

南开大学硕士学位论文/实践成果报告

(匿名评阅学位论文/实践成果报告封面)

学位论文/实践成果报告题目 (含英文题目):

基于深度学习的股票定价实证分析

Empirical Analysis of Stock Pricing based on Deep Learning

专业学位类别: 金融

专业领域:

学位论文/实践成果编号:

摘要

Recurrent Neural Networks

随着多因子资产定价模型的发展，传统统计方法在高维因子建模中的局限性日益凸显，尤其是在处理非线性关系和时序依赖性方面存在显著不足。近年来，机器学习和深度学习技术因其强大的非线性捕捉能力被广泛应用于股票收益预测领域。然而，现有文献对循环神经网络及其衍生模型（RNN、LSTM、GRU）在因子模型风险溢价预测中的系统性比较及其适用条件仍缺乏深入探讨。本文以 1990-2021 年美国股票市场为样本，系统研究循环神经网络模型在股票超额收益预测中的表现，旨在揭示不同模型在不同时间窗口下的性能差异及其经济意义。研究通过对比神经网络（NN）与循环神经网络模型，结合单期因子与多期因子数据，设计抽样训练策略，构建样本外预测评估体系，并基于 Fama-French 五因子模型对预测收益进行 Alpha 检验，以验证深度学习模型在捕捉非线性市场异象中的独特价值。

研究结果表明：深度学习构建的投资组合可超越指数基准收益，其性能受时间窗口和模型复杂度影响。数据噪声显著影响模型泛化能力，使用抽样数据降低噪声后，预测精度明显优于完整数据。循环神经网络能捕捉 Fama-French 五因子无法解释的市场异象，较普通神经网络提取更多未因子化信息，表明引入动态非线性补偿项可优化因子模型对收益关系的刻画。但模型预测能力与组合收益存在非对称性——高预测精度未必转化为高累积收益，市场噪声、模型复杂度和交易成本均制约实际表现。未来需优化因子筛选、数据处理和模型架构，结合更系统的回测分析，以提升深度学习在金融领域的应用效果。

关键词：深度学习；超额收益预测；循环神经网络；因子模型

Abstract

With the development of multi factor asset pricing model, the limitations of traditional statistical methods in high-dimensional factor modeling have become increasingly prominent, especially in dealing with nonlinear relationships and time-series dependence. In recent years, machine learning and deep learning technology are widely used in the field of stock return prediction because of their powerful nonlinear capture ability. However, the existing literature still lacks in-depth discussion on the systematic comparison and application conditions of recurrent neural networks (RNN, LSTM, GRU) in risk premium prediction. Based on the sample of American stock market from 1990 to 2021, this paper systematically studies the performance of recurrent neural network model in the prediction of stock abnormal return, aiming to reveal the performance differences of different models in different time windows and their economic significance. By comparing the neural network (NN) and the recurrent neural network model, combined with the single period factor and multi period factor data, this paper designs the sampling training strategy, constructs the out of sample prediction evaluation system, and alpha test the predicted returns based on the Fama-French five factor model, so as to verify the unique value of the deep learning model in capturing the nonlinear market anomalies.

The results show that the portfolio constructed by deep learning can exceed the index benchmark return, and its performance is affected by the time window and model complexity. The data noise significantly affects the generalization ability of the model, and the prediction accuracy is significantly better than that of the complete data after using the sampling data to reduce the noise. The cyclic neural network can capture the market anomalies that cannot be explained by the Fama-French five factors, and extract more non factorized information than the ordinary neural network, which shows that the introduction of dynamic nonlinear compensation term can optimize the description of the return relationship in the factor model. However, there is asymmetry between the prediction ability of the model and the portfolio return - high prediction accuracy does not necessarily translate into high cumulative return, and market noise, model

complexity and transaction costs all restrict the actual performance. In the future, it is necessary to optimize factor screening, data processing and model architecture, combined with more systematic back test analysis, to improve the application effect of deep learning in the financial field.

Key Words: Deep Learning; Excess Return Prediction; Recurrent Neural Network; Factor Model

目录

1 绪论	1
1.1 研究背景	1
1.2 研究意义	3
1.3 文献综述	4
1.4 研究内容及结构安排	8
1.5 创新点	9
2 相关理论	10
2.1 因子模型理论	10
2.2 深度学习算法	11
2.2.1 神经网络	11
2.2.2 其他辅助优化方法	13
2.3 循环神经网络算法	16
2.3.1 RNN	16
2.3.2 LSTM	18
2.2.4 GRU	19
2.4 性能评估	20
3 数据与实证分析	22
3.1 数据来源与预处理	22
3.2 因子分类	23
3.2.1 宏观因子	23
3.2.2 微观因子	25
3.3 模型设计	30
3.4 实证结果	33
3.4.1 模型预测效果对比	33
3.4.2 选股策略模型构建与结果	34
3.4.3 稳健性检验	38
3.5 进一步分析	39

目 录

3.5.1 完整数据的模型效果对比.....	39
3.5.2 时间步为 3 的模型效果对比.....	40
3.5.3 时间步为 3 的选股模型构建与结果.....	41
3.5.4 模型交叉 α 验证.....	46
4 结论与展望	48
4.1 研究结论	48
4.2 不足与展望	49
参考文献	51
附录	55

1 绪论

1.1 研究背景

股票市场作为现代金融体系的核心，经历了数个世纪的发展，已成为全球经济的重要支柱。自 17 世纪荷兰东印度公司发行世界上第一支股票以来，股票市场不断演变，规模日益壮大，交易机制也日趋复杂。资产的风险溢价预测是资产定价的重要组成部分，精准的预测直接影响投资决策和管理风险，对市场效率、投资者利益和市场稳定具有重要意义。投资者希望通过风险溢价预测制定科学的投资决策并进行相应的资产配置，来实现风险和收益的平衡。监管机构可以通过风险溢价预测监测市场波动、识别潜在风险，指定更精准的政策措施，促进金融市场的稳定与可持续发展。

早期的资本资产定价模型（CAPM）由 Sharpe（1964）提出，基于均值-方差框架，指出资产的预期收益取决于市场风险溢价和资产的贝塔系数，为资产定价奠定了基础。随后，Merton（1973）提出跨期资本资产定价模型（ICAPM），将 CAPM 拓展到多期框架，考虑了投资者的跨期消费和投资决策。Ross（1976）的套利定价理论（APT）则从无套利条件出发，认为资产收益可由多个因子解释，为多因子模型与机器学习方法的结合提供了理论依据。Fama 和 French（1993）的三因子模型在市场因子基础上，加入规模因子和价值因子，显著提升了对股票收益的解释力。后续，他们又在 2015 年提出五因子模型，引入盈利能力因子和投资风格因子，进一步完善了因子定价模型体系。

随着研究的深入，大量新因子的不断涌现，形成了“因子动物园”现象（Cochrane, 2011），其中已有研究识别出多达 316 个因子（Harvey 等, 2016）。由于众多因子之间存在复杂的相关性，资产定价模型的复杂性和不确定性大幅增加。然而，传统资产定价模型的统计推断方法主要针对低维情形设计，在高维情况下，因子的经济解释变得模糊，难以确定哪些因子真正反映了资产的风险与收益关系。数据挖掘和多重检验问题可能导致部分因子只是统计上的偶然现象，缺乏经济意义上的稳健性。

在这种背景下，传统统计方法在处理高维、复杂的因子数据时面临诸多挑战，不仅难以准确估计因子模型参数，还可能降低资产定价的准确性和可靠性。随着

研究逐步向大规模因子集和测试资产的分析过渡，高维统计方法在实证资产定价中的重要性日益凸显。Kelly 等（2019）通过 IPCA 方法尝试解决高维问题，但仍面临因子解释的挑战。

如果把资产超额收益预测任务总结为标准的监督学习和回归任务，以因子作为自变量，资产的超额收益作为因变量，那么上述理论中的线性和高维统计方法则代表二者之间的函数关系。自然地，可以联想到，该函数也可以是应用越来越广泛的机器学习等非线性函数模型。

机器学习和深度学习包含多种类型的研究方法，如监督学习、无监督学习、半监督学习等，主要区别在于对数据样本标签的要求：监督学习需要样本的标签，如股票收益；半监督学习需要部分标签，无监督学习无需标签。由于资产收益预测的任务是得到预期收益，所以资产收益预测类任务属于监督学习任务，监督学习中的回归方法天然适用于该类问题研究。在此背景下，机器学习作为一种能够从数据中高效提取模式和规律的技术，在金融领域展现出了巨大潜力。通过捕捉潜在的非线性关系和复杂的交互效应，机器学习在资产定价、风险管理和投资组合优化等方面取得了显著的成果。惩罚线性回归方法和降维技术应运而生，这些方法可以精准筛选关键变量，控制模型复杂度，提升模型稳定性和预测精度，但是这些方法无法处理因子间的交互作用。回归树方法在实证中表现更为良好，特别是梯度提升回归树和随机森林，能够捕捉复杂的非线性关系和多变量交互作用，且模型能够给出相关特征的重要性，具有良好的可解释性，尤其适应于动态复杂的金融市场环境。Gu 等（2020）在美国市场中系统验证了机器学习方法的有效性，他们通过比较多种机器学习模型在横截面收益预测中的表现，发现非线性模型的预测效果显著优于传统线性模型。

随着数据规模的进一步扩大以及软硬件技术的飞速发展，数据计算能力不断提升，越来越多的学者发现神经网络能够精准捕捉大数据中的非线性关系。Cybenko（1989）首次严格证明单隐层网络（仅需一个隐层）即可逼近任意连续函数，解决了长期存在的开放性问题，虽然可能使用到天文级别数量的神经元。简单的神经网络模型较机器学习模型具有更好的预测能力这一点也被证明（Gu 等，2020）。然而，在深度学习方面，Gu 等仅研究了简单的前馈神经网络。近年来随着复杂的神经网络模型迭代演进，其在金融领域的应用边界不断拓展，在不同结构的金融数据方面展示着独特优势，需要被深入研究。例如，LSTM 能够强化时间序列的长期记忆能力，在预测风险溢价时可以通过调整记忆权重的方式

实现高效预测；CNN 擅长挖掘高维因子间的局部空间关联，构建因子图像，避免人工筛选的主观偏差；Transformer 利用自注意力机制实现跨模态信息的全局交互，在加密货币市场预测、语义关联等领域具有独特优势。

尽管机器学习和深度学习在金融领域的应用前景广阔，但它仍面临一些挑战，例如，“黑箱”模型的可解释性不足，可能让投资者难以理解模型的决策依据，限制了其应用范围；模型训练所需计算资源高；过拟合现象或数据噪声困扰模型的训练过程；非结构化数据的建模缺乏统一框架，限制了其在多资产场景中的应用；不同模型结构适用的场景不同等。因此，结合深度学习和多因子模型的研究，既具有重要的理论意义，也对实际投资实践有深远影响。

1.2 研究意义

本文主要对循环神经网络深度学习模型相关算法（RNN、LSTM、GRU）在股票风险溢价预测方面的应用进行研究。

首先，现有文献在因子模型对深度学习模型的研究中，对不同循环神经网络的系统比较尚缺乏深入探讨。本文通过比较 RNN、LSTM 与 GRU 这三个循环神经网络与普通神经网络在股票超额收益预测中的表现，以及模型构造的投资组合相互之间的解释能力，发现深度学习方法可有效捕捉未被因子化的市场信息，探讨了深度学习方法在资产定价中的应用，为因子模型提供了新的理论视角。通过引入单期因子与多期因子数据来预测未来收益，本文评估了不同循环神经网络模型在长期非线性关系建模中的能力，拓展了现有资产定价模型的时空维度，为后续研究提供了多期因子预测理论框架和分析工具，为资产定价和市场预测研究提供了新的思路。

其次，本文通过比较不同深度学习模型的预测效果，为量化投资和选股策略提供了实证依据，通过深度学习模型构造的投资组合能够获取超越基准指数的收益。本文的研究方法可为金融机构提供更精细的收益预测工具，尽管预测精度与收益表现的关系需要进一步优化，但这一分析框架仍能为风险管理、资产配置和投资组合优化提供数据支持，有助于提升机构投资决策的科学性。

最后，本文探索深度学习在金融市场的应用，有助于推动人工智能在量化投资和资产定价中的发展，加速金融科技的创新进程。对于市场参与者而言，更深入的收益预测研究可以帮助其更好理解资产价格变动规律，减少非理性投资行

为，从而在一定程度上提高市场效率，优化资源配置，促进市场稳定发展；同时本文构建的预测模型在算力需求方面要求不高，有助于中小机构和个人投资者应用前沿 AI 技术进行可行性验证，这种技术民主化尝试有助于缩小机构间、机构与个人间的信息不对称，促进市场公平竞争环境建设。对市场监管者而言，本文通过应用深度学习技术提升数据分析能力，为其提供了新的决策工具，有助于市场透明度的提升，并推动更加科学合理的数据驱动投资理念。

1.3 文献综述

因子模型可以用于横截面分析，研究不同资产之间的收益差异，也可以用于时间序列分析，研究个体资产或组合的收益动态。

资本资产定价模型（CAPM）提供了基础性的视角，假设系统性风险是决定收益的主要因素。然而，该模型存在显著局限性，尤其是在解释不同市场中股票收益的横截面差异方面。因此，为了弥补这些不足，研究者们提出了多种改进方案，主要体现在多因子模型的构建上。1993 年，Fama 和 French 提出的三因子模型在多个市场的实证研究中被证明比 CAPM 具有更强的解释能力，例如在澳大利亚市场的研究中（Chiah 等，2016；Lalwani 和 Chakraborty，2019）。在此基础上，Fama 和 French 进一步扩展模型，提出了五因子模型（FF5），增加了盈利能力因子和投资因子，进一步提升了模型对资产收益的解释能力，尤其是在小盘股和高账面市值比股票中的表现更为突出（Liao 等，2023；Roy 和 Shijin，2018）。研究表明，FF5 在大多数市场上的表现优于其前身 FF3，进一步凸显了其在当代资产定价中的重要性（Machado 和 de Medeiros，2011；Racicot 和 Rentz，2015）。例如，在印度尼西亚股票市场的研究发现，FF5 相较于 FF3 能够更好地解释超额组合收益（Putra 和 Chalid，2021；Wedagama 等，2022）。这些研究结果共同表明，引入更多的风险因子能够显著提高资产定价模型的预测能力。此外，Roy 和 Shijin（2018）提出的六因子模型进一步增加了额外的经济变量，以解释资产价格的变化，反映了学术界不断完善资产定价理论的努力。这些发展趋势表明，资产定价正朝着更加多维度的方向演进，通过整合多个风险因子，使研究者能够更细致地理解金融市场中的收益特征。

尽管多因子模型在实证研究中表现良好，但仍存在一定的局限性。例如，在某些市场条件下，如时间变动的风险因子或市场流动性问题，模型的解释力可能

受到影响 (Galvao 等, 2019)。资产定价理论的持续发展表明, 虽然传统模型奠定了基础, 但未来的进步仍然依赖于整合多因子, 以更全面地捕捉不同市场中的收益动态。

时间序列因子模型在股票超额收益预测中应用广泛。自回归移动平均模型 (ARMA) 及其扩展模型, 如自回归条件异方差模型 (ARCH) 和广义自回归条件异方差模型 (GARCH), 常用于刻画股票收益的时间序列特征和波动聚集性 (Bollerslev, 1986)。这些模型通过分析历史收益数据的自相关性和异方差性, 对未来收益进行预测。然而, 传统时间序列模型在捕捉股票市场复杂的非线性动态关系方面存在局限。Kelly 等 (2021) 使用 ICPA (Instrumented PCA) 模型估计隐含的风险因子, 并发现 Beta 在短时间内存在显著变化, 解释了动量现象主要来源于股票的时间变化风险暴露, 并表示通过正确构建时变条件因子模型, 能更好地解释动量和长期反转现象, 并获得更高的收益预测能力。

资产风险溢价度量本质是预测问题, 而机器学习和深度学习则擅长从数据中提取复杂结构和模式, 减少人为干预, 以辅助预测, 因此在该领域具有天然的应用优势 (李斌等, 2019)。为应对因子模型的困境, 机器学习等非线性模型逐渐被引入资产定价领域。机器学习模型具有强大的非线性拟合能力, 能够捕捉数据中复杂的关系, 有效处理高维数据, 避免传统方法在变量选择和模型设定上的局限性 (Gu 等, 2020)。许杰等 (2022) 将决策树、随机森林、集成学习等算法分类为基于特征处理的机器学习算法, 将 CNN、LSTM、强化学习、文本分析等分类为端到端处理的深度学习方法, 并指出深度学习方法具有强大的复杂特征提取能力和非线性函数拟合能力, 能适应金融大数据与高频金融数据分析的需要。近年来, 机器学习/深度学习在因子投资中的应用主要集中于三个方面: 因子选择、收益预测和投资组合优化。

在因子选择方面, 机器学习方法通过特征选择算法, 能够从海量因子池中筛选出最具解释力的因子。例如, 套索回归 (Lasso) 等稀疏建模方法通过引入正则化约束, 有效地控制了模型复杂性, 同时避免了因子间多重共线性对模型性能的影响。Yuan (2024) 在研究中利用 IC 分析和随机森林算法评估因子重要性, 最终确定线性回归模型的输入变量。深度学习在因子选择中可以利用多个隐藏层对因子进行特征提取, 无需大量的手动特征选择 (Sahu 等, 2023), 这些模型还可以生成辅助稳健投资决策的高维特征 (Yu, 2024)。Lee 和 Moon (2023) 使用深度强化学习方法提出 TACR 模型, 将决策 Transformer 与 Actor-Critic 算法

结合，结合离线学习和正则化，实现了基于历史数据的自动化资产配置决策。这种混合方法可以通过动态调整市场条件并在管理风险的同时实现回报最大化来产生更好的投资结果（Yu, 2024）。在这种情况下，因子选择也受益于深度学习的降维能力。通过各种神经网络层处理大型数据集，深度学习模型可以识别出影响投资绩效的最相关因子（Sahu 等, 2023）。此外，最近的研究强调了深度强化学习在开发自动交易策略中的重要性日益增加，这进一步强调了有效因素选择的相关性（Lee 和 Moon, 2023）：强化学习框架允许根据市场互动的反馈进行持续调整，进一步优化投资模型中的因子选择。

在收益预测方面，非线性机器学习算法如梯度提升决策树、随机森林等通过树结构捕捉因子间的复杂线性关系，不仅提高了预测精度，也具有较好的解释性。李斌等（2019）在中国市场中使用 12 种机器学习算法构建基于异象因子的股票收益预测模型及投资组合，发现机器学习算法能有效识别异象因子-超额收益间的复杂模式，且能获得更好的投资绩效。Gu 等（2020）在美国市场中系统验证了机器学习方法的有效性，他们通过比较多种机器学习模型在横截面收益预测中的表现，发现非线性模型的预测效果显著优于传统线性模型。Gu 等（2021）使用条件自动编码器模型将 Barra 和 Kelly 等（2019）的 IPCA 模型扩展到非线性结构，并增加了额外的解释变量，发现自动编码器方法能够更有效地捕捉资产回报与特征的复杂关系，并在未来收益方面表现出色。Zhang 等（2018）改进传统多因子选股模型，结合 LSTM 处理面板数据，有效捕捉了因子间的动态关系，通过分类方式构造投资策略，预测准确率达 90%，提升了选股策略的盈利能力和稳定性。在中国市场，周亮（2022）将随机森林模型用于股票多因子投资领域，发现随机森林模型能够拟合和预测个股相对收益率；李仁宇等（2023）发现 LightGBM 和随机森林在预测收益率方面优于 OLS 模型；张晗等（2024）通过 CNN-LSTM 模型对数据中时空特征建模，提升了预测精度与鲁棒性。

