

分类号 _____
U D C _____

密级 _____
编号 10741



硕士学位论文

论文题目 基于图卷积神经网络最优投资组合研究

研究生姓名: 唐裕博

指导教师姓名、职称: 韩海波 副教授

学科、专业名称: 应用经济学 数量经济学


研究方向: 金融计量经济分析

提交日期: 2022年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 唐懿博 签字日期： 2022.5.25


导师签名：  签字日期： 2022.5.25

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 唐懿博 签字日期： 2022.5.25

导师签名：  签字日期： 2022.5.25

Research on optimal portfolio based on graph neural network

Candidate : Tang Yubo

Supervisor: Han haibo

摘 要

伴着国民经济水平的提高,股票投资逐渐走进大众视野。众所周知股票投资具有很大的风险,尚待完善的监管体制、投资者从众心理以及恶意坐庄行为都会加大市场的风险。所以对于投资者来讲,最为关心的是如何判定投资风险、在规避风险时如何进行投资组合才获取最大的收益。哈里马科维茨在 20 世纪 50 年代首次提出投资组合理念,并提出均值方差模型。该理论模型为之后投资组合问题的研究奠定了基础。为了获得更加符合我国市场的投资组合模型,并且结合科学的方法更快更准确地为投资者提供合理的投资组合,本文在经典均值方差模型的基础上,结合我国投资市场现状以及我国投资者的需求主要做了以下几个方面的研究:

(1) 基于经典的均值方差模型,提出实用的最优投资组合概念。通过对未来的收益方差进行预测,然后引用均值方差理论计算未来投资的风险,求得风险最小的组合为最优投资组合。

(2) 基于最优投资组合的概念,采用统计学方法 GARCH 模型,以及机器学习方法 LSTM 模型构建投资组合。首先对选中股票池中的数据进行收益率方差的预测,在此基础上计算具有最小收益率方差的组合构建最优投资组合。

(3) 基于最优投资组合的概念,采用图卷积神经网络模型,对股票池中的股票进行收益率方差预测并构造投资组合。首先通过对上证 50 的股票池中,依据地域板块选中主要三个地域板块的 22 支股票,进行图构建之后再放入模型,预测熊市牛市以及震荡区间的收益率方差,之后经过计算求得最优权重,构建具有最小收益率方差的最优投资组合。

研究表明:通过 GCN 模型构建的最优投资组合、GARCH 模型建模的最优投资组合和 LSTM 模型构建的最优投资组合都可以很好的抵抗市场的风险,在牛市能保证可观的收益。在熊市的时候可以取得不错的收益,在股市震荡区间的时候,

三者都可以取得正向的收益。三个模型对比，GCN 构建的最优投资组合效果更好一些，无论是在熊市、牛市以及震荡区间。GCN 模型构建的最优投资组合同 GARCH 模型、LSTM 模型构建的投资组合相比，都具有更好的抗风险能力，更好的经济效益。

关键词：均值方差理论；最优投资组合；GARCH 模型；LSTM 模型；GCN 模型；

Abstract

With the improvement of the national economic level, stock investment has gradually entered the public eye. As we all know, stock investment has great risks, and the yet to be perfected regulatory system, investors' herd mentality and malicious behavior will increase the risk of the market. Therefore, for investors, what they are most concerned about is how to determine investment risks and how to make investment portfolios to obtain maximum returns when avoiding risks. Harry Markowitz first proposed the portfolio idea in the 1950s and proposed the mean-variance model. This theoretical model lays the foundation for the subsequent research on portfolio problems. In order to obtain an investment portfolio model that is more in line with the Chinese market, and combine scientific methods to provide investors with a reasonable investment portfolio faster and more accurately, this paper, on the basis of the classic mean variance model, combines the current situation of my country's investment market and the investment situation of Chinese investors. The needs are mainly studied in the following aspects:

(1) Based on the classical mean-variance model, a practical concept of optimal investment portfolio is proposed. By predicting the variance of future returns, and then using the mean variance theory to calculate the risk of future investment, the portfolio with the smallest risk is obtained as the optimal portfolio.

