普比率等相结合, 可以评估投资收益与风险的匹配程度。计算公式为:

$$\text{Annualized\_Return} = (1 + R)^{\frac{1}{n}} - 1 \quad (3)$$

$R$ 、 $n$  分别表示投资获得的总收益和投资年数。从 166 只股票的年化收益率结果来看, 平均年化收益率约为 2.1541, 但年化收益率大于 1 的股票只有 61 只, 且由表 2 可以看出该指标最大值与最小值差距明显, 这表明根据模型预测的收益率进行投资决策时, 收益存在严重不平衡性。

Table 2. Strategy evaluation indicators

表 2. 策略评价指标

指标	数量	均值	中位数	标准差	最大值	最小值
Cumulative_Return	166	0.0394	0.0392	0.2439	0.7031	-0.5415
Profit_Ratio	166	0.4472	0.4925	0.3032	0.8806	0.0000
Annualized_Return	166	2.1541	0.2288	7.7932	85.2301	-0.9947

#### 4.2. 异象因子的重要性

为了探究文章选用的 4 种异象因子对模型预测能力的重要性, 本文将逐一删除其中一项因子, 使用

剩下的 3 项因子对样本股票进行收益预测进而做出投资决策，最后得到期末各只股票的累计收益率。通过比较删除不同异象因子对平均累计收益率损失的大小来判断因子的重要性程度，若删除该项因子后，损失越大，说明该项因子越重要。

从异象因子重要程度检验结果来看(如表 3 所示)，与 3.94%的累计收益率均值相比，在分别去掉非流动性风险(illq\_risk)、股本增长率(sharechg)、换手率的波动率 turn\_std 之后，股票累计收益率均值均有所下降，其中去掉 sharechg 之后均值变为 0.8%，损失最大，其次是 turn\_std，最后是 illq\_risk。国家统计局数据显示，中国电子商务交易额从 2017 年的 29.16 万亿元增长至 2024 年的 46.41 万亿元，年度同比增速均值达到 8.475%，这体现了电商行业巨大的市场潜力和发展空间。为了推动电商行业健康发展，达到有效助力扩消费、助转型、促合作的目的，一系列政策支持措施的出台降低了电商行业准入门槛。一些成熟的例如淘宝、京东等电商平台为商家提供了标准化的基础设施，互联网上存在大量的开源工具和插件也有助于商家快速搭建基础的电商平台架构，使得进入电商行业的技术门槛相对较低。电商企业可以选择从小规模起步，通过整合供应链资源，降低库存积压和资金占用压力，从而实现以较低的资金进入电商行业门槛。一些新兴市场随着互联网普及率的不断提高，电商行业正处于快速发展的阶段，若电商企业能够提前布局这些市场，将有望获得巨大的增长红利。在电商领域，企业拥有海量的客户数据，通过数据分析技术实现精准营销，提供个性化服务，提高客户转化率，帮助企业获得更高的回报。同时，电商企业还可以通过品牌建设实现较高的品牌溢价，通过开展金融服务等增值服务提高企业的利润率。电商行业的低门槛和高收益不断吸引新企业和传统线下企业的进入，企业需要在技术创新、品牌推广、客户服务等方面加大投入，不断提高自身的核心竞争力以保证在激烈的市场竞争格局中不被淘汰。在全球电商市场规模不断扩大的背景下，电商企业为了实现自身的业务拓展和战略布局，往往会通过收购等活动来扩大商业版图，提升风险抵抗能力。因此，无论是为了提升企业的竞争能力，还是满足业务扩张的需求，上市电商企业都可能通过增发、配股等方式进行股权融资，引入新的资金和资源，从而影响企业的股本增长率。为了验证 sharechg 在预测模型中具有显著重要性的结论，本文将训练集时间长度分别调整为 24 个月和 36 个月，在删除 sharechg 的模型之中，模型盈利能力均有所降低，分别降低了 2.78%和 0.42%，表明股本增长率在预测电商行业股票收益率模型中具有重要作用。

**Table 3.** Importance of the anomaly factor

**表 3.** 异象因子重要性

异象因子	egr	illq_risk	sharechg	turn_std
累计收益率均值	0.0420	0.0344	0.0080	0.0258

而在删除股东权益变化(egr)这项因子之后，通过预测进行的投资决策获得的平均累计收益率有所上升，表明在使用 Lasso 回归算法对电商概念股票收益率预测模型中，这项因子无效且具有干扰作用。因此，本文进一步删除这项异象因子，仅使用 illq\_risk、sharechg、turn\_std 作为输入变量对股票收益率进行预测。此时，平均累计收益率达到 4.20%，较初始值增长了 0.26 个百分点，提高了基于 Lasso 回归算法预测股票收益模型的盈利能力。为了保证模型的稳健性，同样在将训练集调整为 24 个月和 36 个月的基础上，基于删除 egr 的模型，结果显示盈利能力均有小幅度的提升，分别增加了 0.52%和 0.95%，检验了股东权益变化在电商概念股票收益率预测模型中具有干扰作用结论的稳健性。