在投资组合优化方面，深度学习模型能够通过改进决策流程，使投资者更有效地应对金融市场的复杂性，使用大量历史和实时数据提升预测性能。在风险管理方面，卷积神经网络（CNN）等技术被用于预测投资组合中的条件风险值（CVaR），从而支持与投资目标一致的风险管理决策（Wu 等, 2021）。深度学习模型适应金融数据非线性模式的能力，进一步增强了其在识别能够抵御不良市场条件的最优资产配置方面的有效性（Liang 等, 2018）。例如，通过采用深度强化学习（DRL）策略，可以在结构化决策框架下最大化风险调整收益，其中智能

体可以通过与市场环境的持续交互，来学习最优投资行为（Jin, 2023）。Zhang（2023）探索 LSTM、GRU、自注意力和 Transformer 四种深度学习模型在均值-方差优化框架下的资产配置效果，发现 LSTM、GRU 在捕捉时序依赖方面有效，但自注意力机制通过动态权重分配进一步提升了预测准确性。Lee 和 Moon（2023）的研究表明，RNN 可以在更广泛的投资组合管理系统中有效应用，并通过结合 LSTM 和 GRU 等深度学习技术逐步提升预测质量。

特别地，深度学习中的循环神经网络模型，如简单循环神经网络（RNN）及其变体长短期记忆网络（LSTM）和门控循环单元（GRU），在股票收益预测中受到关注。RNN 能够处理具有时间序列结构的数据，通过隐藏层的循环连接捕捉序列中的长期依赖关系（Elman, 1990）。但传统 RNN 存在梯度消失和梯度爆炸问题，限制了其对长期依赖信息的学习能力。LSTM（Hochreiter 和 Schmidhuber, 1997）和 GRU（Cho 等, 2014）通过引入门控机制，有效解决了这一问题。LSTM 的记忆单元和门控结构使其能够更好地保存和更新长期信息，在股票收益预测中表现出更好的性能。任晓萍等（2025）构造并对比了 RNN、LSTM、GRU 指数预测模型，表明深度学习模型比强化学习模型在回测指标方面有明显优势。加密货币方面，Gunarto 等（2023）和 Xiang（2024）均发现深度学习模型可以有效预测加密货币价格，在股票市场预测中有类似作用。不过，上述研究往往使用开盘价、收盘价等数据预测下期价格，即证明在处理时间序列数据的长短期依赖关系上表现出色，能显著提高预测时效性与精度（李星毅等, 2025），与多因子领域涉及多种因子的研究有所不同。除此之外，GRU 在国债期货波动率预测中有较强的适用性（孟祥煜, 2023），LSTM 被整合到混合模型中，通过与其他方法结合以增强股价预测精度，从而构建更稳健的预测系统（Furizal 等, 2024）。许杰等（2022）指出，LSTM 等算法更多被应用于图像、声音和自然语言处理中，在资产定价方面使用较少，未来值得进一步研究。

此外，机器学习/深度学习还可以在复杂和非结构化数据集中使用预测信息来改进预期回报估计。例如，Ke 等（2019）提出了一种新的监督主题模型，用于从原始新闻文本构建回报预测，并展示了其样本外预测的能力。Jiang 等（2023）、Obaid 和 Pukthuanthong（2022）使用计算机视觉文献中的机器学习模型在图像数据中挖掘返回预测信息，结果表明文本和图像数据在短期内能够带来良好的回报预测收益，并且可能是以相对变化较快的市场情绪作为基础。姚加权等（2021）结合词典重组与深度学习，分别构建适用于年报正式用语和社交媒体非正式用

语的情绪词典,发现基于此词典提取的指标能有效预测中国股市收益率。张少军等(2025)结合中文金融情感词典与 BERT-BiLSTM 模型、于孝建等(2024)结合股吧文本情感数据与 SVM-LSTM 模型得到了类似的结论。此类非结构化数据集中的应用并不是本文关注的重点内容,故不再进行赘述。

当然,机器学习/深度学习模型在金融数据中的应用有一定弊端。(1)模型本身比较复杂,会抽取过多的深层次特征,这些深层次特征具有的弱性因子暴露特性与研究发掘金融市场本质规律可能不一致。(2)金融数据时间跨度相对较短,即使是比较完备的美国股市数据,从 1970 年算起,也只有 600 个月度数据,一定程度上会影响模型的准确性。(3)机器学习/深度学习模型复杂,金融规律解释难度较高,模型的泛化能力也需加强,有一定性能瓶颈(李星毅等,2025;交通银行金融服务中信支付清算课题组,2025)。

1.4 研究内容及结构安排

本文旨在通过系统对比循环神经网络模型在股票超额收益预测的准确率、分组收益,探索时间序列在长期非线性关系中的建模能力。

由于本文主要进行方法论探索,利用已有的根据美国数据建模的文章进行对比,会更清楚边际贡献,故本文使用美国排名前三的交易所(NYSE、AMEX、NASDAQ)中的上市股票作为研究对象,使用的因子数据为经学界验证的标准数据集,使用 NN-3、RNN、LSTM、GRU 模型分别进行美股的超额收益预测、分组分析与回测,对模型样本外 R 方、分组预测收益率、真实收益率、波动率、夏普比率、累积收益率进行对比。之后验证分组时参数设置、是否对数据抽样和因子时间步设置对结果的影响。

文章的具体结构安排如下:

第一章为绪论,首先阐述研究背景与研究意义,然后梳理了文献中因子模型的发展现状以及机器学习、深度学习模型在因子投资领域中的应用,最终总结本文研究内容及创新点。

第二章深入介绍了本文涉及的理论基础,首先介绍因子模型理论,接下来介绍深度学习算法、循环神经网络深度学习模型原理,以及深度学习模型的优化算法,最后对模型性能评估进行介绍。

第三章是数据与实证分析。首先介绍数据来源与预处理方法,然后简要介绍

使用的因子及模型设计，之后进行实证分析。实证分析过程中，首先对比模型预测能力，进行选股策略构建，分析结果，并进行稳健性检验，接下来进行进一步分析，从数据完整性、时间步长短角度对比模型效果，最后通过模型交叉 alpha 检验模型间的相互关系。

第四章为研究结论、不足与展望，基于循环神经网络的深度学习模型能够捕捉传统神经网络无法捕捉的信号，且其构造的投资组合能够获取超越指数的收益，但仍需在后续因子筛选、模型架构、回测及非结构化数据方面进行改进。

1.5 创新点

本文在以下几个方面进行了创新和拓展。

首先，本文将循环神经网络模型（RNN、LSTM、GRU）系统性引入股票超额收益预测领域，在已有机器学习方法的基础上进行延申和优化。本文使用经学术界验证的标准数据集，与前人的研究具有对比性，验证了循环神经网络模型可以应用于股票超额收益预测中，并能带来相应的回报。

其次，在深度学习训练方法上进行探索：使用抽样数据完成模型训练，以减少数据噪声对模型学习过程的干扰。实验结果表明，该方法可能有助于模型更好地拟合因子和收益之间的非线性关系，提高模型的泛化能力，减少过拟合风险，为深度学习方法在金融市场中的应用提供了新的优化策略，也为未来在高噪声环境中的深度学习研究提供了新的思路。

最后，本文拓展了多期因子模型的应用。通过对比单期因子与多期因子，分析了不同时间窗口对预测效果的影响。这一研究有助于揭示因子有效性的时间维度特征，为未来资产定价研究提供更丰富的理论依据。

2 相关理论

2.1 因子模型理论

提到经典的因子模型理论，最有代表性的莫过于资本资产定价模型(CAPM)、套利定价理论(APT)。

资本资产定价模型(CAPM)核心假设为市场是完全有效的，市场价格已经反映了所有的可用信息，认为资产的预期回报率取决于资产和市场风险之间的关系，公式为：

$$E(R) = R_f + \beta[E(R_m) - R_f] \quad (2.1)$$

其中， $E(R)$ 表示资产的期望收益率， R_f 表示无风险收益率， β 表示资产相对于市场整体的风险系数， $E(R_m)$ 表示市场组合的期望收益率。

套利定价理论将影响因素由单因子拓展到多因子，认为资产收益受多个宏观经济因子（如利率、通胀、GDP增长等）的驱动，公式为：

$$E(R) = R_f + \sum_{k=1}^n \beta_k \lambda_k \quad (2.2)$$

套利定价理论无需依赖市场组合，更具灵活性，但因子选择依赖实证，且并未明确指出哪些因素影响了资产价格。

如果CAPM模型认为资产的超额收益可以由市场因子解释，那么Fama-French的三因子模型则是将因子拓展到三个：市场组合因子、市值因子SMB和账面市值比因子HML：

$$E(R) = R_f + \beta[E(R_m) - R_f] + sE(SMB) + hE(HML) \quad (2.3)$$

然而，即使是将因子个数扩充为三个，资产的预期回报中仍然有不能用风险暴露解释的部分（即alpha），学者们将相对于现有模型具有显著alpha的投资组合成为异象，并认为仍有大量未被发掘的因子造成了此类异象，于是动量因子（Carhart四因子模型）、盈利因子和投资因子（Fama-French五因子模型）陆续被提出。

如果把资产超额收益预测任务总结为标准的监督学习和回归任务，则可以表示为如下的函数形式：

$$R_{t,i} = f(x_{t-1,i}; \theta) + \varepsilon_{t,i} \quad (2.4)$$

其中， $f(\cdot)$ 定义为一个参数为 θ 的函数， $R_{t,i}$ 为资产 i 在第 t 期的超额收益， $x_{t-1,i}$

为因子向量, $\varepsilon_{t,i}$ 为误差项。若上述经典的因子模型理论中使用的都是线性方法, 那么自然地可以联想到, $f(\cdot)$ 可以是应用越来越广泛的机器学习等非线性模型。

2.2 深度学习算法

机器学习可以看作众多预测函数形式 $f(\cdot)$ 及各种算法的集合, 适用于回归任务的机器学习和深度学习算法均可以被用来建立资产收益预测模型。近年来, 被验证能有效预测资产超额收益的机器学习模型越来越多, 随机森林、提升回归树和深度学习中的神经网络算法均被证明有良好的预测能力。由于 Gu 等 (2020) 文章中, 效果最好的模型为 3 层神经网络 (简称 NN-3 模型), 故本节主要参考 Chollet (2021) 对神经网络模型进行介绍。

2.2.1 神经网络

神经网络模型模仿人脑神经元的工作方式, 通过分层非线性变换将输入数据映射到输出, 学习数据中的复杂模式。基础的前馈神经网络 (Feedforward Neural Network, FNN) 中, 每一层的神经元可以接收前一层神经元的信号, 并产生信号输出到下一层。第 0 层叫做输入层, 最后一层叫做输出层, 其他中间层叫做隐藏层。整个网络中信号从输入层向输出层单向传播, 可用下图表示:

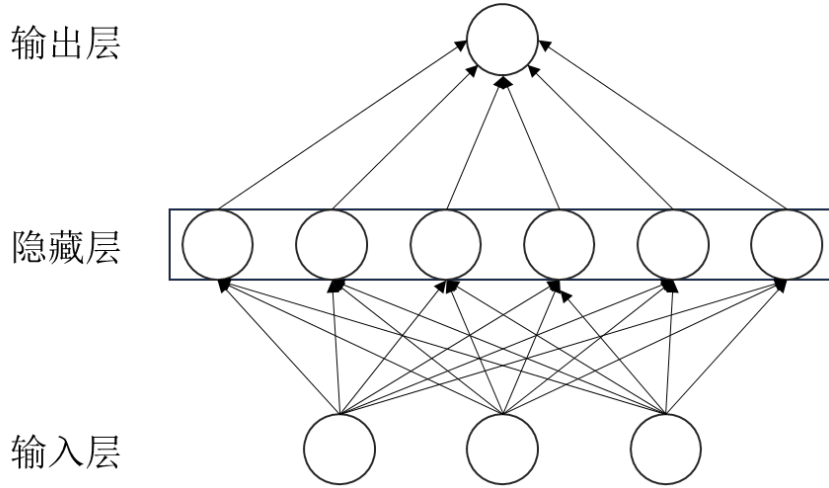


图 2.1 神经网络结构示意图

前馈神经网络通过如下公式进行信息的前向传播:

$$z^{(l)} = W^{(l)} \cdot a^{(l-1)} + b^{(l)} \quad (2.5)$$

$$a^{(l)} = f_l(z^{(l)}) \quad (2.6)$$

其中, L 表示神经网络的层数, $f_l(\cdot)$ 表示第 l 层神经元的激活函数, $W^{(l)}$ 表示第 $l-1$ 层到第 l 层的权重矩阵, $b^{(l)}$ 表示第 $l-1$ 层到第 l 层的偏置, $z^{(l)}$ 表示第 l 层的净输入, $a^{(l)}$ 表示第 l 层的净输出, 特殊的, 当 l 为 1 时, a 可以视为输入数据 x 。常见激活函数如下:

- Sigmoid 函数: 将输出压缩到 $[0,1]$ 区间, 常用于二分类任务的输出层, 但容易导致梯度消失问题。

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2.6)$$

- tanh 函数: 输出范围为 $[-1,1]$, 在零均值数据上表现更好, 但仍可能出现梯度消失。

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (2.7)$$

- ReLU 函数 (Rectified Linear Unit): 最常用的激活函数, 因其非饱和性质有效缓解了梯度消失问题。

$$f(z) = \max(0, z) \quad (2.8)$$

- Leaky ReLU 函数: 引入负斜率参数, 避免 ReLU 在 $z \leq 0$ 时完全失活。

$$f(z) = \begin{cases} z, & z > 0 \\ \alpha z, & z \leq 0 \end{cases} \quad (2.9)$$

数据从输入层逐层传递至输出层后, 得到预测值, 之后模型则会通过损失函数计算预测误差, 并逆向调整权重, 优化模型。损失函数主要用于衡量不同网络输出 a 与目标值 \hat{a} 的差距, 用于优化目标函数, 不同任务使用的损失函数不同:

- 均方误差 (MSE): 用于回归问题。

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.10)$$

- 交叉熵损失: 用于分类问题。

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.11)$$

显然, 在利用神经网络预测资产超额收益问题中, 损失函数应使用均方误差。逆向调整权重的过程又叫做反向传播, 核心步骤如下:

1. 计算输出层误差 $\delta^{(2)}$, 即损失对输出层输入的梯度。

$$\delta^{(2)} = \frac{\partial \mathcal{L}}{\partial z^{(2)}} = (\hat{y} - y) \odot g'(z^{(2)}) \quad (2.12)$$

其中, \odot 表示逐元素乘法, $g'(z)$ 表示输入层激活函数的导数, 例如 Sigmoid 的导数为 $g(z)(1 - g(z))$ 。

2. 计算隐藏层误差 $\delta^{(1)}$ 。

$$\delta^{(1)} = (W^{(2)T} \delta^{(2)}) \odot f'(z^{(1)}) \quad (2.13)$$

其中 $f'(z^{(1)})$ 为隐藏层激活函数的导数（如 ReLU 导数为 1 或 0）。

3. 利用误差项计算权重和偏置的梯度。

$$\frac{\partial L}{\partial W^{(2)}} = \delta^{(2)} a^{(1)T} \quad (2.14)$$

$$\frac{\partial L}{\partial b^{(2)}} = \delta^{(2)} \quad (2.15)$$

$$\frac{\partial L}{\partial W^{(1)}} = \delta^{(1)} x^T \quad (2.16)$$

$$\frac{\partial L}{\partial b^{(1)}} = \delta^{(1)} \quad (2.17)$$

4. 使用梯度下降法更新参数：

$$W^{(k)} \leftarrow W^{(k)} - \eta \frac{\partial L}{\partial W^{(k)}} \quad (2.18)$$

$$b^{(k)} \leftarrow b^{(k)} - \eta \frac{\partial L}{\partial b^{(k)}} \quad (2.19)$$

其中， η 为学习率 Learning Rate。

当然，上述模型的反向传播优化流程仅为最基础标准的优化流程，在算法的不断应用中，一系列辅助提升模型预测效果的方法被提出。

2.2.2 其他辅助优化方法

1. 归一化

现有研究普遍认为，训练样本的数量和质量对深度学习模型的泛化能力对未见数据的预测性能具有决定性作用。数据归一化等预处理方法对预测结果有显著影响，在样本量较小时，其作用尤为突出。数据归一化可以消除量纲差异，加速收敛，提高模型稳定性。常用归一化方法有三种：最小-最大归一化、Z-Score 归一化、批量归一化。

最小-最大归一化（Min-Max Scaling）可以将数据缩放到 $[0,1]$ 或 $[-1,1]$ 区间，适用于数据分布有明确边界、需要保留原始数据相对关系的任务。该法对异常值敏感，极端值会导致大部分数据压缩到狭窄区间，新数据超出原范围时需重新计算。公式如下：

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.20)$$

Z-Score 标准化（Standardization）可以将数据缩放到均值为 0、标准差为 1 并保留原始分布形态。该法适用于数据近似正态分布且模型假设数据均值为 0 的情况，对异常值鲁棒性较强，且新数据可动态调整，无需固定范围。公式如下：

$$X_{std} = \frac{X - \mu}{\sigma} \quad (2.21)$$

批量归一化（Batch Normalization）是深度学习中的归一化技术。其原理是对每个神经层的输入按批次进行标准化。公式如下：

$$\hat{X} = \gamma \frac{X - \mu_{batch}}{\sqrt{\sigma_{batch}^2 + \epsilon}} + \beta \quad (2.22)$$

其中 μ_{batch} , σ_{batch} 为批次统计量, γ , β 为可学习参数。批量归一化方法能缓解模型内部协变量偏移, 加速训练, 且允许更大学习率, 提升模型对初始化敏感型的鲁棒性, 适用于全连接层、卷积层。

2. 样本抽样

抽样策略对模型的学习效果有重要影响, 研究表明, 在不平衡数据场景下, 合理的抽样策略可以有效改善模型的预测能力, 使其能够从更具代表性的样本中学习 (Zhu 等, 2018)。样本抽样是从总体数据集中选择子集的过程, 其核心目标是降低计算成本 (避免全数据集训练的硬件资源压力)、提升模型泛化性 (通过数据多样性抑制过拟合)、适应动态场景 (处理流式数据或数据分布偏移)。样本抽样有如下方法:

- 简单随机抽样 (Simple Random Sampling): 每个样本以等概率被选中, 无任何偏好, 适用于数据分布均匀的基准数据, 公式如下:

$$P(x_i \in S) = \frac{M}{N}, \forall i \in 1, 2, \dots, N \quad (2.23)$$

- 分层抽样 (Stratified Sampling): 按类别或特征划分层级, 每层按比例抽取样本, 能够保证类别均衡, 适用于长尾分布数据 (如医疗图像分类), 公式如下: 设数据分为 K 个层级, 第 k 层样本数为 N_k , 则抽样数:

$$M_k = \text{round} \left(M \times \frac{N_k}{N} \right) \quad (2.24)$$

- 小批量抽样 (Mini-Batch Sampling): 遍历所有数据, 划分为固定大小的批次, 逐批训练。其损失函数近似为:

$$\mathcal{L}(\theta) \approx \frac{1}{M} \sum_{(x,y) \in S} \ell(f_{\theta}(x), y) \quad (2.25)$$