(2) Based on the concept of the optimal investment portfolio, the statistical method GARCH model and the machine learning method LSTM model are used to construct the investment portfolio. Firstly, the data in the selected stock pool is forecasted for the return variance, and on this basis, the optimal portfolio is constructed by calculating the portfolio with the smallest return variance.

(3) Based on the concept of optimal investment portfolio, a graph convolutional neural network model is used to predict the return variance of stocks in the stock pool and construct an investment portfolio. First of all, by selecting 22 stocks in the main three regional sectors from the stock pool of the Shanghai Stock Exchange 50 according to the regional sectors, after constructing the graph, they are put into the model to predict the bear market, bull market and the yield variance of the shock range. The optimal weights are used to construct the optimal portfolio with the smallest return variance.

The research shows that the optimal investment portfolio constructed by the GCN model, the optimal investment portfolio constructed by the GARCH model and the optimal investment portfolio constructed by the LSTM model can well resist the risks of the market, and can guarantee considerable returns in the bull market. In a bear market, you can achieve good returns, and when the stock market fluctuates, all three can achieve positive returns. Comparing the three models, the optimal portfolio

constructed by GCN is better, no matter in bear market, bull market and shock range. Compared with the investment portfolio constructed by GARCH model and LSTM model, the optimal investment portfolio constructed by GCN model has better anti-risk ability and better economic benefits.

Key words : Mean-variance theory; optimal portfolio; GARCH model; LSTM model; GCN model;

目 录

1 绪论	1
1.1 研究背景	1
1.2 国内外研究现状	2
1.2.1 股市预测研究现状	2
1.2.2 股市选股研究现状	3
1.2.3 图卷积神经网络研究现状	4
1.3 研究意义	5
1.4 研究内容及安排	5
1.4.1 研究内容	5
1.4.2 结构安排	6
1.5 创新之处	7
2 相关概念与方法基础	8
2.1 深度学习方法	8
2.1.1 卷积神经网络	8
2.1.2 循环神经网络	9
2.1.3 长短期记忆人工神经网络	10
2.1.4 GRU 神经网络	11
2.1.5 图神经网络	12
2.1.6 图卷积神经网络	12
2.2 投资组合理论	13
2.2.1 均值方差理论	13
2.2.2 最优投资组合理论	14
3 最优投资组合实证研究	17
3.1 数据整理	17
3.1.1 股价走势分析	18
3.1.2 股票收益率分析	20
3.2 股票风险估计	22
3.2.1 GARCH 模型股票风险估计	22
3.2.2 LSTM 模型股票风险估计	25
3.2.3 GCN 模型股票风险估计	29
3.3 最优投资组合的构建	34
3.3.1 GARCH 构建最优投资组合	35
3.3.2 LSTM 构建最优投资组合	38
3.3.3 GCN 构建最优投资组合	40
3.4 最优投资组合收益的计算	43
4 三种模型构建组合对比分析	45
4.1 组合结构对比分析	45

4.2 组合收益对比分析	48
4.2.1 牛市收益率对比分析	48
4.2.2 熊市收益率对比分析	49
4.2.3 震荡区间收益率对比分析	50
5 总结与展望	51
5.1 总结	51
5.1.1 收益方面	51
5.1.2 抗风险能力方面	51
5.1.3 股市波动方面	52
5.2 展望	52
参考文献	54
致 谢	58
附录	59