### 4.3. 替换模型算法

为了探究不同机器学习算法在预测电商概念股票上的差异性，本文首先使用岭回归(Ridge)算法替代

Lasso 回归算法, 仍然在 4 项异象因子的条件下预测股票收益率, 并据此做出投资决策, 在期末通过计算累计收益率均值的方式来比较算法的适用性。在 Lasso 回归中, 通过在损失函数中加入一个 L1 正则化项来降低模型的复杂程度, 其目标是最小化函数  $\sum_{i=1}^n (y_i - X_i^T W)^2 + \lambda \sum_{j=1}^p |w_j|$ 。其中,  $\lambda \sum_{j=1}^p |w_j|$  是 L1 正则化项,  $\lambda$  是正则化参数, 控制正则化强度。而在岭回归中, 是通过在损失函数中加入一个 L2 正则化项来减少模型的复杂程度, 其目标是最小化函数  $\sum_{i=1}^n (y_i - X_i^T W)^2 + \lambda \sum_{j=1}^p w_j^2$ 。其中,  $\lambda \sum_{j=1}^p w_j^2$  是 L2 正则化项,  $\lambda$  仍然用于控制正则化强度。Lasso 回归通过绝对值项惩罚权重, 可以将某些股票特征权重完全置为 0, 从而实现特征的选择。而岭回归是通过平方项惩罚权重, 只能使一些股票特征权重变小, 而不能完全将其置为 0, 因此岭回归会将全部的股票特征都保留在模型之中。另外, 岭回归在特定条件下能够使用解析解直接求解, 而不需要迭代优化方法, 能够提高计算效率。因此, 本文在股票特征数量较少的情况下, 选择岭回归算法替代 Lasso 回归, 研究改变模型算法能否提高投资收益率。计算结果显示, 166 只股票期末累计收益率均值为 98.91%, 较使用 Lasso 回归算法增加了 94.97%, 明显提高了根据预测结果做出投资决策的收益率。

进一步地, 使用将岭回归与 Lasso 回归结合起来的弹性网络(Elastic Net)回归算法用于模型之中, 探究其是否能改进单独使用上述两种算法的预测能力。弹性网络回归也是一种正则化线性回归方法, 其结合了 Lasso 回归和岭回归的正则化项来平衡两者的特点, 从而更够更好地处理股票特征选择和多重共线性的问题。弹性网络回归的目标是最小化函数  $\sum_{i=1}^n (y_i - X_i^T W)^2 + \lambda_1 \sum_{j=1}^p |w_j| + \lambda_2 \sum_{j=1}^p w_j^2$ 。其中,  $\lambda_1 \sum_{j=1}^p |w_j|$  是 L1 正则化项, 用于选择股票特征;  $\lambda_2 \sum_{j=1}^p w_j^2$  是 L2 正则化项, 用于处理股票特征之间存在的多重共线性。弹性网络回归可以通过调整  $\lambda_1$  和  $\lambda_2$  的值来平衡股票特征选择和处理多重共线性, 当  $\lambda_1 = 0$  时, 弹性网络回归转化为岭回归, 而  $\lambda_2 = 0$  时, 则转化为 Lasso 回归。从使用弹性网络算法替换 Lasso 回归算法的结果来看, 样本股票累计收益率均值变为 99.35%, 高于使用 Lasso 回归算法的 3.94% 和使用岭回归算法的 98.91%。比较结果表明, 将 Lasso 回归算法和岭回归算法共同用于股票收益率预测模型中的效果要优于单独使用 Lasso 回归算法和岭回归算法的效果。

## 5. 主要结论

本文以 2009~2020 年 166 只电商概念股票为研究样本, 选取 4 项异象因子和监督机器学习中的 Lasso 回归算法构建量化投资模型, 探究了机器学习驱动的投资模型在电商概念股票中取得的实证绩效。研究结果发现: (1) 机器学习驱动的量化投资模型能够在电商概念股票中取得一定的投资收益, 但投资收益率较低。在未扣除无风险收益率的情况下, 基于 Lasso 回归算法的投资模型平均累计收益率为 3.94%。(2) 4 种异象因子对模型的预测重要性存在显著差异, 在模型使用 Lasso 回归算法的条件下, 股本增长率的重要程度最高。(3) 对比不同算法在电商概念股票中的实证绩效, 弹性网络回归算法表现优于岭回归算法, 岭回归算法优于 Lasso 回归算法, 且弹性网络回归算法和岭回归算法均显著优于 Lasso 回归算法, 能够大幅度提升电商企业股票收益率预测的准确性。