在深度学习中, 样本抽样还可以有特殊应用: 类别不平衡时可以选择加权抽样, 即在损失函数中赋予不同类别权重; 可以根据模型训练状态实时改变抽样策略, 例如设置高损失样本在后续批次中被抽中的概率提升、使用强化学习 Agent 决策抽样权重。

3. 正则化技术

正则化是防止深度学习模型过拟合的关键, 在金融数据低信噪比的背景下更为重要, 可分为 L1/L2 惩罚、早停法和集成学习三种。

L1/L2 惩罚是在损失函数中增加参数范数惩罚项，例如：

$$\mathcal{L} = MSE + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta\|_2^2 \quad (2.26)$$

L1 正则化 (λ_1) 推动部分权重归零，实现特征选择，例如筛选出对因子暴露贡献显著的财务指标（如现金流波动率）。L2 正则化 (λ_2) 约束权重幅度，防止单一特征过度主导模型，提升泛化能力。具体到在因子模型中，L1 正则化可自动剔除冗余因子，如高度相关的动量和反转指标；L2 正则化平滑因子暴露的时间序列，避免极端值，例如危机时期的 Beta 突变。

早停法 (Early Stopping) 通过监控验证集的损失变化，在连续 K 个 epoch 损失未改善时终止训练，从而避免模型过拟合噪声，同时保留对经济周期的泛化能力。

集成学习 (Ensemble Learning) 通过训练多个独立模型，这些模型可能由于不同初始化方式或子样本选择，导致模型超参数不同。最终，集成方法通过平均或投票集成预测结果。集成模型的优点是在极端市场环境下表现更稳健，提高预测的可靠性，在相关研究中也有许多应用。Gu 等 (2020) 使用平均的方法对同一模型进行集成，提升模型预测准确性。Zhao 等 (2022) 使用机器学习中的 stacking 方法，使用多种线性和非线性模型提升其对股票超额收益的预测能力。

4. 自适应矩估计

深度学习的优化过程需要在收敛速度和稳定性之间取得平衡，而金融数据的高噪声使这一挑战更加严峻。自适应矩估计 (Adam) 在提升训练效率和模型稳定性方面发挥着重要作用。

自适应矩估计 (Adam) 的更新规则如下：

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.27)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.28)$$

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon} \quad (2.29)$$

其中 g_t 为梯度， β_1 和 β_2 是估计的指数衰减率， η 为学习率。Adam 是一种自适应学习率的优化算法，其优点包括：

- 自适应性：能够自适应地调整学习率，根据梯度的一阶矩估计和二阶矩估计自适应地调整每个参数的学习率，使得算法能够适应各种不同的数据集和模型。
- 收敛速度快：利用梯度一阶矩估计和二阶矩估计来更新参数，能够有效地降低训练过程中的震荡和噪声，从而提高收敛速度。

- 对内存需求低：对比其他优化算法更低，能够处理大规模数据和深度神经网络模型。
- 鲁棒性强：对初始值的选择比较鲁棒，对于不同的初始化方式都能够得到比较好的结果。

Gu 等（2020）三层的 NN 模型中，隐藏层为 3 层，节点数分别为 32, 16, 8，使用的激活函数为 ReLU，输出层为线性激活，并使用到了随机梯度下降、早停法、集成学习、Adam、批量归一化等方法优化训练过程。

此外，参数优化也能提高模型预测能力，例如江雨燕等（2024）使用 SSA 算法进行网络参数学习，实现训练集参数最优化，能够达到提升预测模型稳健性的目的。

2.3 循环神经网络算法

上一节分析了本文的基准模型 NN-3 模型和一系列优化算法，本节主要介绍后续使用的三种时间序列深度学习算法。

2.3.1 RNN

循环神经网络（recurrent neural network, RNN）是一种具有内部循环的神经网络，可以通过迭代序列元素，并维护一个包含与已处理元素相关信息的内部状态来处理序列。类比来说，人类阅读文字时，就是以过去阅读的内容作为基础模型，随着新的文字输入，不断更新阅读到的文字。RNN 的核心思想是通过循环连接（Recurrent Connection）维护一个隐藏状态（Hidden State），该状态会随着时间步（Timestep）更新，并捕捉序列中的时序依赖关系。

简单循环神经网络（Simple RNN）其核心结构如下：

- 输入序列：假设序列长度为 T ，每个时间步的输入为 $x_t (t = 1, 2, \dots, T)$ 。
- 隐藏层：每个时间步的隐藏状态 h_t 是记忆单元，初始状态 h_0 通常为全 0 向量，后续 h_t 由 x_t 和 h_{t-1} 共同决定。
- 输出序列：每个时间步可以生成输出 o_t ，通常用于预测或传递到下一层。

对每个时间步 t ，隐藏状态的更新过程可用如下公式表示：

$$h_t = \text{activation}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (2.30)$$

其中， W_{xh} 为输入到隐藏层的权重矩阵， W_{hh} 为隐藏层到隐藏层的权重矩阵， b_h 为隐藏层偏置， activation 是激活函数，通常为 \tanh 和 ReLU 。

输出过程为：

$$o_t = \text{activation}(W_{ho}h_t + b_o) \quad (2.31)$$

其中 W_{ho} 为隐藏层到输出层的权重矩阵， b_o 为输出层偏置。

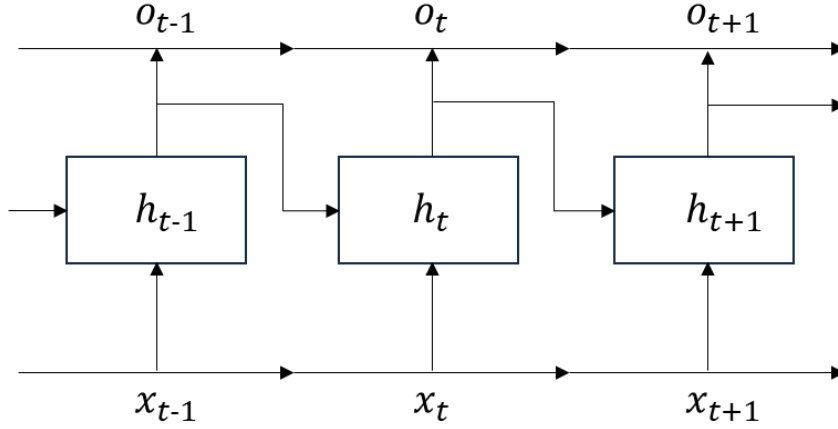


图 2.2 简单 RNN 结构示意图

简单 RNN 因其结构相对简洁，仅由一个隐含状态（Hidden State）进行信息传递，因此参数较少，计算开销低，易于实现和训练。此外，由于网络结构较轻量，简单 RNN 在小规模数据集或计算资源受限的环境下表现较好，适用于某些对长期依赖要求不高的任务，如基础的时间序列建模和短时预测。

理论上，简单 RNN 虽然能够在 t 时间保留许多步之前的输入信息，但由于梯度消失问题的存在，当不断向网络添加层时，损失函数收敛缓慢甚至下降，网络会变得不可训练，且记忆能力有限，隐藏状态容易被新输入覆盖，难以捕捉长序列中的远距离关联。

梯度消失问题出现的原因有多种。第一个是链式法则的连乘效应，指简单 RNN 这类模型在反向传播时，梯度值随着网络层数或时间步的增加呈指数级衰减，导致浅层网络的参数几乎无法更新的现象。梯度是多个导数的乘积：

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h_T} \cdot \frac{\partial h_T}{\partial h_{T-1}} \cdot \frac{\partial h_{T-1}}{\partial h_{T-2}} \cdots \frac{\partial h_1}{\partial W} \quad (2.32)$$

每个局部梯度 $\frac{\partial h_t}{\partial h_{t-1}}$ 的值小于 1，多次连乘后整体梯度趋近于 0。

第二个是激活函数的饱和性，一些激活函数的导数值在两端接近 0，如 \tanh 函数的导数为 $1 - \tanh^2(x)$ ，当输入较大时导数趋近于 0。

第三个是 RNN 的循环权重共享，其循环权重矩阵 W_{hh} 在每个时间步被重复使用，如果 W_{hh} 的特征值小于 1，梯度会指数级衰减；如果大于 1，可能引发梯度爆炸。

2.3.2 LSTM

为解决上述梯度消失问题, Hochreiter 和 Schmidhuber 在 1997 年最早提出了 LSTM 模型, 首次成功解决了梯度消失问题, 且能够有效捕捉长序列中的长期依赖问题。LSTM 跨多个时间步长传输信息的方式类似于: 一条“传送带”与正在处理的数据序列平行运行, 序列中的信息可以在任何一点跳到传送带上, 传输到以后的时间步, 并在需要时完好无损地跳下来。其核心思想是: 通过增加细胞状态 (Cell State) 和门控机制 (Gates), 动态控制信息的保存与遗忘。细胞状态可以理解为“传送带”, 贯穿所有时间步, 储存长期信息, 通过线性传递和门控机制更新, 缓解梯度消失。

LSTM 流程及门控机制的经典结构如下:

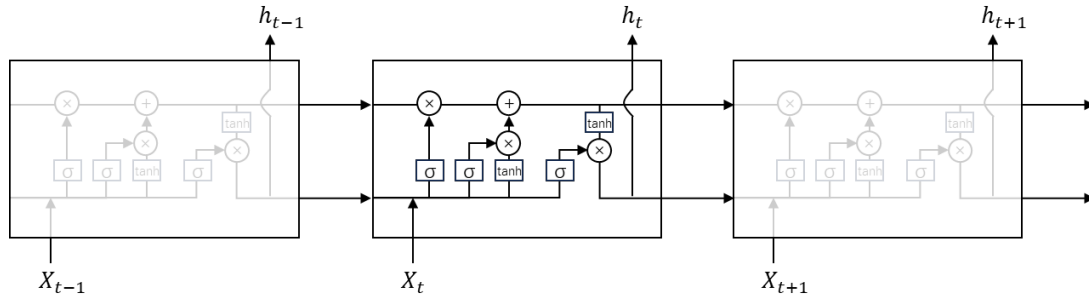


图 2.3 LSTM 结构示意图

- 遗忘门 (Forget Gate) 决定从细胞状态中丢弃哪些历史信息, 公式如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.34)$$

根据当前输入 x_t 和前一隐藏状态 h_{t-1} , 利用 sigmoid 函数 (即 σ) 生成 0-1 的权重, 控制细胞状态 C_{t-1} 的遗忘比例, 得到 C_t^{temp} 。此时 $C_t^{temp} = f_t \cdot C_{t-1}$, 如果 $f_t = 0$, 则完全遗忘; 若 $f_t = 1$, 则全部保留。

- 输入门 (Input Gate) 决定将哪些新信息存入细胞状态, 首先计算输入门权重 i_t , 然后生成候选记忆 \tilde{C}_t , 更新细胞状态 C_t , 具体公式如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.35)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.36)$$

$$C_t = C_t^{temp} + i_t \odot \tilde{C}_t \quad (2.37)$$

其中 \tanh 函数可以将候选记忆压缩至 $(-1, 1)$, 增强非线性表达能力, 第三个公式是用加法操作, 使得梯度可以沿时间步线性传递。

- 输出门 (Output Gate) 决定当前隐藏状态的输出内容, 首先计算输出门权重 o_t , 生成隐藏状态 h_t , 再将 h_t 和 C_t 传递到下一时间步, 具体公式如下:

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \quad (2.38)$$

当然,整个过程中用于门控和状态更新的激活函数主要使用 Sigmoid 和 tanh, 因为 Sigmoid 函数的输出范围为 (0, 1), 适合表示开关比例, tanh 输出范围为 (-1, 1), 对称且梯度较大, 适合表示有正负方向的信息, 缓解梯度消失。实际工程中激活函数可以根据需求调整, 本文选择使用经典设计完成长序列任务, 避免因激活函数改变引入不稳定风险。

LSTM 具有如下优点: 第一, 时序建模深度突破: LSTM 通过细胞状态的总想传导机制, 理论上可捕捉超过 1000 步的长期依赖关系 (Hochreiter & Schmidhuber, 1997), 相较传统 RNN 的 5-10 步有效记忆长度实现数量级突破。第二, 门控动态选择机制: 三重门控结构形成微分数据流控制器, 使得模型在股票价格预测任务中可以达到 54.3% 的准确率。

尽管 LSTM 通过门控机制缓解了梯度消失问题, 但其结构复杂, 仍存在一定据典型。首先, LSTM 依赖遗忘门、输入门和输出门这三个门控单元, 以及细胞状态 (Cell State) 来控制信息流动, 这些组件需要大量参数, 使得模型在训练和推理过程中计算开销较大。其次, 门控单元之间存在较强的依赖关系, 导致模型的调试和优化难度增加, 同时工程实现相对复杂, 训练过程较难收敛。最后, 在处理长序列或大规模数据时, LSTM 的计算效率较低, 对硬件资源的需求更高, 可能影响实际应用的可扩展性。

2.3.3 GRU

GRU (Gated Recurrent Unit) 由 Cho 等在 2014 年提出, 是更精简版本的 LSTM。其主要在两个方面进行修改: 一是将 LSTM 的输入门和遗忘门合并为更新门 (Update Gate) 减少参数; 二是移除细胞状态, 直接通过隐藏状态传递信息, 简化记忆管理。GRU 通过牺牲部分精细控制能力, 换取了更快的训练速度和更少的资源消耗。

GRU 单元在时间步 t 内包含以下组件:

- 更新门 (Update Gate), 主要控制当前隐藏状态中有多少信息来自前一状态 h_{t-1} , 有多少来自候选状态 \tilde{h}_t , 替代了 LSTM 的输入门和遗忘门, 实现“遗忘”与“新增”的联合控制。

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (2.39)$$

其中, h_{t-1} 为前一隐藏状态, x_t 为当前输入, $[h_{t-1}, x_t]$ 表示这两项的拼接向

量。

- 重置门 (Reset Gate), 主要控制历史信息对候选状态的贡献 (过滤无关信息), 公式如下。

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2.40)$$

- 生成候选隐藏状态 (Candidate Hidden State), 公式如下。

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t] + b) \quad (2.41)$$

在更新门和重置门完成门控计算后, 重置门使用 $r_t \odot h_{t-1}$ 选择性地遗忘历史信息, 若 $r_t \approx 0$, 则忽略 h_{t-1} , 候选状态仅依赖当前输入 x_t , 若 $r_t \approx 1$, 则完全保留历史信息。

- 更新候选隐藏状态 (Candidate Hidden State), 公式如下。

$$\tilde{h}_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.42)$$

若 $z_t \approx 0$, 则表示保留历史信息, 隐藏状态几乎保留前一状态 h_{t-1} , 若 $z_t \approx 1$, 则表示添加新信息, 隐藏状态主要由候选状态 \tilde{h}_t 更新。

由于更新候选隐藏状态时, GRU 使用了加法更新机制, 使得梯度可以通过加法传递, 从而有效避免梯度在长序列中出现指数级衰减的问题, 提升了长程依赖的学习能力。同时, GRU 的更新门和重置门能够根据输入动态调整信息的保留与遗忘程度, 而非采用固定比例的衰减方式, 使模型在不同时间步具有更强的适应性。此外, 与 LSTM 相比, GRU 通过减少门控单元的数量, 合并遗忘门和输入门, 从而减少了约三分之一的参数。这不仅提高了训练速度, 降低了内存占用, 还使得 GRU 更易于实现, 适用于计算资源有限的场景。

尽管 GRU 在计算效率和参数优化方面具有优势, 但也存在一定的局限性。首先, 由于 GRU 没有独立的细胞状态 (Cell State), 其长期记忆能力相对较弱, 在处理超长序列任务时, 可能会丢失关键信息, 影响对细节的捕捉。其次, GRU 的更新门同时负责信息的遗忘与新增, 而不像 LSTM 采用独立的遗忘门和输入门进行精细控制, 因此在某些复杂任务中, GRU 的信息更新机制可能不如 LSTM 灵活, 限制了对长期依赖关系的建模能力。

2.4 性能评估

为了评估模型对股票超额收益的预测效果, 我们将样本外 R^2 设置为:

$$R_{oos}^2 = 1 - \frac{\sum_{(i,t) \in \mathcal{T}_3} (r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{(i,t) \in \mathcal{T}_3} r_{i,t+1}^2} \quad (2.43)$$

其中 \mathcal{T}_3 表示仅对样本外测试集的评估拟合值。

这个 R^2 指标的分母是未经去均值处理的超额收益平方和。在许多样本外预测应用中,预测结果通常会与历史平均收益进行比较,但历史平均收益的噪声非常高,导致其人为地降低了“良好”预测表现的标准。为避免这一问题,本文假设未来超额收益为 0,将 0 作为基准预测值。这个做法同样参考了 Gu 等(2020)的研究。

若 $R_{oos}^2 > 0$,则说明模型的预测误差平方和小于实际收益的平方和,即模型预测效果优于“未来收益为 0”的基准;若 $R_{oos}^2 < 0$,说明模型的预测误差平方和大于实际收益的平方和,即模型预测效果弱于基准。

3 数据与实证分析

3.1 数据来源与预处理

本文主要研究 1990 至 2021 年间的美股市场，数据来源有三个部分：

1. 证券价格研究中心 (CRSP)：该库由芝加哥大学商学院研究生院于 1960 年成立，是证券领域极具权威的数据库。该库广泛收录了美国上市公司的股票价格和交易数据，提供自 1926 年以来美国上市公司单日、月度、年度的股票价格、收益率、红利、交易信息等数据。本文使用到的是美国排名前三的交易所 (NYSE、AMEX、NASDAQ) 中的上市股票的月度收益率数据，以及 30 日美国国库券利率作为无风险利率，用于计算个股当期超额收益。

2. Dacheng Xiu 的个人主页：该网站中提供了该学者的所有研究成果，包括其补充材料。在该网站中，可以找到 Gu 等 (2020) 使用到的 94 个个股因子实证数据，并已经更新至 2021 年 12 月，因子具体内容及处理方式稍后将进行说明。

3. Amit Goyal 的个人主页：该网站中提供了该学者的所有研究成果及其补充材料，在该网站中，可以找到 Welch 和 Goyal (2008) 中使用的宏观因子原始数据及更新至 2023 年的数据，因子具体内容及处理方式将稍后进行说明。

在本文中，我们选择沿用 Gu 等 (2020) 构建的标准化数据集，这一决策基于以下两个角度考虑：

1. 保持学术一致性与可比性：该数据集经过学术界严格的验证，其变量定义、缺失值处理与标准化方法已获得学界广泛认可。采用统一数据集可确保研究结论与既有文献在一定程度上的可比性，避免因因子选择的差异导致结果偏差过大，从而精准聚焦于模型方法本身的创新性评估。

2. 可复现性与科学严谨性：实证研究中，数据构造的主观性（如异常值剔除规则、行业分类标准等）往往会导致结果的不同。通过采用公开透明的标准化数据集，探讨循环神经网络模型的预测效果，有助于为后续研究者提供一定的增量贡献。

Gu 等 (2020) 的研究时间范围覆盖 1957 至 2016 年，共 60 年。然而，由于计算资源的限制，以及深度学习模型相较于传统机器学习模型训练时间更长，本