1 绪论

1.1 研究背景

伴随着改革开放，我国经济水平的不断提高，证券市场也在飞速发展。在 1984 年我国开始以分割的区域性试点为主，开始个别股票的发行。在个别股票实行发行后，1987 年深证正确公司成立，1990 年上海证券所开始营业，在 1991 年中国证券业协会成立，我国的证券市场开始走向规范步入大众视野。现在已经是改革开放 52 年，我国的证券已经逐步走向成熟。在 2020 年末，上市、深市两市股票的数量已经超过 4100 支，我国的股市的总市值也突破 776 万亿元，两市的日交易额也是非常的巨大，每日都会有千亿成交额。证券市场中的股票的数量也是呈直线上涨。因为证券市场的波动性的存在，股市就存在投资风险。如果可以对证券市场进行深度的分析，不但可以很好的帮助证券市场监管工作，同时还可以给股票投资者有参考性的建议。众所周知，证券市场的波动性时时刻刻都存在，非常的难以掌握。如果此时可以有一个很好的方法，可以对证券市场的波动趋势做一个精准的预测，这样就利于降低投资的风险，还有利于相关部门做好金融风险防控工作。股市的风险对于市场的监管以及个人的投资者的决定的影响相当巨大的，因此相关的问题研究也是学者们、商业人士们研究热点之一。

在证券市场，学者以及相关的专业人士对证券市场的预测也是有了众多的研究，相关的预测方法也是越来越多，针对不同类型的数据，不同频率的数据都有相关的预测方法。伴随着计算机技术的突飞猛进，投资者们可以实时获得众多股票相关信息，计算机也可以帮助投资者获取以及处理实时的海量数据。众所周知，数据量的增加使其包含有价值的信息更多，可以更为精准全面的表达数据内包含的特征以及更多的微观结构，可以更好的分析以及预测证券波动的变化。在未来的金融市场，利用计算机对证券的价格波动预测必定成为未来的主流方向，具有重要的现实意义。

从现实层面来讲，我国证券市场仍处于发展阶段，市场整体的风险程度较高，我国证券市场上的投资者对于市场风险认知水平很低，对一些市场中释放的危险信号不敏感。他们大多金融知识准备不足，对股市中认知以及在股市中的操作都

是凭借“经验”、“感觉”。他们非常容易受到市场波动的影响造成经济损失。因此对金融市场的风险进行量化,让投资者可以更清楚自己所面临的风险是非常重要的。通过对金融市场风险的量化,更好的让投资者进行股票的投资组合,更有利于投资者规避风险获取收益。

综上所述,本文将对金融风险进行深入研究,通过 GCN、GARCH、LSTM 模型对风险的预测,在此基础上构建风险最小最优投资组合并对比不同组合的收益情况,为投资者提供一定参考。

1.2 国内外研究现状

1.2.1 股市预测研究现状

股市预测领域,众多学者使用了诸多的模型对股价进行建模预测,期望预测的结果可以更为精准。Varfis 等人在上世纪九十年代对金融领域时间序列数据,引用人工神经网络进行分析^[1]。Hammad 在 2007 年使用了向前多层神经网络对众多公司的股价进行预测^[2]。Bertail 在 2009 年计算了马尔科夫链的极值并将该值应用到股票市场^[3]。张伟在 2010 年应用遗传算法优化的 SVM 进行股票的趋势预测^[4]。李巍在 2012 年从宏观经济的角度,采用机器学习的方法对上证综合指数进行预测^[5]。宋敏晶在 2013 年将情感分析应用于股票预测模^[6]。张希影在 2014 年通过遗传算法优化 BP 神经网络,可以防止 BP 神经网络模型陷入局部极小值点,从而可以有效预测股价^[7]。高天在 2015 年采用最优小波包变换,将股票价格序列进行分解,之后在采用 ARIMA 和 SVR 进行对股价的预测^[8]。张贵勇在 2016 年使用卷积神经网络来预测股票指数^[9]。吕涛等人在 2017 年把 K 线图的信息引入机器学习,通过 K 线滑动搜索算法对股票价格进行预测,取得不错的效果^[10]。余传明等人在 2018 年提出新的文本价格融合模型,通过对股票贴吧中的评论进行处理在使用 Kmeans 进行聚类分类,再结合股票多个价格指标,使用多层感知机对股票进行预测研究^[11]。冉杨帆等人在 2018 年采集股票新闻数据,利用 BP 神经网络和 SVR 对股价进行预测研究^[13]。谢琪等人在 2019 年用 Bagging 方法集成 8 个 LSTM 模型,应用该模型对中国股市进行预测研究^[14]。宋刚等人在 2019 年提出一种基于自适应粒子群优化的 LSTM 模型,通过 PSO 算法对 LSTM 模型的一些关键参数进行寻优,提高模型预测股价的精确度^[15]。王禹在 2019 年,应用 Boosting