## 参考文献

- [1] 李子煜, 张金珠, 高青山. 基于模型和算法的量化投资方法股票预测研究综述[J/OL]. 计算机工程与应用: 1-12. <https://link.cnki.net/urlid/11.2127.TP.20250529.1505.008>, 2025-06-30.
- [2] 丁鹏. 量化投资: 策略与技术[M]. 北京: 电子工业出版社, 2016.
- [3] 黄鑫, 周永刚. 从阿里巴巴上市看中国电子商务的历史、现状与发展[J]. 商业研究, 2015(12): 11-17.
- [4] 候杰, 张心雨. 跨境电商综合试验区对外贸高质量发展的影响研究[J]. 商业经济研究, 2025(8): 123-128.
- [5] 王立勇, 张晨阳, 袁小康. 跨境电商产业政策对出口开放度的影响: 来自合成控制法的证据[J]. 国际贸易, 2024(3): 42-57.

- [6] 金丽静. “双循环”背景下跨境电商发展对我国对外贸易的影响及建议[J]. 商业经济研究, 2025(4): 120-123.
- [7] 梁娅青. 二元创新对跨境电商企业绩效的影响研究——基于管理者过度自信的调节作用[J]. 商业经济研究, 2024(10): 152-155.
- [8] 常梅, 王美强. 中国上市电商企业综合绩效研究——基于生产运营和资本运作视角[J]. 技术经济, 2022, 41(10): 149-160.
- [9] 陈千里, 周少甫. 上证指数收益的波动性研究[J]. 数量经济技术经济研究, 2002(6): 122-125.
- [10] 李胜利. 中国股票市场杠杆效应研究[J]. 证券市场导报, 2002(10): 10-14.
- [11] 闫政旭, 秦超, 宋刚. 基于 Pearson 特征选择的随机森林模型股票价格预测[J]. 计算机工程与应用, 2021, 57(15): 286-296.
- [12] 包振山, 郭俊南, 谢源, 等. 基于 LSTM-GA 的股票价格涨跌预测模型[J]. 计算机科学, 2020, 47(S1): 467-473.
- [13] 张晨希, 张燕平, 张迎春, 等. 基于支持向量机的股票预测[J]. 计算机技术与发展, 2006(6): 35-37.
- [14] 李斌, 邵新月, 李玥阳. 机器学习驱动的基本面量化投资研究[J]. 中国工业经济, 2019(8): 61-79.
- [15] 任君, 王建华, 王传美, 等. 基于 ELSTM-L 模型的股票预测系统[J]. 统计与决策, 2019, 35(21): 160-164.
- [16] 李辉, 化金金, 邹波蓉. 基于 RF-LSTM 组合模型的股票价格预测[J]. 河南理工大学学报(自然科学版), 2022, 41(1): 136-142.
- [17] 张鹏, 莫仕茵, 曹卿. 基于机器学习的多目标投资组合优化研究[J]. 华南师范大学学报(自然科学版), 2024, 56(4): 100-110.
- [18] DeMiguel, V., Garlappi, L. and Uppal, R. (2007) Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *Review of Financial Studies*, **22**, 1915-1953. <https://doi.org/10.1093/rfs/hhm075>
- [19] Richardson, S.A., Sloan, R.G., Soliman, M.T. and Tuna, İ. (2005) Accrual Reliability, Earnings Persistence and Stock Prices. *Journal of Accounting and Economics*, **39**, 437-485. <https://doi.org/10.1016/j.jacceco.2005.04.005>
- [20] Amihud, Y. (2002) Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets*, **5**, 31-56. [https://doi.org/10.1016/s1386-4181\(01\)00024-6](https://doi.org/10.1016/s1386-4181(01)00024-6)
- [21] Pontiff, J. and Woodgate, A. (2008) Share Issuance and Cross-sectional Returns. *The Journal of Finance*, **63**, 921-945. <https://doi.org/10.1111/j.1540-6261.2008.01335.x>
- [22] Chordia, T., Subrahmanyam, A. and Anshuman, V.R. (2001) Trading Activity and Expected Stock Returns. *Journal of Financial Economics*, **59**, 3-32. [https://doi.org/10.1016/s0304-405x\(00\)00080-5](https://doi.org/10.1016/s0304-405x(00)00080-5)
- [23] 王子豪. 基于深度学习的股票价格预测模型与交易策略研究[D]: [硕士学位论文]. 广州: 广东技术师范大学, 2024.