文将研究时间范围设定为 1990 至 2021 年。在此期间，样本中的股票数量共计 24729 只，月均股票数量超过 6800 只，以确保数据量充足，同时提高研究的可行性和计算效率。

在股票筛选方面，本文遵循 Gu 等（2020）的处理方法，未对价格低于 5 美元或高于 10 美元的股票进行剔除，旨在保留尽可能广泛的资产池。较大的样本规模不仅有助于提高模型的泛化能力，还能在训练过程中有效降低过拟合风险。最终筛选得到的股票数量变化趋势如图 3.1 所示。

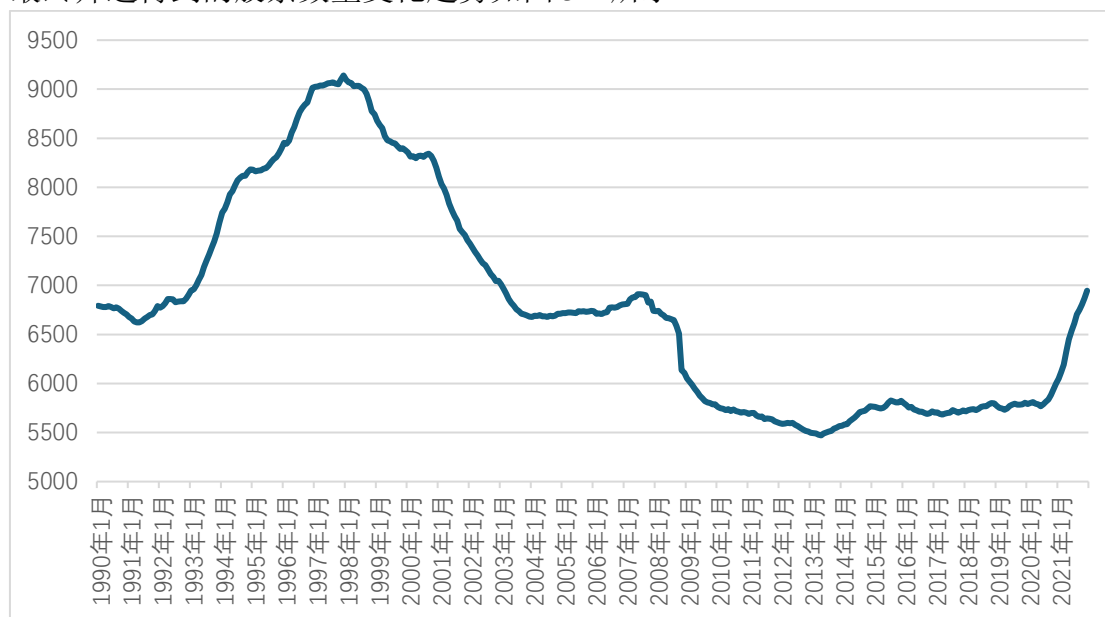


图 3.1 1990-2021 年间股票数量变化趋势

3.2 因子分类

个股收益率的影响因素可以分为宏观因子和微观因子两大类，二者共同作用于股票市场，决定了个股的表现。

3.2.1 宏观因子

宏观因子指来自整体经济环境和市场体系的因素，这些因素往往对所有公司或行业产生广泛影响。本文使用的 8 个宏观因子来自 Welch 和 Goyal (2008)，其计算方式及其经济含义如下。

(1) 股息价格比率 (dp)

$$dp = \log(\text{标普 500 股息的 12 个月移动总和}) - \log(\text{标普 500 指数收益})$$

股息价格比率是衡量股息回报率与股市整体收益之间的差异的指标，通过计算标准普尔 500 指数的股息总和与其收益的比率，反映了市场的股息回报情况。当股息收益率较高时，通常表示市场相对较为低迷，投资者对于股息收益有更强的偏好。

(2) 收益价格比率 (ep)

$$ep = \log(\text{标普 500 的 12 个月移动总收益}) - \log(\text{标普 500 指数价格})$$

收益价格比率衡量市场整体盈利能力，高 EP 表明企业盈利相对市场价格较高；在经济扩张期通常下降，衰退期上升，可作为经济周期的领先指标。

(3) 账面市值比 (bm)

账面市值比率 (b/m) 是道琼斯工业平均指数的账面价值与市场价值的比率。本质反映指数成分股整体资产负债表质量与市场定价的偏离程度。该指数反映了市场对价值股与成长股的风格偏好，高 bm 指数偏向“价值型”，成分股多为传统制造业、能源等低估值板块，反映市场对资产安全性的偏好；低 bm 指数偏向“成长型”，科技、消费等板块占优，体现流动性宽松下的增长预期。

(4) 净股权扩张 (ntis)

$$\text{净股权扩张} = (\text{新股发行总额} - \text{股票回购总额}) / \text{市场总价值}$$

该指标通过衡量股权融资的净流量，揭示市场供需结构变化。高 ntis 表明净融资增加，在供给端企业加速 IPO 或增发，市场扩容，需求端投资者风险偏好高，愿意承接新股，通常对应牛市中后期；低 ntis 表明净回购占优，企业大规模回购来支撑股价。

(5) 美国 3 月期国库券利率 (tbl)

作为无风险收益的基准，反映央行货币政策立场。

(6) 期限利差 (tbs)

$$\text{期限利差} = \text{长期政府债券收益率} - \text{国库券利率}$$

反映市场对未来利率走势的预期。tbs>0 代表长期利率大于短期利率，市场预期未来经济扩张；tbs<0 代表长期利率小于短期利率，预示经济衰退。tbs 扩大，预示企业融资成本下降、信贷扩张，tbs 倒挂，预测市场进入衰退期。

(7) 违约利差 (dfy)

$$\text{违约利差} = \text{BAA 级公司债收益率} - \text{AAA 级公司债收益率}$$

是信用周期的“压力计”，衡量市场对信用风险的定价。BAA 级公司债代表高风险端，AAA 级公司债代表低风险端。正常状态下，dfy 较小，代表经济稳定，

dfy 升高时，往往代表系统性风险爆发，经济周期出现衰退信号。

(8) 股票波动率 (svar)

股票波动率 = 标普 500 指数收益的月度方差

衡量市场整体波动性，方差值越大，代表市场价格波动越剧烈。高 svar 代表市场恐慌加剧，低 svar 代表市场过度乐观。

这八个宏观因子从不同维度刻画了市场状态和经济环境，多年来被学者广泛应用于资产定价、收益预测与风险管理中。通过结合这些因子，投资者可更全面地评估市场风险与收益潜力。

3.2.2 微观因子

微观因子共 94 个，其中 61 个因子每年更新，13 个因子每季度更新，20 个因子每月更新，详细的因子介绍及来源见附录 1。本文结合相关参考文献，将所选因子划分为以下七类：

(1) 估值类因子 (16 个)

主要包括规模、账面市值比、股息率、现金价格比、销售价格比、销售现金比率等。估值类因子在资产定价中扮演核心角色，其核心作用在于揭示市场定价偏差、衡量安全边际并反映投资者对企业未来现金流的预期等。学术研究证实，以市净率 (PB，即账面市值比的倒数)、股息率为代表的估值指标具有显著超额收益：Fama-French 三因子模型表明低市净率组合 (HML) 长期跑赢市场，Campbell 和 Shiller (1988) 的研究则证明估值指标对长期收益率具有预测能力。例如市净率，高 PB 反映市场对企业未来盈利能力的乐观预期或行业特性（如科技股），但需警惕资产泡沫；低 PB 则可能暗示行业周期底部或资产低估，需验证资产质量。股息率方面，高股息率通常体现企业稳定的现金流能力与成熟行业特征，但也可能隐含增长停滞；低股息率多出现于成长型企业，反映其高再投资需求，需关注自由现金流充足性。

(2) 风险类因子 (11 个)

主要包括 beta、回报波动率、流动性波动率、收益波动率、现金波动率等。风险类因子在资产定价与组合管理中承担风险度量与收益预测的核心功能，其重要性体现在揭示资产波动特性、优化风险收益配比及构建风险对冲策略。学术研究表明，以 beta 系数、回报波动率为代表的风险因子对资产收益率具有显著解释力：CAPM 模型证实 beta 因子能解释系统性风险溢价 (Sharpe, 1964)，Fama-

French 五因子模型将投资风格风险纳入框架。例如 beta 系数，高 beta 值（大于 1）表明资产与市场波动高度相关（如科技股 $\beta \approx 1.5$ ），低 beta 值（小于 1）则体现防御属性（如公用事业股 $\beta \approx 0.7$ ）。回报波动率（Return Volatility）方面，高波动率（如纳斯达克 100 指数年波动率 25%+）反映资产价格剧烈波动，低波动率（如标普 500 红利指数年波动率 15%）则暗示稳定收益特征。实践中需结合流动性风险与尾部风险（VaR）指标综合评估，注意波动率聚类效应与结构性变化对模型有效性的影响。

（3）动量与趋势类因子（10 个）

主要包括 1/6/12/36 月动量、行业动量，以及资产/经营资产/债务/销售的增长。动量与趋势类因子在量化投资中承担趋势识别与惯性收益捕捉的核心功能，其重要性源于市场对信息反应不足或过度的行为特征。学术研究证实，以中期动量（6-12 个月）、销售增长率为代表的趋势因子具有显著超额收益：Carhart(1997) 将动量因子纳入四因子模型解释了 31 个基点的收益差异。例如 12 月动量因子，高动量值（过去 12 个月收益率分位数前 20%）反映资产价格趋势延续性（如新能源板块动量值常超 80%），低动量值则暗示反转可能（如周期性行业动量值低于 20%）。销售增长率方面，高增长率（如科技股年均增长 25%+）体现企业市场份额扩张与营收质量提升，低增长率（如传统制造业低于 5%）可能预示需求疲软或竞争加剧。研究发现，动量策略在所有资产类别中均表现出显著的超额收益，且在流动性充足市场表现更佳（Asness, 2013）；资产增长因子与未来股票收益显著负相关，且在经济复苏期可能因市场预期偏差和过度投资而强化其效应（Cooper 等, 2008）。

（4）盈利能力与质量类因子（17 个）

主要包括资产/股东权益/投资回报率、毛利率、营业利润率、现金生产率、收益公告回报等。盈利能力与质量类因子在量化投资中承担企业基本面评估与可持续增长识别的核心功能，其重要性在于揭示企业真实盈利质量、现金流健康度及经营效率。学术研究证实，以净资产收益率（ROE）、毛利率为代表的因子具有显著超额收益：Fama-French 五因子模型将盈利能力（RMW）与投资风格（CMA）纳入框架（2015）。具体如净资产收益率，高 ROE（大于 15%）反映企业高效利用股东权益创造价值（如白酒行业 ROE 常超 25%），低 ROE（小于 8%）则可能暗示资本配置低效或行业竞争激烈。毛利率（Gross Profit Margin）方面，高毛利率（如软件行业大于 70%）体现产品差异化与定价权优势，低毛

利率（如零售业小于 20%）可能面临成本压力或同质化竞争。

（5）运营效率与资产配置类因子（13 个）

主要包括资产周转率变化、资本支出变化、销售存货比率变化、这就与固定资产比率、房地产持有等。运营效率与资产配置类因子在量化投资中承担企业资源优化评估与经营效率分析的功能，其重要性体现在揭示企业资产利用效率、现金流管理能力及战略调整方向。具体如资产周转率变化，高周转率增长（如零售业同比提升 15%）反映企业资产利用效率改善，低周转率增长（如制造业低于 5%）可能暗示产能过剩或需求疲软。资本支出变化方面，高增长（如科技行业年均超 20%）体现企业扩张意愿与未来产能布局，低增长（如公用事业低于 8%）则可能表明收缩战略或现金流压力。

（6）资本结构与杠杆类因子（8 个）

主要包括杠杆、担保债务、可转换债务、流通股变化、资本支出增长、公司投资等，资本结构与杠杆类因子在量化投资中承担企业财务风险评估与资本配置效率分析的核心功能，其重要性体现在揭示企业负债水平、融资策略及风险收益特征。具体如杠杆率，高杠杆（如房地产行业 $D/E > 100\%$ ）反映财务风险较高但潜在收益放大效应，低杠杆（如科技行业 $D/E < 30\%$ ）则体现财务稳健性但可能抑制增长弹性。

（7）流动性与市场微观结构类因子（8 个）

主要包括股票换手率、交易量、买卖价差、流动比率、速动比率、流动比率变化百分比等。流动性与市场微观结构类因子在量化投资中承担资产定价效率评估与交易成本控制的核心功能。具体如股票换手率，高换手率（如科技股月均大于 30%）反映交易活跃度高但可能伴随投机泡沫（如 2021 年新能源板块换手率峰值达 60%），低换手率（如公用事业小于 5%）则暗示交易清淡或长期持有特征；买卖价差，高价差（如小盘股大于 2%）体现流动性供给不足（如科创板新股首日价差中位数达 5.2%），低价差（如蓝筹股小于 0.5%）表明市场深度良好。

（8）成长与创新类因子（7 个）

主要包括研发增加、存货变化、研发与市值比率、普通股股东权益增长、税收费用变化等。成长与创新类因子在量化投资中承担企业长期增长潜力评估与技术壁垒识别的核心功能。学术研究表明，以研发强度、股东权益增长为代表的成长因子具有显著选股有效性：Chan 等（2001）发现研发投入高的公司表现出

显著的超额收益，可能反映市场对其长期价值的低估。具体如研发与市值比率，高比率（如医药行业大于 15%）体现企业技术领先优势（如恒瑞医药 2023 年研发占比 24%），低比率（如公用事业小于 3%）可能面临技术迭代风险；普通股股东权益增长，高增长（如新能源行业年均大于 30%）反映企业通过内生或外延式扩张提升资本基础，低增长（如传统制造业小于 8%）可能暗示增长瓶颈或资本配置效率不足。

（9）事件驱动与特殊情境类因子（5 个）

主要包括股息初始、股息遗漏、罪恶股票、上市年数、收益增加次数。事件驱动与特殊情境类因子在量化投资中承担市场异常捕捉与短期定价偏差修正的功能，其重要性源于行为金融学框架下的事件冲击反应不足（Jegadeesh 和 Titman, 1993）与市场摩擦带来的套利机会。学术研究表明，以股息事件、罪恶股票为代表的特殊情境因子具有显著超额收益：Baker 和 Wurgler（2004）发现股息宣告日存在显著的正异常收益。例如股息初始，高股息初始（如标普 500 成分股首次派息）通常引发 2-3 个月的正收益，低股息初始可能反映公司现金流紧张或成长优先策略。

在上述微观因子中，绝大多数是利用财报数据得到的，往往会延迟向公众发布。为避免前瞻性误差，月度因子特征最多延迟 1 个月，季度因子值至少延迟 4 个月，年度因子值至少延迟 6 个月。此外，对于该月缺失的特征值，使用每只股票每个月的横截面中位数替换。

除了微观层面的因子本身外，为观测个股微观因子与宏观因子间的交互作用，本文将微观因子与宏观因子分别相乘：

$$z_{i,t} = x_{i,t} \otimes c_{i,t} \quad (3.1)$$

其中， $c_{i,t}$ 是形状为 $P_c \times 1$ 的向量，表示微观因子， $x_{i,t}$ 是形状为 $P_x \times 1$ 的向量，表示宏观因子； $z_{i,t}$ 是形状为 $P \times 1$ 的向量（ $P = P_c P_x$ ）。

同时，本文使用标准行业分类（sic）代码中的前两位数字作为虚拟变量，表示个股的行业特征，得到额外的 75 个行业虚拟变量。此处与 Gu 等（2020）得到的 74 个行业虚拟变量不同，原因是在实际处理过程中，发现部分股票的 SIC 代码为空值，对于此类样本，本文未进行删除，作为单独的因子加入到特征向量中。总体来看，因子的总数为 $94 \times (8+1) + 75 = 921$ 。

之后，相关因子数据采用 Kelly 等（2019）的标准化处理方法，在每个截面对每个因子进行横截面排名，并将排名映射到 $[-1, 1]$ 中，代替原本因子值。这

种处理方式有如下好处：

第一，不同因子的原始量纲差异显著，其分布状态也并不相同（如图 3.2），直接输入模型会导致数值范围较大的因子主导梯度更新过程，通过排序归一化处理，所有因子被约束在相同尺度内，能够有效消除量纲差异对模型训练的干扰，同时保留因子间的序数关系。相较于传统的 Z-Score 标准化，该方法的排序转换对分布形态的敏感性更低，更适用于金融数据中普遍存在的尖峰厚尾分布特征。

第二，金融数据常受极端值影响，传统标准化方法易受离群值影响，而排序归一化方法确保了结果对异常值不敏感，并且因子值考虑横截面中的相对大小，避免随时间变化，因子大小的含义发生变化，能提升模型在不同市场环境下的泛化能力。对于处理后的因子空值，采用截面中位数填充，由于已经归一化至 $[-1,1]$ ，故直接使用 0 进行填充。

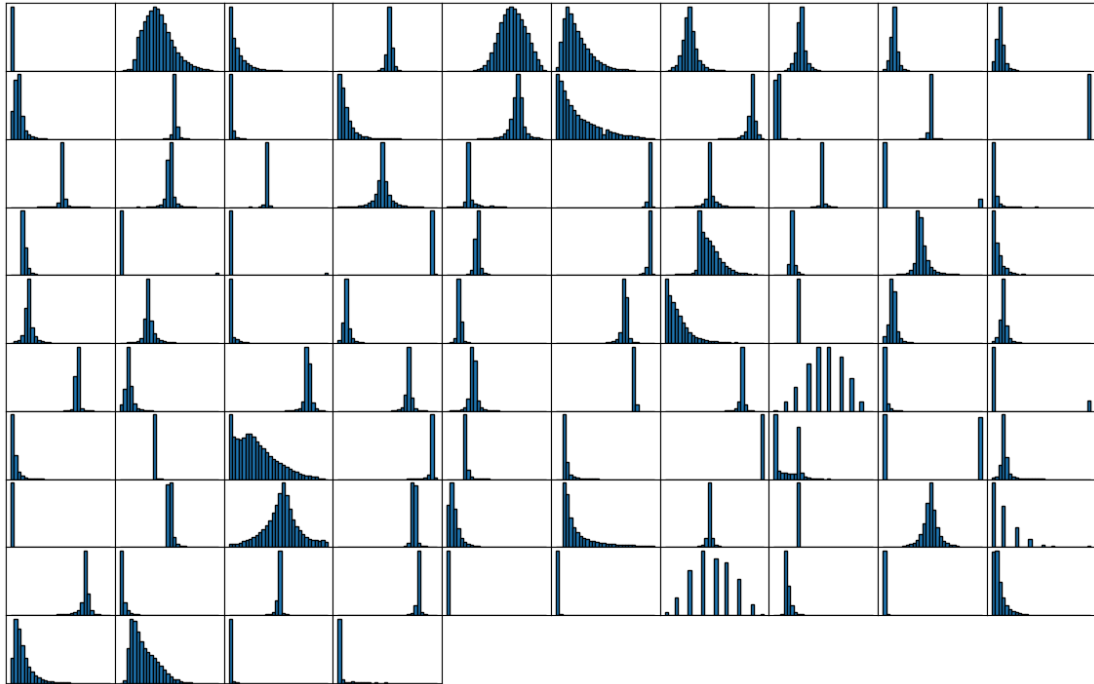


图 3.2 94 个微观因子分布图

通过数据库得到的收益率计算出的是个股当期超额收益，为得到下期超额收益作为预测值，在处理时对每个股票代码采用 `shift(-1)` 操作。

由于样本量较大，在完成上述处理后，按照因子年份逐年保存因子数据，后续读取数据时采用自定义的数据加载函数 `load_data(year)`，将各年的特征数据和目标变量分别合并。这样操作有如下好处：第一，使数据管理更加高效：将庞大