算法将多棵决策树进行级联来预测股价，提高了预算精确度降低了均方误差^[16]。罗鑫等人在 2020 年提出多时间尺度 CNN-LSTM 模型，通过 CNN 和 LSTM 对时间尺度进行提取特征，应用于股票涨跌预测研究^[17]。肖雯在 2020 年使用多种机器学习算法对股票收益率进行预测^[18]。

1.2.2 股市选股研究现状

学者方面，骆桦和秦艳艳在 2011 年采用 Jegadeesh 方法，对上证 50 指数和深圳中小板企业板块的股票进行研究，认为我国的股票市场有中长期动量效应和反转效应，并且认为采用反转策略以及动量策略可以获取超额收益^[19]。陈健等人在 2011 年研究投资者不同的风险偏好，制定了四种投资组合的策略，结果表明投资股票数量在 30 个左右的时候，就可以消除非系统风险^[20]。林德发和杨潇宇在 2014 年利用 PCA 方法对可能影响普票价格的众多指标进行降维处理，建立了多因子模型，继而对沪深指数 300 内的成分股按照多因子模型选股进行投资组合，得出有效的多因子模型选股可以有很好的收益，跑赢大盘^[21]。柯原和郑双阳在 2014 年研究量化交易，发现可以通过价值投资和行业轮动两个结合，可以同时兼顾两种方法的优点，降低风险，还可以让资金使用效率更高，可以让投资具有更稳定的收益^[22]。巨红岩等人在 2015 年使用 SPSS 实证分析，得出股票资金强度与股价波段环比增长率呈线性趋势，可以利用股票资金流的强度来预测未来股价的短期趋势^[23]。黄宏运等人在 2016 年利用了多元线性回归的相关理论建立了相对收益率，再利用影响股票重要的十个因子建立了多因子选股模型^[24]。吴瑜琪在 2017 年利用因子分析在 TMT 板块实证，发现经过因子分析法选出的优质股票大部分收益都是不错的^[25]。Enke 和 Zhong 在 2017 年采用 PCA, KPCA 以及 RPCA 三种降维技术进行降维，对处理过后的数据作为输入变量与 ANN 结合来预测标准 500 指数的涨跌^[26]。刘述忠在 2018 年讨论了单一预测方法的局限性，并提出了一种名为 GM-RBF 神经网络的组合神经网络预测模型，相对单一预测模型，GM-RBF 神经网络预测模型的预测速度和收敛精度都得到提高^[27]。Hu et al 在 2018 年应用改进的正弦余弦算法与 BP 神经网络结合 (ISCA-BPNN)，同时引入谷歌趋势数据预测标准普尔指数以及道琼斯指数开盘价的变化方向^[28]。周亮在 2019 年认为过去的 Garch 以及 VAR 指标都不能很好的反应金融市场的风险，随着方法的进步，他选取了 IVX-QR 回归的方法，构造了 17 个预测变量，来对股票未来的股价进行预

测来研究市场风险因素^[29]。王慎敏在 2020 年通过利用随机森林的优点，利用多因子选股对中证 800 成分股的股票池进行挑选，选取优秀的股票进行投资回测，分别投资前 1/10, 1/5 的股票，获取的收益率也是远超大盘，最高的超额年化收益率也达到了 22%^[30]。