数据分割成较小的年度文件，尽管在一定程度上会增加读取时间，但便于存储、备份和检索，降低单次加载数据时的内存负担。第二，便于分阶段加载和处理：在后续增量训练的过程中，只需加载对应年份的数据即可，无需一次性处理所有数据，从而提高数据处理和模型更新的效率。第三，增强数据追溯与异常检测能力：当模型训练过程出现异常（如内存溢出导致进程终止）时，可以快速定位到具体年份的数据进行排查，若同时保存了模型参数，则无需重新从头开始训练，在断点读取数据及模型即可。

3.3 模型设计

本节主要介绍用于训练模型的数据集的划分方法，以及不同模型在构造层面的细节，使用 Python 语言完成。

1. 时间框架

在机器学习训练中，为了遏制过度拟合，通常将整个样本划分为三个不相交的子样本，以保持数据的时间顺序。第一个子样本（训练集 **Train**）用于估计受一组特定优化超参数值约束的模型。第二个子样本（验证集 **Validation**）子样本用于优化超参数，根据训练集中的估计模型为验证样本中的数据点构建拟合值，根据验证样本中的误差计算目标函数，然后选择超参数以优化验证目标。验证集拟合不是真正的样本外拟合，往往用于模型调整，而调整又是估计的输入。因此，第三个子样本（测试集 **Test**）用于评估方法的样本外性能。

本文的训练集-验证集-测试集时间框架如下：

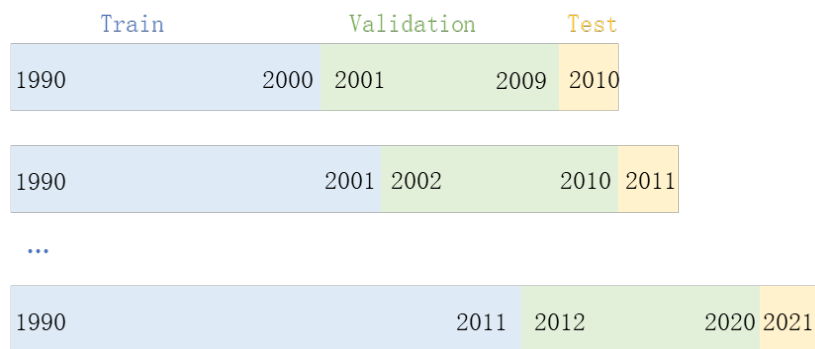


图 3.3 训练集-验证集-测试集时间框架划分

由于深度学习模型是计算密集型的，因此本文避免每月递归重新拟合模型，而选择一年更新一次，因为大多数的因子每年更新一次。例如，第一期，以 1990-

2000 的数据作为训练集，2001-2009 的数据作为验证集，预测 2010 年的超额收益；第二期，以 1990-2001 年的数据作为训练集，2002-2010 年的数据作为验证集，预测 2011 年的超额收益……最后，得到 2010-2021 年的预测超额收益序列。

2. 增量训练

在实际训练过程中，若完全按照上述框架进行，由于因子量较多、时间跨度长，数据累积增加，内存需求要求极高，无法完成训练。作为替代方法，本文选择增量训练方式，方法如下：

在初始训练阶段，即 1990-2000 的数据作为训练集，2001-2009 的数据作为验证集，预测 2010 年的超额收益中，得到训练集后，随机采样 70% 的样本作为后续训练集。这个做法的好处除了降低资源消耗、提高训练效率外，还可以模拟更广泛的市场环境，提高模型对未知数据的适应能力，并且在样本量充足的情况下，舍弃部分数据并不会显著降低模型的学习能力。总的来说，随机采样是值得采用的。

之后，对训练集和验证集采用 **Standard Scaler** 进行归一化处理，确保数据特征处于相同尺度。虽然在数据预处理阶段已经进行了因子的截面排序归一化，但使用 **Standard Scaler** 进行全局归一化可以让输入数据在网络中更加平衡，进一步消除跨年度的数据尺度不一致问题，提高训练稳定性和泛化能力。

构造集成学习框架，即使用五个参数设置相同的网络模型，对每个模型设置不同的随机种子，以保证模型多样性，训练时分别训练五个模型，最终得到五个的预测结果。每个模型采用早停法（**Early Stopping**）监控验证集损失，防止过拟合，在初始训练阶段设置 50 个 **epoch** 的训练。利用辅助函数 **ensemble_predict** 获取各模型对测试数据的预测结果，并计算单模型及整体集成模型的 R^2 评价指标。集成学习和早停法的优点在理论部分已经进行了介绍，在此不过多赘述。初始训练完成后，将中间变量与模型参数保存至本地文件，方便后续增量训练。

在增量更新时，为了避免完全丢弃历史数据造成模型偏离原有特征，本文选择保留少部分上期训练集同时加入新一期训练数据的方式构造新的数据集，可以表示为：

$$all_train = [all_train * ratio, new_train] \quad (3.2)$$

其中，初始时 **all_train** 即为初始训练的训练集，**ratio** 为保留比例。例如，当 2011 年作为样本外测试集时，**new_train** 则为 2001 年的数据，这样在初始模型参数的基础上完成增量训练。这种训练方式在节省计算资源和时间的同时，灵活调

整保留比例，减少因数据分布突变带来的性能波动，在保持模型稳定性和适应新数据之间取得了平衡。在增量更新中，依旧使用早停法提高预测效率。

计算因子重要性和模型调参是深度学习算法中运算量较大的环节，学者们往往调用数十台机器的计算机集群实现（李斌等，2019）。由于算力有限，本文并未进行因子重要性分析和模型参数及结构的调整，所使用到的深度学习框架及节点设置也尽量简单。

整体结构介绍完毕，接下来介绍最重要的模型设置部分，后续深度学习框架使用 Tensorflow 包进行搭建。

NN-3 作为本文的基准模型，相关设置均与 Gu 等（2020）保持一致，保证一定程度上的可比性。模型首先采用学习率衰减策略（函数为 `ExponentialDecay`），设置初始学习率为 $1e-3$ ，属于中等水平，适用于大多数优化问题，以便快速收敛。每 1000 个训练步后，学习率衰减 0.9 倍，确保训练后期学习率降低，防止梯度振荡。之后采用 `Sequential` 函数构建三层（`Dense` 层）神经网络，激活函数均为 `ReLU`，神经元数分别为 32，16，8，并添加 L2 正则化限制权重大小。最后输出层神经元数为 1，无激活函数（即线性回归）。每个隐藏层后都有 `BatchNormalization()` 进行批量归一化。

模型编译时采用 Adam 优化器，损失函数使用均方误差（MSE），评价指标采用绝对平均误差（MAE）。

单独介绍 NN 模型是因为在输入数据维度方面，它与后续的循环神经网络模型并不相同。NN 模型的输入数据形状要求为二维：（样本数 `batch_size`，特征数 `num_features`），在本文中，特征数即为因子数，共 921 个，样本数取决于当期截面股票数据的个数。

时间序列模型作为本文的研究内容，其在设置上尽量与基准模型有所关联。在学习率衰减策略方面，RNN、GRU、LSTM 与 NN-3 模型保持一致，在此不进行赘述。

其他模型结构设置中，将 NN3 的第一个神经元数为 32 的隐层，分别修改为 32 个神经元的 `SimpleRNN` 层、32 个神经元的 `LSTM` 层、32 个结点的 `tf.keras.layers.GRU` 层。

上述模型中，关于 L2 正则化的相关设置均相同。

在输入数据维度方面，这三个循环神经网络模型要求的输入形状均为三维：（样本数 `batch_size`，时间步 `time_steps`，特征数 `num_features`），在基础模型训练

中, 时间步暂时设置为 1, 表示每个样本仅用当期因子值预测下一期超额收益率。

3.4 实证结果

3.4.1 模型预测效果对比

根据 3.3 节的模型设置, 经训练后得到模型性能评估的 R_{oos}^2 :

表 3.1 模型性能评估的 R_{oos}^2

	模型 1	模型 2	模型 3	模型 4	模型 5	集成
NN-3	0.158%	0.299%	0.168%	-0.128%	0.319%	0.345%
RNN	-1.226%	0.837%	0.202%	0.229%	-0.234	0.272%
LSTM	0.375%	-8.045%	-1.187%	-3.450%	-2.615%	-0.016%
GRU	0.094%	-2.225%	-0.311%	-0.132%	-0.156%	0.137%

最后一列集成表示每种模型的集成学习模型, 模型 1-5 均单独训练的模型。除 LSTM 的集成模型 R_{oos}^2 为负外, 其它三种模型均为正, 表明 LSTM 的预测误差的平方和超过了实际超额收益的平方和, 可能的原因如下: 第一, 存在过拟合现象: LSTM 在训练集上可能过度拟合了噪声, 导致样本外泛化能力差; 第二, 时序模式缺失: 由于时间步设置为 1, 导致数据中缺乏有效的时序依赖性, 而 LSTM 因复杂结构引入了额外噪声, 导致预测效果较差。从集成学习模型的 R_{oos}^2 结果来看, 另外三种模型的 R_{oos}^2 均为正, 表明模型捕捉到了部分有效信息, 但预测能力有限 (数值小)。NN-3 的表现最好, 与 Gu 等 (2020) 的结果相呼应, $R_{oos}^2 = 0.345\%$, 表明在时间步=1 时, 非线性静态特征比时序特征更适用于当前任务; RNN 与 GRU 的表现依次变差, R_{oos}^2 分别为 0.272% 与 0.137%, 因为二者的结构逐渐复杂, 超参数调整的空间越来越大, 在时间步为 1 时可能存在参数冗余, 逐渐过拟合噪声, 而结构简单的模型反而泛化能力会更强。

从前五列的结果来看, 同一结构训练的五個模型在样本外测试中差异较大, 既有预测效果较好的, 也有较差的, 这是因为单个模型可能因随机初始化、数据采样或训练中的随机性, 导致预测结果不稳定。而在集成平均后, 单个模型在不同方向上的预测误差可能相互抵消, 从而整体预测的准确性得到了提升。此外, 单个模型可能对训练数据过拟合, 导致样本外表现不佳, 集成多个模型的预测可能可以缓解过拟合, 提升泛化能力。

前五列结果中, R_{oos}^2 表现最好的为 RNN 的模型 2, 为 0.837%, 但此时 RNN

模型的其他模型表现较差，并如 NN-3 模型（仅模型 4 表现较差）有更加稳定的表现。从一定程度上可以认为，RNN 模型的潜力更大，但仍需细致调整其结构与参数。

3.4.2 选股策略模型构建与结果

在上一节中，通过滑动窗口模型训练的过程，已经得到 2010-2021 年的模型预测超额收益。本节将构造选股策略模型，根据每个集成模型的预测超额收益率从低到高将股票分类为十个组合，分别构造等权重和价值加权的投资组合，并计算各组合的月均预测收益率（月均超额收益+当月无风险利率）、月均收益率、标准差、年化夏普比率。

表 3.2 价值加权组合表现

NN-3					RNN				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1 (Low)	0.26%	0.74%	4.14%	0.620	1 (Low)	-0.66%	0.68%	5.74%	0.412
2	0.43%	0.84%	3.98%	0.727	2	-0.02%	0.88%	4.74%	0.648
3	0.53%	1.05%	3.93%	0.928	3	0.27%	0.77%	4.44%	0.603
4	0.62%	1.04%	4.10%	0.880	4	0.61%	0.94%	4.03%	0.811
5	0.73%	1.28%	4.19%	1.057	5	0.87%	0.96%	4.19%	0.797
6	0.84%	1.33%	4.75%	0.970	6	1.07%	1.26%	4.22%	1.036
7	0.96%	1.26%	4.84%	0.901	7	1.26%	1.32%	4.56%	1.001
8	1.11%	1.35%	5.35%	0.873	8	1.47%	1.25%	4.93%	0.876
9	1.31%	1.35%	5.85%	0.800	9	1.81%	1.27%	5.11%	0.863
10 (High)	1.72%	1.18%	7.16%	0.572	10 (High)	2.54%	1.42%	6.23%	0.786
H-L	1.46%	0.44%	4.79%	0.319	H-L	3.20%	0.73%	3.78%	0.672

3 数据与实证分析

(续) 表 3.2 价值加权组合表现

LSTM					GRU				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1 (Low)	0.13%	0.45%	5.59%	0.277	1 (Low)	-0.68%	0.61%	4.64%	0.457
2	0.38%	0.84%	4.73%	0.613	2	-0.10%	0.69%	4.38%	0.540
3	0.49%	1.04%	3.95%	0.908	3	0.38%	0.78%	4.11%	0.655
4	0.57%	0.96%	3.98%	0.831	4	0.69%	1.07%	4.19%	0.887
5	0.65%	1.18%	4.13%	0.987	5	0.89%	1.13%	4.26%	0.921
6	0.74%	1.21%	4.31%	0.970	6	1.06%	1.26%	4.56%	0.960
7	0.83%	1.26%	4.42%	0.991	7	1.23%	1.22%	4.80%	0.883
8	0.98%	1.16%	5.09%	0.786	8	1.44%	1.12%	5.08%	0.766
9	1.43%	1.31%	5.72%	0.790	9	1.81%	1.69%	5.68%	1.029
10 (High)	3.16%	1.40%	6.80%	0.714	10 (High)	2.87%	1.74%	7.69%	0.783
H-L	3.03%	0.96%	4.28%	0.772	H-L	3.33%	1.12%	5.56%	0.700

从结果中可以看出，样本外价值加权投资组合性能与深度学习模型预测收益在整体趋势上一致，实际收益通常随模型预测结果单调增加，表明模型的预测能力在一定程度上能够有效捕捉资产回报的横截面特征，从而影响投资组合的表现。虽然存在偶尔例外，如 NN-3 模型的第 7 组、RNN 模型的第 8/9 组，并未完全符合这一单调递增的趋势，可能与市场噪声、模型误差有关。

H-L 策略表现最好的模型是 GRU，月均收益率为 1.12%（年化 13.44%），月度标准差为 5.56%（年化 19.23%），夏普比率为 0.202（年化 0.7），表明 GRU 模型的组间分类效果最强，能较为有效地区分不同资产组合地收益水平。

3 数据与实证分析

表 3.3 等权重组合表现

NN-3					RNN				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	0.27%	0.74%	3.61%	0.710	1	-0.78%	0.22%	6.27%	0.118
2	0.42%	0.78%	3.96%	0.686	2	-0.02%	0.52%	3.99%	0.454
3	0.52%	0.77%	4.54%	0.589	3	0.24%	0.75%	4.17%	0.627
4	0.62%	0.80%	4.78%	0.582	4	0.60%	0.92%	5.04%	0.634
5	0.73%	0.89%	5.07%	0.606	5	0.87%	1.02%	5.01%	0.703
6	0.84%	0.99%	5.31%	0.644	6	1.07%	1.16%	5.18%	0.779
7	0.97%	1.07%	5.59%	0.658	7	1.26%	1.13%	5.22%	0.748
8	1.12%	1.21%	6.04%	0.693	8	1.48%	1.24%	5.43%	0.790
9	1.34%	1.25%	6.58%	0.662	9	1.84%	1.40%	6.00%	0.807
10	1.78%	1.49%	7.69%	0.672	10	2.73%	1.63%	7.20%	0.783
H-L	1.51%	0.75%	5.00%	0.520	H-L	3.51%	1.41%	3.49%	1.399
LSTM					GRU				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	0.11%	0.23%	6.17%	0.128	1	-0.74%	0.30%	3.56%	0.294
2	0.38%	0.61%	4.80%	0.436	2	-0.19%	0.65%	4.39%	0.509
3	0.48%	0.87%	4.21%	0.717	3	0.36%	0.84%	4.84%	0.603
4	0.57%	0.98%	4.56%	0.748	4	0.68%	1.01%	4.84%	0.721
5	0.65%	1.09%	4.86%	0.776	5	0.89%	1.08%	4.95%	0.759
6	0.74%	1.14%	4.87%	0.811	6	1.06%	1.14%	5.19%	0.759
7	0.83%	1.28%	5.03%	0.883	7	1.23%	1.19%	5.62%	0.731
8	0.99%	1.20%	5.45%	0.762	8	1.45%	1.27%	5.96%	0.738
9	1.49%	1.24%	6.08%	0.703	9	1.86%	1.11%	6.48%	0.592
10	3.56%	1.36%	7.28%	0.644	10	2.99%	1.41%	7.87%	0.624
H-L	3.46%	1.13%	3.74%	1.046	H-L	3.73%	1.11%	5.23%	0.738

在等权重投资组合中，实际收益同样随深度学习预测效果单调增加（同样少部分除外），H-L 策略表现最好的模型是 RNN，月均收益率为 1.14%，月度标准差为 3.49%，夏普比率为 0.404，较价值加权组合的最好表现更好。

由于在价值权重组合中，投资组合按照股票市值加权，大市值股票权重大，

能够主导组合表现，而在等权重组合中，每支股票的权重相同，所以小盘股的影响力会增加。从价值权重到等权重，组合收益上升，说明小盘股的收益较高，符合“小盘股效应”。小盘股效应指在长期投资中，小市值公司（小盘股）的股票平均收益显著高于大市值公司（大盘股）的现象。Banz（1981）首次系统性发现，美国股市中最小市值的股票在 1936-1975 年期间的年化收益比最大市值股票高出约 4.8%，后续全球市场研究也普遍观察到类似现象，但强度和持续性存在差异；后 Fama-French 三因子模型将小盘股效应归为市值因子，认为超额收益是对风险的合理补偿。

做多 group 为 10 的组合和做空 group 为 1 的组合得到的价值加权累计收益曲线及多空组合收益曲线如图 3.4、图 3.5，可以看出四种模型的 group 为 1 的组合累积收益表现几乎相同，NN-3 的 group 为 10 的累积收益表现最差。四个模型多空组合表现均优于 SP500 指数表现，说明深度学习构造的投资组合能够获取超越基准的收益。

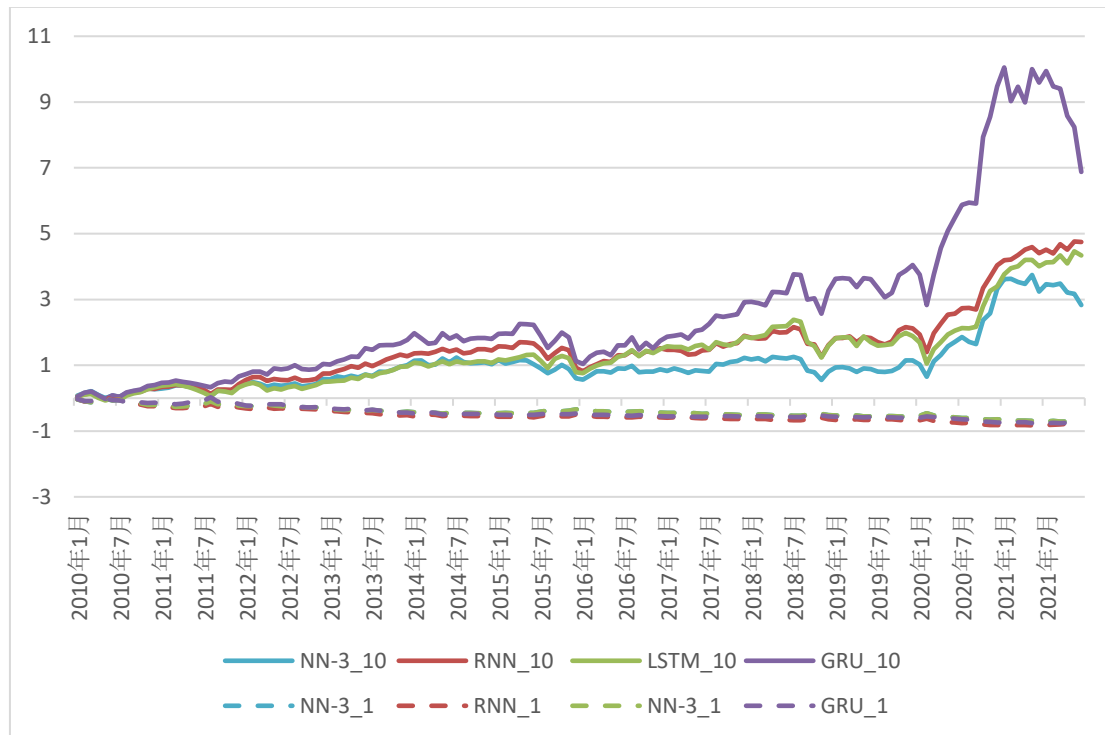


图 3.4 时间步为 1 时的分组价值加权累积收益

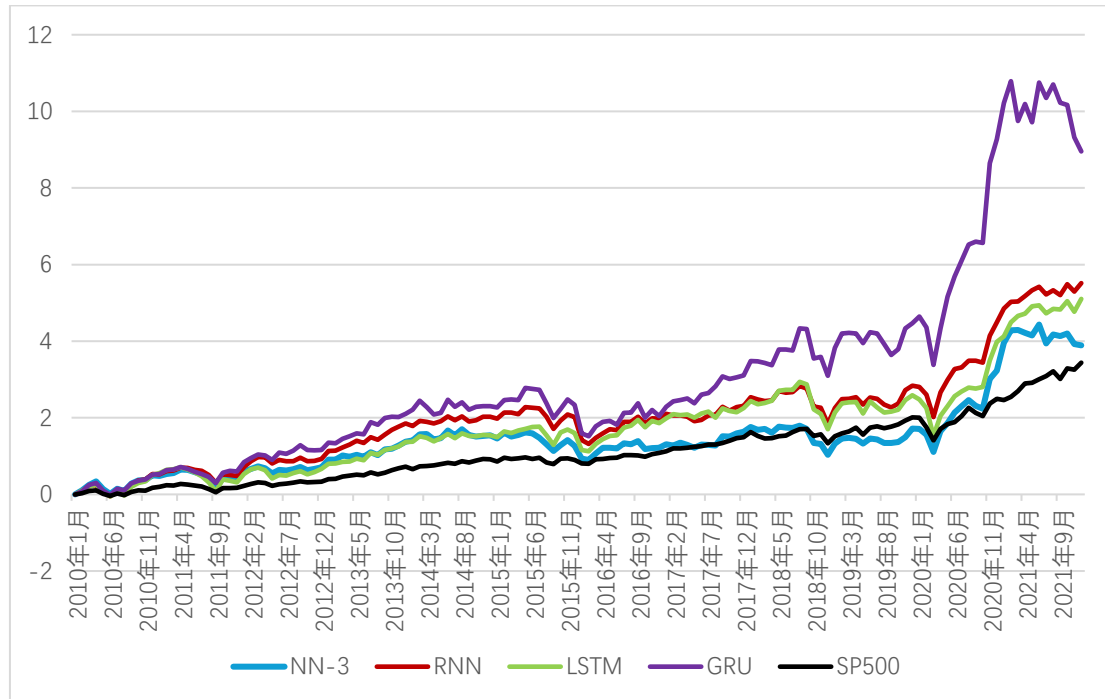


图 3.5 时间步为 1 时的多空组合累积收益率

3.4.3 稳健性检验

上述实证分析中，分组数设置为 10，当样本量有限时，10 分组可能会导致有效信号被稀释；从行为金融学视角来看，投资者对 5 分类体系的决策准确率显著高于 10 分类，反映认知负荷与决策质量的倒 U 型关系，5 分组天然形成“优秀-良好-中性-较差-极差”的心里锚点，增强策略信号的传播效率。因此，在稳健性检验部分，通过调整分组数量（由划分 10 组变为划分 5 组）进行参数敏感性测试，得到分组后 H-L 组合表现的变化如表 3.4。

通过对比其分组后的变化，可以发现：

（1）出现收益压缩现象。由于分组减少，导致高-低组合收益差收窄，极端值被稀释，H-L 月均收益率变化幅度下降，符合预期。虽然收益差收窄，但模型仍保持正向预测能力，证明策略具有实际经济价值。

（2）不同分组下模型间的表现不变。价值权重组合中 GRU 在 group=5 时仍然具有最高的月均收益率 1.10%，等权重组合中 RNN 在 group=5 时仍有最高的月均收益率 1.14%。表示分组数量变化并未改变模型表现得相对排序，原始结论仍有统计稳健性。

3 数据与实证分析

表 3.4 五分组组合表现

价值权重								
	group=10				group=5			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR
NN-3	1.46%	0.44%	4.79%	0.092	1.04%	0.50%	3.16%	0.159
RNN	3.20%	0.73%	3.78%	0.194	2.31%	0.47%	3.09%	0.152
LSTM	3.03%	0.96%	4.28%	0.223	1.68%	0.60%	3.37%	0.180
GRU	3.33%	1.12%	5.56%	0.202	2.32%	1.10%	3.63%	0.304
等权重								
	group=10				group=5			
	Pred	Avg	SD	SR	Pred	Avg	SD	SR
NN-3	1.51%	0.75%	5.00%	0.15	1.21%	0.61%	4.03%	0.152
RNN	3.51%	1.41%	3.49%	0.404	2.69%	1.14%	3.13%	0.365
LSTM	3.46%	1.13%	3.74%	0.302	2.28%	0.88%	3.26%	0.270
GRU	3.73%	1.11%	5.23%	0.213	2.89%	0.79%	3.88%	0.204

3.5 进一步分析

3.5.1 完整数据的模型效果对比

在基础的模型训练过程中，数据采用抽样数据，为了避免数据抽样使结果变差，在模型验证时采用完整数据进行重新训练。由于使用完整数据后，对内存空间占用变大，所以无法保存前期历史数据，每次增量模型训练仅使用当期新增的数据集，例如：初始训练使用 1990-2000 作为训练集后，第一个滑动窗口增量模型训练则使用 2001 年的数据。集成后的模型预测效果如下：

表 3.5 完整数据的模型预测 R_{os}^2

	抽样历史数据集成	完整数据集成模型
NN-3	0.345%	-0.386%
RNN	0.272%	-1.843%
LSTM	-0.016%	-3.095%
GRU	0.137%	-1.597%

NN-3 的 R_{oos}^2 由 0.345% 变为 -0.386%，RNN 由 0.272% 变为 -0.183%，LSTM 由 -0.016% 变为 -0.395%，GRU 由 0.137% 变为 -1.597%，证明在原始实证中使用的抽样历史数据集成方法的模型预测能力优于完整数据集成模型。进一步证明了，完整数据中引入了更多的低信噪比样本，稀释有效信号，采用抽样方法恰好避免这一点。此外，也说明增量训练确实存在覆盖历史知识的现象，完全丢弃历史数据进行增量训练会使模型底层参数发生较大变化。

3.5.2 时间步为 3 的模型效果对比

在基础的模型训练过程中，设置的时间步（time_step）均为 1，发现此时，结构越复杂的模型预测效果越差。为了检验时间步的长度对模型预测效果的影响，本文使用时间步为 3 的数据重新训练各个模型。

首先，修改训练集的特征。选择时间步为 3 而不是更长主要基于如下两点原因：第一，时间步为 3 的含义是，对每支股票，使用连续 3 个月的因子值来预测下一月的超额收益，输入形状为（样本数 batch_size, 3, 921），现实中。由于时间步由 1 变为 3，数据集的大小也扩大了三倍，仍旧使用基础模型的设置对内存要求过高，在后续的模型训练中，保留的数据比例 *ratio* 会适当调低。

其次，对于输入数据为二维的 NN-3 模型，将三个时间步的所有特征同时作为输入特征，其输入形状为（样本数 batch_size, 3*921=2763）。四种模型的其他设置不变，得到模型预测 R_{oos}^2 如下：

表 3.6 模型性能评估的 R_{oos}^2 -时间步为 3

	模型 1	模型 2	模型 3	模型 4	模型 5	集成
NN-3	0.217%	0.181%	0.195%	0.189%	0.252%	0.337%
RNN	0.005%	-5.614%	-1.126%	-0.520%	-0.845%	-0.480%
LSTM	0.280%	0.321%	0.145%	0.095%	0.398%	0.373%
GRU	-0.078%	-0.250%	-4.750%	0.050%	-0.362%	-0.245%

从表中最后一列的集成准确率可以看出，LSTM 模型的效果有了非常显著的提升， R_{oos}^2 由 -0.016% 增至 0.373%，成为了预测效果最好的模型。这说明 LSTM 可以通过提供的时间序列信息，区分噪声与有效信号，对噪声有更强的鲁棒性，例如预测收益率时，单日波动可能代表噪声，但连续三天的趋势可能反映真实模式。此外，由于 LSTM 的参数量大，时间步增加后模型拟合效果变好，可能与模型复杂度更加匹配数据复杂度匹配有关，说明在时间步为 1 时，LSTM 模型可能

存在欠拟合现象，而时间步为 3 时避免了欠拟合现象。

由于 NN-3 并不是专门为时间序列设计的，步长的增加在该模型中代表历史数据增加，其预测效果略有下降（由 0.345% 降至 0.337%），表明时间步增加并不会显著改善其预测效果，甚至可能因包含过多历史数据、多重共线性等原因导致效果变差。

RNN、GRU 的集成模型样本外预测能力显著下降，分别由 0.272%、0.137% 降至 -0.480%、-0.245%，原因可能如下：第一，可能存在梯度消失/爆炸问题：当时间步为 3 时，RNN/GRU 可能无法捕捉到第 1 期到第 3 期的关联，导致预测失效。第二，超参数未匹配：RNN/GRU 可能需要调整隐藏层大小、学习率或正则化参数来适配更长的时间步。若超参数未优化、隐藏层数不够，模型可能欠拟合或过拟合。

时间步为 3 时的模型训练稳定性有明显的改进，NN-3 与 LSTM 每个模型的预测准确率都为正值，RNN 与 GRU 每个模型的预测准确率都为负值。

3.5.3 时间步为 3 的选股模型构建与结果

接下来，利用时间步为 3 的模型超额收益预测结果，以上文相同的方式分别构造价值加权投资组合与等权重投资组合，结果如下：

表 3.7 价值加权组合表现-时间步为 3

NN-3					RNN				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	-0.08%	0.72%	6.88%	0.360	1	-1.21%	0.72%	4.89%	0.513
2	0.29%	0.93%	5.61%	0.576	2	-0.44%	0.84%	3.89%	0.746
3	0.44%	1.02%	4.47%	0.793	3	-0.03%	0.98%	4.04%	0.838
4	0.55%	0.97%	4.48%	0.747	4	0.29%	1.08%	4.23%	0.881
5	0.64%	1.32%	4.07%	1.119	5	0.57%	1.25%	4.39%	0.986
6	0.73%	1.23%	3.94%	1.086	6	0.86%	1.37%	4.56%	1.038
7	0.84%	1.18%	4.22%	0.968	7	1.19%	1.40%	4.70%	1.034
8	0.99%	1.59%	4.52%	1.218	8	1.60%	1.38%	4.78%	1.000
9	1.32%	1.32%	6.02%	0.761	9	2.20%	1.44%	5.71%	0.874
10	2.20%	1.30%	7.06%	0.639	10	3.46%	1.30%	6.99%	0.645
H-L	2.28%	0.59%	4.41%	0.460	H-L	4.67%	0.58%	5.72%	0.349

3 数据与实证分析

(续) 表 3.7 价值加权组合表现-时间步为 3

LSTM					GRU				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	-0.43%	0.77%	6.42%	0.418	1	-1.18%	0.42%	4.79%	0.301
2	0.15%	0.76%	5.02%	0.522	2	-0.39%	0.80%	4.90%	0.566
3	0.35%	1.05%	4.46%	0.817	3	0.09%	0.74%	4.10%	0.625
4	0.47%	1.20%	3.94%	1.055	4	0.41%	1.14%	3.93%	1.001
5	0.57%	0.92%	4.09%	0.778	5	0.66%	1.16%	4.29%	0.934
6	0.68%	1.27%	4.04%	1.091	6	0.92%	1.26%	4.23%	1.035
7	0.81%	1.34%	4.22%	1.099	7	1.22%	1.49%	4.93%	1.049
8	0.96%	1.48%	4.57%	1.123	8	1.68%	1.39%	5.59%	0.858
9	1.21%	1.50%	5.28%	0.988	9	2.41%	1.41%	6.20%	0.788
10	2.09%	1.41%	7.04%	0.694	10	3.66%	1.40%	7.25%	0.670
H-L	2.52%	0.64%	4.08%	0.540	H-L	4.84%	0.99%	5.73%	0.596

与时间步为 1 时的结果类似, H-L 策略表现较的模型依旧是 GRU, 月均收益率为 0.99%, 月度标准差为 5.73%, 年化夏普比率为 0.596。

表 3.8 等权重组合表现-时间步为 3

NN-3					RNN				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	-0.11%	0.32%	7.28%	0.153	1	-1.27%	0.23%	4.87%	0.165
2	0.29%	0.81%	5.27%	0.529	2	-0.45%	0.75%	4.24%	0.614
3	0.44%	0.84%	4.42%	0.660	3	-0.03%	0.99%	4.27%	0.800
4	0.54%	0.93%	4.36%	0.741	4	0.29%	1.06%	4.67%	0.785
5	0.64%	1.17%	4.55%	0.891	5	0.57%	1.08%	4.95%	0.753
6	0.73%	1.15%	4.50%	0.888	6	0.87%	1.14%	5.15%	0.765
7	0.84%	1.17%	4.79%	0.845	7	1.19%	1.20%	5.37%	0.772
8	1.02%	1.22%	4.98%	0.846	8	1.62%	1.43%	6.01%	0.823
9	1.39%	1.30%	5.67%	0.795	9	2.28%	1.41%	6.69%	0.729
10	2.44%	1.77%	7.41%	0.830	10	3.97%	1.41%	7.86%	0.623
H-L	2.54%	1.45%	3.90%	1.292	H-L	5.24%	1.18%	6.22%	0.658

3 数据与实证分析

(续) 表 3.8 等权重组合表现-时间步为 3

LSTM					GRU				
group	Pred	Avg	SD	SR	group	Pred	Avg	SD	SR
1	-0.52%	0.17%	6.77%	0.085	1	-1.25%	0.24%	4.96%	0.165
2	0.14%	0.65%	5.35%	0.420	2	-0.42%	0.68%	4.64%	0.506
3	0.35%	0.79%	4.39%	0.624	3	0.07%	0.99%	4.37%	0.782
4	0.47%	1.00%	3.97%	0.868	4	0.40%	1.10%	4.32%	0.880
5	0.57%	1.04%	4.31%	0.838	5	0.66%	1.07%	4.59%	0.810
6	0.68%	1.22%	4.75%	0.892	6	0.92%	1.19%	4.99%	0.827
7	0.81%	1.27%	5.06%	0.873	7	1.24%	1.19%	5.49%	0.753
8	0.97%	1.30%	5.31%	0.850	8	1.71%	1.40%	6.15%	0.786
9	1.30%	1.48%	5.90%	0.871	9	2.48%	1.43%	7.04%	0.706
10	2.38%	1.76%	7.62%	0.799	10	4.25%	1.41%	7.99%	0.609
H-L	2.91%	1.59%	3.68%	1.495	H-L	5.51%	1.17%	6.40%	0.633

在等权重投资组合中, 实际收益同样随深度学习预测结果单调增加 (同样少部分除外), H-L 策略表现最好的模型是 LSTM, 月均收益率为 1.59%, 月度标准差为 3.68%, 年化夏普比率为 1.495, 较价值加权组合的最好表现更好。

在时间步为 3 时, 同样存在小盘股效应, 各个模型从价值加权到等权重的过程中, 实际收益均有不同程度的上升。

group10 和 group1 的组合得到的价值加权累计收益曲线与时间步为 3 时的多空收益曲线如图 3.6 与图 3.7。

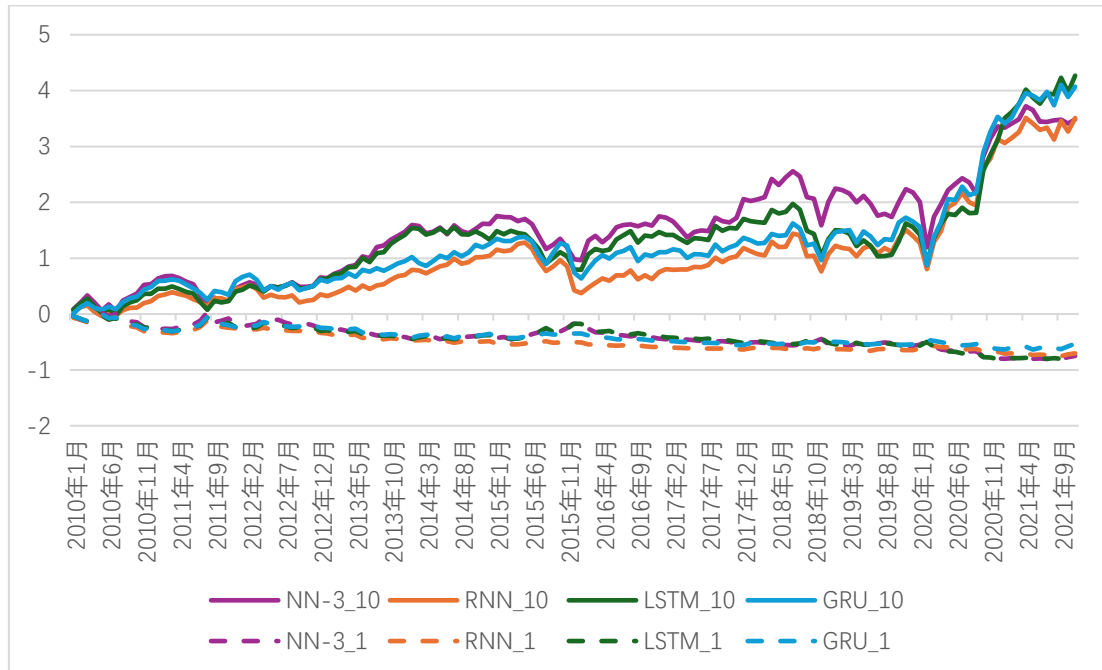


图 3.6 时间步为 3 时的分组价值加权累积收益

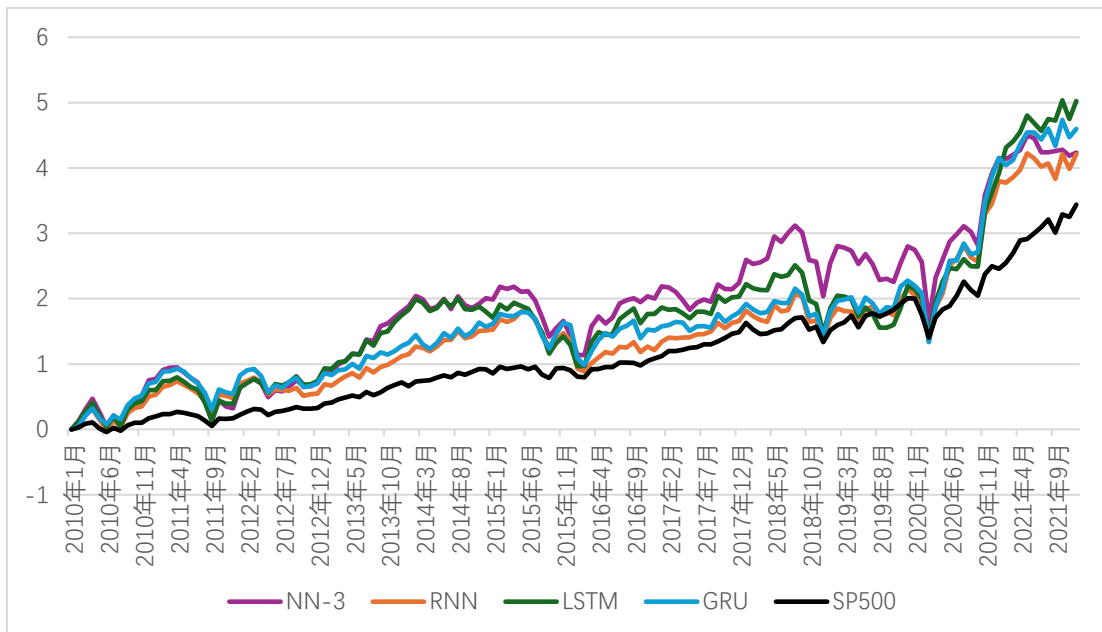


图 3.7 时间步为 3 时的多空组合收益

在累计收益曲线中，四个模型的 group1 的累积收益表现相差不大，group10 的累积收益曲线中，LSTM 整体表现较好，RNN 的累积收益表现最差。

截至 2021 年 11 月，时间步为 1 的 group10 的累积收益分别为 3.17、4.76、

4.46、8.24，时间步为 3 的 group10 的累积收益分别为 3.49、3.51、4.27、4.07，仅有 NN 模型累积收益升高；虽然 LSTM 模型在时间步为 3 时预测准确率及稳定性有大幅度提升，但累积收益角度的效果略有下降。