1.2.3 图卷积神经网络研究现状

深度学习具有极强的学习能力，同时还具有覆盖范围广适应性好、数据驱动上限高的优点，所以在诸多的领域，学者们开始使用图神经网络进行研究，图神经网络也成了最近几年的热点。随着互联网的发展，出现越来越多的非欧数据，也就是具有图形式的数据，例如：通信网络、社交网络、化学领域分子间结构、论文领域论文引用关系等^[31]。对于这种非欧空间的图数据，传统的 CNN、RNN 是不能处理的。学者们为了研究这类图数据，就应运而生了图神经网络。Gori 等在 2005 年提出一种用于图的神经网络模型，是神经网络在图数据的首次应用^[32]。在最近几年越来越多的学者开始对图神经网络进行研究，也因此提出了很多新的模型。Kipf 等在 2016 年提出一种基于频谱的图卷积网络模型（GCN），该模型结合图的结构信息使用图卷积方式来学习节点特征，并应用于半监督的分类任务^[33]。由于 GCN 需要结合全图全部节点信息进行节点特征的学习表示，所以该模型在处理大规模的图的时候效率非常低，而且还不能修改图的结构。为了解决效率问题，Hamilton 等在 2017 年提出一种基于空间的图卷积网络模型（GraphSAGE），该模型通过邻接点采样方法以消息传递来学习节点特征^[34]。Hamilton 等提出一种基于空间的图卷积网络模型（GraphSAGE），该模型通过邻接点采样方法以消息传递来学习节点特征^[34]。Velickovi 等在 2017 年提出了一种图注意力网络（GAT），在 GCN 网络中加入了注意力机制。通过这种方式可以让节点在聚合过程中舍弃一些非重要信息，可以让模型更好的学习节点特征^[35]。随着图神经网络的发展，越来越多的学者将图神经网络应用在不同的领域，比如交通、计算机视觉、股票价格的预测、推荐系统等。Xu 等在 2017 年提出一种场景图生成模型，通过对该模型输入图像，让模型自动进行识别与分析，预测对象之间的关系^[36]。Yan 等在 2018 年在动作识别领域提出了时空图卷积神经网络（ST-GCN）^[37]。李志帅等人在 2019 年提出图卷积神经网络和注意力机制的短时交通流预测模型^[38]。闫旭范晓亮等人在 2020 年提出改进的时空图卷积神经网络，

基于自由流动可达矩阵的时空图卷积深度神经网络 (FAST-GCN)，预测准确率优于 GCN^[39]。王天保等人在 2021 年基于 TP-GCN 建立人体交互模型，并进行轨迹预测^[40]。

1.3 研究意义

最优投资组合是基于均值方差理论提出的，通过对未来股票的收益率方差进行预测，通过计算出具有最小收益率方差的组合，即为最优投资组合。通过不同模型构建最优投资组合具有以下理论及现实意义。

(1) 理论意义

传统的统计模型和当前比较火的机器学习模型，都被应用于股票波动预测，两种的预测效果具有很大的差异。近年来，各种机器学习模型 CNN、RNN、LSTM 等模型在分类与预测方面都有非常好的性能，也因此也被广泛应用。应用传统的时间序列模型 GARCH 模型、机器学习时间序列预测模型 LSMT 模型与最近比较热的 GCN 模型对股票的波动进行预测同时构建最优投资组合比较少。本文分别使用三个模型进行收益率方差预测，基于预测的基础上计算未来的最优投资组合。对不同模型构建的最优投资组合的收益率进行对比研究。在学术研究中具有一定的价值。

(2) 现实意义

股票的波动反应了很多的生活现状，其中不仅包括公司的经营发展状况、所处行业的发展前景，还有宏观经济政策。同时研究最优投资组合不仅可以很好的研究股票的波动，还可以指导股票市场投资者投资行为，让投资者在进行投资时可以有效的避免风险获取稳定收益。