四个模型的累积收益均高于 SP500 指数收益，证明在时间步为 3 时，深度学习模型仍能获取超越基准的收益。

为了对比时间步增加对每个模型的累积收益的影响，令 $\text{Cumulative return}_{\text{time_step}=3} - \text{Cumulative return}_{\text{time_step}=1}$ 得到下图：

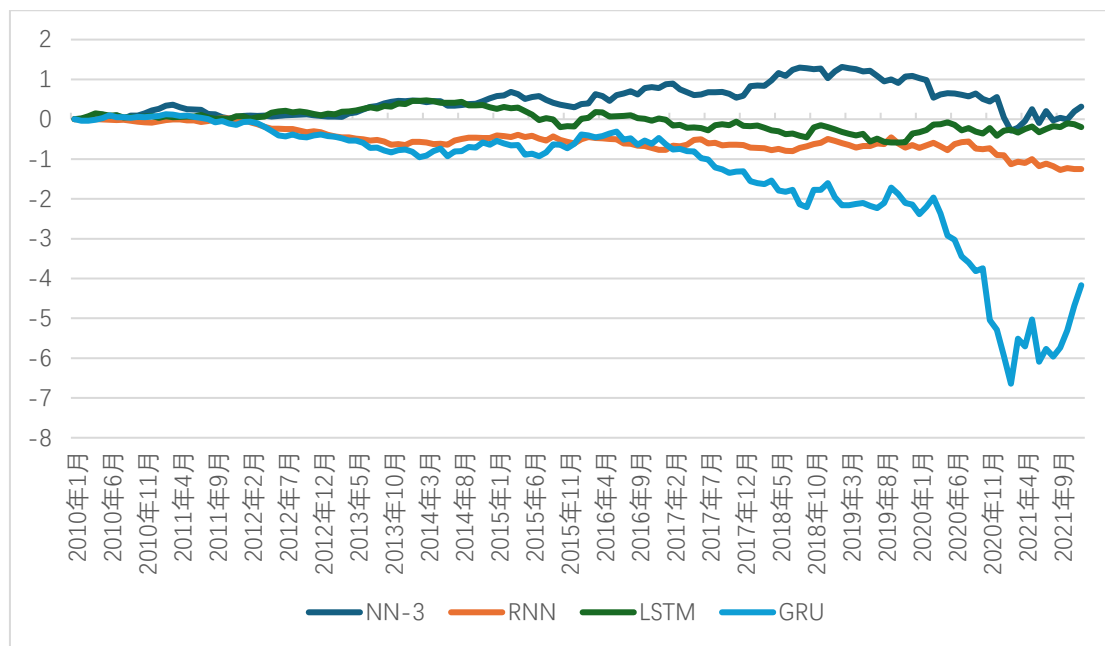


图 3.8 各模型在不同时间步时的累积收益之差

研究发现，当输入序列时间步从 1 期扩展至 3 期时，NN-3 模型展示出持续增强的时间序列预测能力，其累积收益在样本期内稳定为正，证明多期历史因子可提升模型的累积收益获取能力；RNN 与 GRU 模型的累积收益差全程为负，说明其在处理中长期序列依赖时存在梯度传到失真的风险。

LSTM 呈现阶段性特征：2015 年 8 月前累积收益差为正，之后转为负。发现该特征后，对不同时段模型预测准确率进行检验：时间步为 3 时，LSTM 集成模型 2015 年前预测准确率为 0.803%，2015 年后预测准确率降至 0.210%。这与其门控机制对市场状态切换过度敏感相关。2015 年 8 月，世界股市大幅下跌，道琼斯工业平均指数在 8 月 24 日开盘下跌 1000 点，并出现大幅波动。在 2015 年 8

月前的低波动环境下,LSTM 的门控机制能够有效捕捉预测信号,8 月后,其细胞状态更新规则难以适应急剧的机制转换,导致预测偏差累积。

3.5.4 模型交叉 alpha 验证

本节检验不同模型间的相互关系。在 Moskowitz、Ooi 和 Pedersen (2012) 的研究中,将时间序列动量策略(TS)的收益率回归到截面动量策略(CS)的收益率上,得到的截距项显著为正,而将 CS 回归到 TS 时,截距项在统计上不显著。他们将此解释为时间序列动量能够解释截面动量,但反之则不成立。本文使用每种模型的预测收益最高的第 10 组构成的投资组合,对四种模型相互间的交互作用,以及 Fama-French 五因子模型对四种模型的解释作用进行验证。

表 3.9 Cross-Alpha 检验

时间步为 1						
		X:NN-3	X:RNN	X:LSTM	X:GRU	X:FF-5
y:NN-3	α	--	(0.0026)	(0.0010)	(0.0022)	0.0131**
	t-value	--	(1.08)	(0.38)	(0.77)	2.02
y:RNN	α	0.0050**	--	0.0022	0.0017	0.0137**
	t-value	2.21	--	1.19	0.54	2.56
y:LSTM	α	0.0043	(0.0003)	--	0.0010	0.0147**
	t-value	1.48	(0.15)	--	0.26	2.38
y:GRU	α	0.0063**	0.0018	0.0039	--	0.0174***
	t-value	2.05	0.57	1.04	--	2.62
时间步为 3						
		X:NN-3	X:RNN	X:LSTM	X:GRU	X:FF-5
y:NN-3	α	--	0.0017	-0.0003	0.0012	0.0118*
	t-value	--	0.4974	-0.1188	0.4048	1.8269
y:RNN	α	0.0019	--	0.0006	0.0002	0.0131**
	t-value	0.6246	--	0.1958	0.0946	2.1471
y:LSTM	α	0.0019	0.0025	--	0.002	0.0143**
	t-value	0.8799	0.7514	--	0.6466	2.0858
y:GRU	α	0.0025	0.0012	0.0011	--	0.0133**
	t-value	0.97	0.63	0.40	--	2.20

从表中回归结果可以看出,在时间步为 1 时,NN-3 模型构造的投资组合产生的收益序列可以被 RNN、LSTM、GRU 模型产生的收益率序列完全解释,但反过来不能完全解释 RNN 和 GRU 产生的收益率序列(存在经 Newey-West 调整后统计显著的未解释 α 成分),说明此时 RNN、GRU 产生的收益序列中存在 NN 模型不能解释的信息。以 Fama-French 五因子作为自变量,各模型产生的收益率序列作为因变量时,截距项均为显著,说明四个模型的收益均不能被 Fama-French 五因子完全解释,四个模型产生的收益有其他来源。在时间步为 3 时,四个模型产生的收益序列两两之间可以完全解释,说明此时模型之间的信息差异不大。四个模型的收益均不能被 Fama-French 五因子模型完全解释。

上述结果表明,深度学习模型能够捕捉未被因子化的市场异象,因子模型中可以引入动态非线性补偿项来精准刻画复杂市场环境下的收益关系。

4 结论与展望

4.1 研究结论

本文以美股市场为研究对象，采用滑动窗口方法进行样本外预测，来评估深度学习模型在股票超额收益预测中的有效性。研究结论如下：

（1）模型集成能力与稳定性：通过模型集成，可以提升预测能力，在不同时间窗口下，模型的训练稳定性有所不同：窗口期为 1，即用当月因子预测下月收益时，神经网络模型（NN-3）的训练稳定性较高，而循环神经网络模型（RNN、LSTM、GRU）整体稳定性较差；窗口期为 3，即用前三个月因子预测下个月收益时，各模型的训练稳定性均有所提高，其中 NN-3 与 LSTM 预测能力较高，而 RNN 与 GRU 虽稳定性提升，但预测能力降低。

（2）循环神经网络模型与神经网络模型预测能力比较：NN 模型的预测能力整体较为稳定，时间步增加对其稳定性影响较小；在循环神经网络模型中，时间步为 3 的 LSTM 预测能力优于 NN 模型，说明 LSTM 能更好地捕捉较长时间跨度的非线性信息。

（3）抽样数据和完整数据预测能力比较：使用抽样数据训练的模型在预测超额收益时表现出更高的 R_{oos}^2 ，而使用完整数据训练的模型预测能力较低。这可能由于抽样数据在一定程度上能降低数据噪声的干扰，使模型更专注于学习因子与收益之间的核心非线性关系，提高泛化能力。相反，完整数据可能包含较多随机波动和市场噪声，导致模型在训练过程中受到干扰，从而降低预测准确率。合理的数据预处理与抽样策略可能是优化深度学习模型在金融市场应用的一种有效方法。

（4）同一时间步下不同复杂度的循环神经网络模型预测能力对比：时间步为 1 时（短期预测），随着模型复杂度增加，预测能力反而下降，特别是 LSTM 模型的预测结果呈负效应。这可能是因为短期内信息较少，复杂模型易受噪声影响，导致过拟合或学习不到有效模式。时间步为 3 时，随着模型复杂度增加，预测能力上升，说明在较长的时间序列中，复杂模型能够更好地学习到收益因子的动态关系。RNN 与 GRU 在该时间步下的预测效果仍然较差。

（5）不同时间步下循环神经网络模型预测能力的变化趋势：随着时间步增

加,模型预测能力与模型复杂度之间的关系发生逆转。短时间步下(时间步为1),简单模型更优,而长时间步下(时间步为3),复杂模型更优。在超额收益预测中,循环神经网络模型的有效性可能依赖于适当的时间窗口选择。

(6) 投资组合构建与收益获取:研究结果表明,基于深度学习模型预测信号可以构建投资组合,并能获取显著超越基准指数的收益。这说明尽管模型的预测能力存在局限性,但仍可用于选股策略,为量化投资提供参考价值。

(7) 预测能力与超额收益的关系:研究结果表明,尽管某些模型在样本外预测中表现出较高的准确率,但其构建的分组投资组合未必能实现较高的平均收益。同样,即便某一模型的分组平均收益较高,也不意味着其在长期投资策略中能够获得更优的累积超额收益。这一现象可能受多种因素影响,如模型未调参导致模型不收敛、回测时市场环境变化、交易成本、策略执行的稳定性等,表明单纯依赖预测能力来优化投资策略可能存在局限性,需要结合更全面的回测分析与风险控制机制进行优化。

(8) 通过模型交叉 alpha 验证可知,深度学习模型获取的收益中存在 Fama-French 五因子不能解释的部分、循环神经网络深度学习模型中存在 NN 模型不能解释的信息。说明深度学习模型能够捕捉未被因子化的市场异象,因子模型可以引入动态非线性补偿项来精准刻画复杂市场环境下的收益关系;循环神经网络深度学习模型可以捕捉传统神经网络无法获得的信息。

4.2 不足与展望

本研究在以下方面存在改进空间,这些局限也为后续研究指明了方向:

第一,因子筛选机制缺失。本研究未引入动态因子筛选机制,所有模型均采用统一的基础因子集。这种设计虽保证了跨模型可比性,但可能削弱对市场时变特征的适应性。未来研究可探索分层因子筛选框架,在基准因子的基础上引入 CNN、注意力机制等结构进行因子筛选。

第二,非结构化信息融合不足。当前模型仅处理结构化财务数据,尚未整合新闻舆情、供应链关系图谱等非结构性因子。后续可引入多模态深度学习架构,例如采用 BERT 模型提取财经文本语义特征、结合图神经网络(GNN)建模产业链拓扑结构等,从而提升模型对复杂市场环境的解析能力。

第三,模型架构设计简易。由于算力限制和研究目标(分析网络结构差异的

影响), 研究采用标准参数配置, 未进行超参数优化, 也未尝试混合架构(如 CNN-LSTM 融合模型)。这可能导致模型潜力未完全释放。随着分布式计算技术的普及, 未来可架构两阶段优化框架: 首先通过贝叶斯优化确定各模型独立的最优参数空间, 继而设计门控机制自适应融合不同网络的预测效果。

第四, 回测环节理想化假设。回测环节未考虑滑点、限价单成交概率、大额订单市场冲击等现实约束, 可能存在理论收益与实操表现的偏差。

综上所述, 虽然本研究在循环神经网络深度学习模型预测股票超额收益方面取得了一定成果, 但依然有许多方面可以进一步优化。通过不断改进模型和策略, 未来能够为投资者提供更加精准、稳健的投资工具, 并在更加复杂多变的 market 环境中取得优异的表现。

参考文献

- [1] Sharpe W F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk[J]. The Journal of Finance, 1964, 19(3): 425.
- [2] Merton R C. An Intertemporal Capital Asset Pricing Model[J]. Econometrica, 1973, 41(5): 867.
- [3] Ross S A. The arbitrage theory of capital asset pricing[J]. Journal of Economic Theory, 1976, 13(3): 341-360.
- [4] Fama E F, French K R. Common risk factors in the returns on stocks and bonds[J]. Journal of financial economics, 1993, 33(1): 3-56.
- [5] Fama E F, French K R. A five-factor asset pricing model[J]. Journal of Financial Economics, 2015, 116(1): 1-22.
- [6] Cochrane J H. Presidential address: Discount rates[J]. The Journal of finance, 2011, 66(4): 1047-1108.
- [7] Harvey C R, Liu Y, Zhu H. ... and the Cross-Section of Expected Returns[J]. Review of Financial Studies, 2016, 29(1): 5-68.
- [8] Kelly B T, Pruitt S, Su Y. Characteristics are covariances: A unified model of risk and return[J]. Journal of Financial Economics, 2019, 134(3): 501-524.
- [9] Gu S, Kelly B, Xiu D. Empirical Asset Pricing via Machine Learning[J]. The Review of Financial Studies, 2020, 33(5): 2223-2273.
- [10] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals, and Systems, 1989, 2(4): 303-314.
- [11] Chiah M, Chai D, Zhong A, et al. A Better Model? An empirical investigation of the Fama–French five-factor model in Australia[J]. International Review of Finance, 2016, 16(4): 595-638.
- [12] Lalwani V, Chakraborty M. Multi-factor asset pricing models in emerging and developed markets[J]. Managerial Finance, 2020, 46(3): 360-380.
- [13] Liao Z, Lin Z, Quan X. Analysis of the Evolution of the Multiple Factors Financing Model[J]. Highlights in Business, Economics and Management, 2023, 19: 450-458.
- [14] Roy R, Shijin S. A six-factor asset pricing model[J]. Borsa Istanbul Review, 2018, 18(3): 205-217.
- [15] Machado M A V, De Medeiros O R. Asset pricing model and the liquidity effect: Empirical evidence in the Brazilian stock market[J]. Brazilian Review of Finance, 2011, 9(3): 383-412.
- [16] Racicot F E, Rentz W F. Testing Fama–French’s new five-factor asset pricing model: evidence from robust instruments[J]. Applied Economics Letters, 2016, 23(6): 444-448.
- [17] Putra K, Chalid D A. Accuracy Level Analysis of Pricing Model on State-Owned Enterprises Stocks[C]//5th Global Conference on Business, Management and Entrepreneurship (GCBME

- 2020). Atlantis Press, 2021: 73-76.
- [18] Wedagama D N W, Hakim D B, Juanda B, et al. Implementation of Capm, Fama-French Three-Factor, and Five-Factor in Indonesia Stock Exchange and Cement Industry Sector[J]. International Journal of Economics and Financial Issues, 2022, 12(2): 52-60.
- [19] Galvao A F, Montes-Rojas G, Olmo J. Tests of asset pricing with time-varying factor loads[J]. Journal of Applied Econometrics, 2019, 34(5): 762-778.
- [20] Bollerslev T. Generalized autoregressive conditional heteroskedasticity[J]. Journal of econometrics, 1986, 31(3): 307-327.
- [21] Kelly B T, Moskowitz T J, Pruitt S. Understanding momentum and reversal[J]. Journal of financial economics, 2021, 140(3): 726-743.
- [22] 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.
- [23] 许杰, 祝玉坤, 邢春晓. 机器学习在金融资产定价中的应用研究综述[J]. 计算机科学, 2022, 49(6): 276-286.
- [24] Yuan M. Research on Multi Factor Quantitative Investment Strategies Based on Machine Learning[J].
- [25] Sahu S K, Mokhade A, Bokde N D. An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges[J]. Applied Sciences, 2023, 13(3): 1956.
- [26] Yu J. An Effective Composite Investment Quantitative Strategy with Machine Learning Method[J]. Highlights in Business, Economics and Management, 2024, 31: 65-70.
- [27] Lee N, Moon J. Offline Reinforcement Learning for Automated Stock Trading[J]. IEEE Access, 2023, 11: 112577-112589.
- [28] Gu S, Kelly B, Xiu D. Autoencoder asset pricing models[J]. Journal of Econometrics, 2021, 222(1): 429-450.
- [29] Zhang R, Huang C, Zhang W, et al. Multi Factor Stock Selection Model Based on LSTM[J]. 2018, 10(8).
- [30] 周亮. 基于随机森林模型的股票多因子投资研究[J]. 金融理论与实践, 2021(7): 97-103.
- [31] 李仁宇, 叶子谦. 基于机器学习的基金收益预测[J]. 统计与决策, 2023, 39(11): 156-161.
- [32] 张晗, 王维国. 用于金融时间序列预测的神经网络集成模型[J]. 北京邮电大学学报: 1-6.
- [33] Wu M E, Syu J H, Lin J C W, et al. Portfolio management system in equity market neutral using reinforcement learning[J]. Applied Intelligence, 2021, 51(11): 8119-8131.
- [34] Liang Z, Chen H, Zhu J, et al. Adversarial deep reinforcement learning in portfolio management[J]. arXiv preprint arXiv:1808.09940, 2018.
- [35] Jin B. A mean-VaR based deep reinforcement learning framework for practical algorithmic trading[J]. IEEE Access, 2023, 11: 28920-28933.
- [36] Zhang E. Portfolio Optimization Strategy Based on Four Deep Learning Models[J]. Advances in Economics, Management and Political Sciences, 2023, 47: 295-302.

- [37] Elman J L. Finding Structure in Time[J]. Cognitive Science, 1990, 14(2): 179-211.
- [38] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [39] Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv preprint arXiv:1409.1259, 2014.
- [40] Gunarto D M, Sa'adah S, Utama D Q. Predicting cryptocurrency price using rnn and lstm method[J]. Jurnal Sisfokom (Sistem Informasi dan Komputer), 2023, 12(1): 1-8.
- [41] 任晓萍, 陈志平. 采用门控循环单元与深度进化策略的股票指数量化模型[J]. 西安交通大学学报, 2025, 59(2): 146-155.
- [42] Xiang S. Evaluations of the machine learning schemes for cryptocurrency prediction[J]. Applied and Computational Engineering, 2024, 54(1): 82-89.
- [43] 李星毅, 李仲飞, 李其谦, 等. 基于机器学习的资产收益率预测研究综述[J]. 中国管理科学, 2025, 33(1): 311-322.
- [44] 孟祥煜. 我国国债期货价格波动率预测研究——基于 HAR-GARCH-GRU 混合模型的分析[J]. 价格理论与实践, 2022(7): 84-88.
- [45] Furizal F, Ritonga A, Ma'arif A, et al. Stock Price Forecasting with Multivariate Time Series Long Short-Term Memory: A Deep Learning Approach[J]. Journal of Robotics and Control (JRC), 2024, 5(5): 1322-1335.
- [46] Ke Z T, Kelly B T, Xiu D. Predicting returns with text data[R]. National Bureau of Economic Research, 2019.
- [47] Jiang J, Kelly B, Xiu D. (Re-) Imag (in) ing price trends[J]. The Journal of Finance, 2023, 78(6): 3193-3249.
- [48] Obaid K, Pukthuanthong K. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news[J]. Journal of Financial Economics, 2022, 144(1): 273-297.
- [49] 姚加权, 冯绪, 王赞钧, 等. 语调、情绪及市场影响:基于金融情绪词典[J]. 管理科学学报, 2021, 24(5): 26-46.
- [50] 张少军, 苏长利. 基于情绪词典和 BERT-BiLSTM 的股指预测研究[J]. 计算机工程与应用, 2025, 61(4): 358-367.
- [51] 于孝建, 刘国鹏, 刘建林, 等. 基于 LSTM 网络和文本情感分析的股票指数预测[J]. 中国管理科学, 2024, 32(8): 25-35.
- [52] 交通银行金融服务中心支付清算课题组, 郑英姿. 深度学习模型在跨境支付信息体系中的应用探究——基于自由格式报文分类场景[J]. 新金融, 2025(3): 48-57.
- [53] Zhao A B, Cheng T. Stock return prediction: Stacking a variety of models[J]. Journal of Empirical Finance, 2022, 67: 288-317.
- [54] Chollet F. Deep learning with Python[M]. Simon and Schuster, 2021.
- [55] Zhu B, Baesens B, Backiel A, et al. Benchmarking sampling techniques for imbalance learning in churn prediction[J]. Journal of the Operational Research Society, 2018, 69(1): 49-65.

- [56] 江雨燕, 邵金, 陈梦凯, 等. 融合分解集成和深度学习的金融时间序列预测模型[J]. 统计与决策, 2023, 39(24): 152-156.
- [57] Welch I, Goyal A. A Comprehensive Look at The Empirical Performance of Equity Premium Prediction[J]. Review of Financial Studies, 2008, 21(4): 1455-1508.
- [58] Campbell J Y, Shiller R J. Stock prices, earnings, and expected dividends[J]. the Journal of Finance, 1988, 43(3): 661-676.
- [59] Carhart M M. On Persistence in Mutual Fund Performance[J]. The Journal of Finance, 1997, 52(1): 57-82.
- [60] ASNESS C S, MOSKOWITZ T J, PEDERSEN L H. Value and Momentum Everywhere[J]. The Journal of Finance, 2013, 68(3): 929-985.
- [61] Cooper M J, Gulen H, Schill M J. Asset growth and the cross-section of stock returns[J]. the Journal of Finance, 2008, 63(4): 1609-1651.
- [62] Chan L K, Lakonishok J, Sougiannis T. The stock market valuation of research and development expenditures[J]. The Journal of finance, 2001, 56(6): 2431-2456.
- [63] Jegadeesh N, Titman S. Returns to buying winners and selling losers: Implications for stock market efficiency[J]. The Journal of finance, 1993, 48(1): 65-91.
- [64] Baker M, Wurgler J. A catering theory of dividends[J]. The Journal of finance, 2004, 59(3): 1125-1165.
- [65] Moskowitz T J, Ooi Y H, Pedersen L H. Time series momentum[J]. Journal of Financial Economics, 2012, 104(2): 228-250.

附录

简称	名称	含义	文献	来源	更新频率
abasec	绝对应计项目	衡量公司应计项目的绝对值。	Bandyopadhyay, Huang & Wujanto (2010, WP)	Compustat	年度数据
acc	营运资本应计项目	衡量公司营运资本的应计项目。	Sloan (1996, TAR)	Compustat	年度数据
aoavol	异常收益公告量	衡量公司收益公告时的异常交易量。	Lerman, Livnat & Mendenhall (2007, WP)	Compustat+ CRSP	季度数据
age	自首次 Compustat 覆盖以来的年数	衡量公司自首次被 Compustat 覆盖以来的年数。	Jiang, Lee & Zhang (2005, RAS)	Compustat	年度数据
agr	资产增长	衡量公司资产的总增长。	Cooper, Gulen & Schill (2008, JF)	Compustat	年度数据
baspread	买卖价差	衡量股票的买卖价差。	Amihud & Mendelson (1989, JF)	CRSP	月度数据
beta	贝塔	衡量股票相对于市场的系统性风险。	Fama & MacBeth (1973, JFE)	CRSP	月度数据
betasq	贝塔平方	衡量贝塔的平方值。	Jones, MacBeth (1973, JFE)	CRSP	月度数据
bm	账面市值比	衡量公司账面价值与市场价值的比	Rosenberg, Reid &	Compustat+	年度数据

附 录

		率。	Lanstein (1985, JPM)	CRSP	
bm_ia	行业调整后的账面 市值比	衡量行业调整后的账面市值比。	Asness, Porter & Stevens (2000, WP)	Compustat+ CRSP	年度数据
cash	现金持有	衡量公司持有的现金量。	Palazzo (2012, JFE)	Compustat	季度数据
cashdebt	现金流与债务比率	衡量公司现金流与债务的比率。	Ou & Penman (1989, AJE)	Compustat	年度数据
cashpr	现金生产率	衡量公司现金的生产效率。	Chandrashekar & Rao (2009, WP)	Compustat	年度数据
cfp_ia	现金流与价格比率	衡量公司现金流与股票价格的比率。	Desai, Rajgopal & Venkatachalam (2004, TAR)	Compustat	年度数据
cfb_ia	行业调整后的现金 流与价格比率	衡量行业调整后的现金流与价格比 率。	Asness, Porter & Stevens (2000, WP)	Compustat	年度数据
chasda	行业调整后的资产 周转率变化	衡量行业调整后的资产周转率变化。	Soliman (2008, TAR)	Compustat	年度数据
chenpoa	流通股变化	衡量公司流通股的变化。	Pontiff & Woodgate (2008, JF)	Compustat	年度数据
chenpha	行业调整后的员工 变化	衡量行业调整后的员工数量变化。	Asness, Porter & Stevens (1994, WP)	Compustat	年度数据
chinva	存货变化	衡量公司存货的变化。	Thomas & Zhang (2002,	Compustat	年度数据

附 录

			RAS)		
chromn	第四动量变化	衡量股票动量的变化。	Gottleman & Marks (2008, WP)	CRSP	月度数据
chpmia	行业调整后的利润率变化	衡量行业调整后的利润率变化。	Soliman (2008, TAR)	Compustat	年度数据
chtx	税收费用变化	衡量公司税收费用的变化。	Thomas & Zhang (2011, JAR)	Compustat	季度数据
chx	公司投资	衡量公司的投资水平。	Titman, Wei & Xie (2004, JFQA)	Compustat	季度数据
convind	可转换债务指标	衡量公司是否有可转换债务。	Valta (2010, JFQA)	Compustat	年度数据
currat	流动比率	衡量公司流动资产与流动负债的比率。	Ou & Penman (1989, AJE)	Compustat	年度数据
depr	折旧与固定资产比率	衡量公司折旧与固定资产的比率。	Holthausen & Larcker (1992, JAE)	Compustat	年度数据
divi	股息初始	衡量公司是否首次发放股息。	Michaely, Thaler & Womack (1995, JF)	Compustat	年度数据
divo	股息遗漏	衡量公司是否遗漏了股息发放。	Michaely, Thaler & Womack (1995, JF)	Compustat	年度数据
doivol	美元交易量	衡量股票的交易量（以美元计）。	Chordia, Subrahmanyam &	CRSP	月度数据

附 录

			Anshuman (2001, JFE)		
dy	股息与价格比率	衡量公司股息与股票价格的比率。	Litzenberger & Ramaswamy (1982, JF)	Compustat	年度数据
ear	收益公告回报	衡量公司收益公告时的股票回报。	Kishore, Brandt, Santa- Clara & Venkatachalam (2008, WP)	Compustat+ CRSP	季度数据
egr	普通股股东权益增 长	衡量普通股股东权益的增长。	Richardson, Sloan, Soliman & Tuna (2005, JAE)	Compustat	年度数据
ep	收益政策	衡量公司收益政策。	Basu (1977, JP)	Compustat	年度数据
gma	毛利率	衡量公司的毛利率。	Novy-Marx (2013, JFE)	Compustat	年度数据
grCAPX	资本支出增长	衡量公司资本支出的增长。	Anderson & Garcia-Feijoo (2006, JP)	Compustat	年度数据
gfrInoa	长期净经营资产增 长	衡量公司长期净经营资产的增长。	Fairfield, Whisenant & Yohn (2003, TAR)	Compustat	年度数据
href	行业销售集中度	衡量行业销售集中度。	Hoberg & Phillips (2006, JP)	Compustat	年度数据
hire	员工增长率	衡量公司员工的增长率。	Bazdresch, Belo & Lin (2014, JFE)	Compustat	年度数据

附 录

idlovol	异质回报波动率	衡量股票的异质回报波动率。	Ali, Hwang & Trombley (2003, JFE)	CRSP	月度数据
ill	非流动性	衡量股票的非流动性。	Amihud (2002, JFM)	CRSP	月度数据
indmom	行业动量	衡量行业的动量效应。	Moskowitz & Grinblatt (1999, JP)	CRSP	月度数据
invest	资本支出和存货	衡量公司的资本支出和存货水平。	Chen & Zhang (2010, JP)	Compustat	月度数据
lev	杠杆	衡量公司的财务杠杆。	Bhandari (1988, JP)	Compustat	年度数据
lgr	长期债务增长	衡量公司长期债务的增长。	Richardson, Sloan, Soliman & Tuna (2005, JAE)	Compustat	年度数据
maxret	最大日回报	衡量股票的最大日回报。	Bali, Cakici & Whitelaw (2011, JFE)	CRSP	月度数据
mom12m	12 个月动量	衡量股票过去 12 个月的动量。	Jegadeesh (1990, JP)	CRSP	月度数据
mom1m	1 个月动量	衡量股票过去 1 个月的动量。	Jegadeesh & Titman (1993, JP)	CRSP	月度数据
mom36m	36 个月动量	衡量股票过去 36 个月的动量。	Jegadeesh & Titman (1993, JP)	CRSP	月度数据
mom6m	6 个月动量	衡量股票过去 6 个月的动量。	Jegadeesh & Titman (1993, JP)	CRSP	月度数据

附 录

ms	财务报表评分	衡量公司财务报表的评分。	Mohanram (2005, RAS)	Compustat	季度数据
mvel	行业调整后的规模	衡量行业调整后的公司规模。	Banz (1981, JFE)	CRSP	月度数据
mve_ia	行业调整后的规模	衡量行业调整后的公司规模。	Asness, Porter & Stevens (2000, WP)	Compustat	年度数据
nincr	收益增加次数	衡量公司收益增加的次数。	Barth, Elliott & Finn (1999, JAR)	Compustat	季度数据
opeprof	营业利润率	衡量公司的营业利润率。	Fama & French (2015, JFE)	Compustat	年度数据
orgap	组织资本	衡量公司的组织资本。	Falato & Papanikolaou (2013, JP)	Compustat	年度数据
pchenyx_ia	行业调整后的资本支出变化百分比	衡量行业调整后的资本支出变化百分比。	Abarbanell & Bushee (1995, TAR)	Compustat	年度数据
pchenyrat	流动比率变化百分比	衡量公司流动比率的变化百分比。	Ou & Penman (1999, JAE)	Compustat	年度数据
pchdepr	折旧变化百分比	衡量公司折旧的变化百分比。	Holthausen & Larcker (1992, JAE)	Compustat	年度数据
pchgn_pchsale	毛利率变化百分比与销售额变化百分比的差异	衡量毛利率变化百分比与销售额变化百分比的差异。	Abarbanell & Bushee (1998, TAR)	Compustat	年度数据

附 录

pchqufda	速动比率变化百分比	衡量公司速动比率的变化百分比。	Ou & Penman (1999, JAE)	Compustat	年度数据
pchsale_pchinv t	销售额变化百分比 与存货变化百分比 的差异	衡量销售额变化百分比与存货变化百分比的差异。	Abarbanell & Bushee (1998, TAR)	Compustat	年度数据
pchsale_pchrec t	销售额变化百分比 与应收账款变化百分比 的差异	衡量销售额变化百分比与应收账款变化百分比的差异。	Abarbanell & Bushee (1998, TAR)	Compustat	年度数据
pchsale_pchxega	销售额变化百分比 与销售管理费用变化 百分比的差异	衡量销售额变化百分比与销售管理费用变化百分比的差异。	Abarbanell & Bushee (1998, TAR)	Compustat	年度数据
pchsaleinv	销售额与存货比率 变化百分比	衡量销售额与存货比率的变化百分比。	Ou & Penman (1989, JAE)	Compustat	年度数据
petace	应计项目百分比	衡量公司应计项目的百分比。	Hribar, Lundholm & Van Winkle (2011, TAR)	Compustat	年度数据
pricedelay	价格延迟	衡量股票价格对信息的反应延迟。	Hou & Moskowitz (2005, RFS)	CRSP	月度数据
ps	财务报表评分	衡量公司财务报表的评分。	Piotroski (2000, JAR)	Compustat	年度数据
quick	速动比率	衡量公司速动比率。	Ou & Penman (1989, JAE)	Compustat	年度数据

附 录

rd	研发增加	衡量公司研发支出的增加。	Eberhart, Maxwell & Siddique (2004, JF)	Compustat	年度数据
rd_mve	研发与市值比率	衡量公司研发支出与市值的比率。	Guo, Lev & Shi (2006, JBFA)	Compustat	年度数据
rd_sale	研发与销售额比率	衡量公司研发支出与销售额的比率。	Guo, Lev & Shi (2006, JBFA)	Compustat	年度数据
realestate	房地产持有	衡量公司持有的房地产资产。	Tuzel (2010, RFS)	Compustat	年度数据
retvol	回报波动率	衡量股票回报的波动率。	Ang, Hodrick, Xing & Zhang (2006, JF)	CRSP	月度数据
roaq	资产回报率	衡量公司资产的回报率。	Balakrishnan, Bartov & Faurel (2010, JAE)	Compustat	季度数据
roavol	收益波动率	衡量公司收益的波动率。	Francis, LaFond, Olsson & Schipper (2004, TAR)	Compustat	季度数据
roeq	股东权益回报率	衡量公司股东权益的回报率。	Hou, Xue & Zhang (2015, RFS)	Compustat	季度数据
roic	投资资本回报率	衡量公司投资资本的回报率。	Brown & Rowe (2007, WP)	Compustat	年度数据
rsup	收入惊喜	衡量公司收入与预期的差异。	Kama (2009, JBFA)	Compustat	季度数据
salecash	销售额与现金比率	衡量公司销售额与现金的比率。	Ou & Penman (1989, JAE)	Compustat	年度数据

附 录

saleinv	销售额与存货比率	衡量公司销售额与存货的比率。	Ou & Penman (1989, JAE)	Compustat	年度数据
salerec	销售额与应收账款比率	衡量公司销售额与应收账款的比率。	Ou & Penman (1989, JAE)	Compustat	年度数据
secured	担保债务	衡量公司的担保债务水平。	Valta (2016, JFQA)	Compustat	年度数据
securedind	担保债务指标	衡量公司是否有担保债务。	Valta (2016, JFQA)	Compustat	年度数据
sgr	销售增长	衡量公司销售额的增长。	Lakonishok, Shleifer & Vishny (1994, JF)	Compustat	年度数据
sin	罪恶股票	衡量公司是否属于“罪恶”行业（如烟草、酒精等）。	Ling & Kacperczyk (2009, JFE)	Compustat	年度数据
sp	销售额与价格比率	衡量公司销售额与股票价格的比率。	Barbee, Mukherji & Raines (1996, FAJ)	Compustat	年度数据
std_dolvol	流动性波动率	衡量股票流动性（以美元交易量计）的波动率。	Chordia, Subrahmanyam & Anshuman (2001, JFE)	CRSP	月度数据
std_turn	流动性波动率	衡量股票流动性（以股票换手率计）的波动率。	Chordia, Subrahmanyam & Anshuman (2001, JFE)	CRSP	月度数据
stdacc	应计项目波动率	衡量公司应计项目的波动率。	Bandyopadhyay, Huang & Wirjanto (2010, WP)	Compustat	季度数据
stdcf	现金流波动率	衡量公司现金流的波动率。	Huang (2009, JEF)	Compustat	季度数据
tang	债务能力/公司有形	衡量公司债务能力和有形资产的比	Almeida & Campello	Compustat	年度数据

附 录

	性	例。	(2007, RFS)		
tb	税收收入与账面收入比率	衡量公司税收收入与账面收入的比率。	Lev & Nissim (2004, TAR)	Compustat	年度数据
turn	股票换手率	衡量股票的换手率。	Datar, Naik & Radcliffe (1998, JFM)	CRSP	月度数据
zerotrade	零交易日	衡量股票在某个时间段内没有交易的天数。	Liu (2006, JFE)	CRSP	月度数据