1.4 研究内容及安排

1.4.1 研究内容

本文采用 GARCH 模型、LSTM 模型、GCN 模型构建最优投资组合，在此基础上主要研究一下内容：

(1) 建立 GARCH 模型对选中的股票进行数据的检验，在数据通过检验后，

对数据进行 GARCH (1,1) 建模, 来对未来的收益率方差进行预测。基于模型预测数据, 计算未来五个交易日股票池中个股的最优投资权重, 构建最优投资组合, 计算组合收益。

(2)通过阅读文献, 选取收益率方差、开盘价、最高价、最低价、收盘价、成交量、成交金额七个指标进行 LSTM 模型构建, 预测股票收益率方差。基于模型预测数据, 计算未来五个交易日股票池中个股的最优投资权重, 构建最优投资组合, 计算组合收益。

(3)通过东方财富网, 筛选股票板块之间的关系, 再通过板块及之间的关系构建关联矩阵, 再将整理好的收益率方差数据放入 GCN 模型中进行训练, 来对未来的收益率方差进行预测, 基于模型预测数据, 计算未来五个交易日股票池中个股的最优投资权重, 构建最优投资组合, 计算组合收益。

(4) 通过对比 GARCH 模型、LSTM 模型、GCN 模型的收益对比, 分析三种方法在股市熊市、牛市、震荡区间的收益的差距, 同时对比三种方法的优劣处, 提出相对最优的投资组合方法。

1.4.2 结构安排

第一部分为引言.介绍选题的研究背景、研究现状及意义与运用的方法和创新之处等.

第二部分为预备知识.介绍了图卷积网络涉及的相关内容、均值方差理论以及最优投资组合概念。

第三部分为最优投资组合的实证。(1)基于最优投资理论, 开始对实证数据进行整理。(2)应用 GARCH 模型构建最优投资组合, 计算组合收益并进行分析。(3)应用 LSTM 模型构建最优投资组合, 计算组合收益并进行分析。(4)应用 GCN 模型构建最优投资组合, 计算组合收益并进行分析。

第四部分为三种模型构建的投资组合的对比分析.对比 GARCH 模型、LSTM 模型、GCN 模型构建的最优投资组合, 分析组合成分股以及组合收益率。

第五部分为研究的总结及展望.总结了全文要点,并且对未来研究方向和内容做了简单展望。

1.5 创新之处

针对于最优投资组合，本文主要的工作及创新在于以下三点：

第一，定义最优投资组合。通过对均值方差理论模型进行梳理，该模型是基于历史的收益率波动方差数据的计算出收益方差均值，用来判定股票未来的风险。而本篇文章则是在此理论之上，提出通过预测未来一天的收益率方差，来预测风险，通过最小化组合风险构造最优投资组合。这样可以有效的帮助投资者，规避风险实现最优投资。

第二，引入图卷积神经网络模型，构建最优投资组合。在机器学习比较热的情况下，各种神经网络都是被广泛使用。而图卷积神经网络作为新兴的网络，具有其独特的优势，可以很好的刻画股票之间的内在联系。同时图卷积神经网络在股市方差预测领域中研究较少，使用该方法在股票收益率方差预测是个创新。

第三，将图卷积神经网络同 GARCH 模型、LSTM 模型构建的最优投资组合作对比，分析构建三个模型构建组合的差异。分析 GARCH 模型、LSTM 模型以及 GCN 模型构建最优投资组合的优劣势。三种模型构建最优投资组合对比研究是个创新。

2 相关概念与方法基础

2.1 深度学习方法

2.1.1 卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNNs)是一种应用非常广的神经网络。早期的CNN由于受到计算机的限制以及实验所需标签数据的匮乏,很难通过训练得到性能好的网络。随着时代的进步,计算机的计算能力提升以及实验所需标签数据的不断发展,CNN在解决图像分类问题、图像识别问题中有突出的表现。所以,卷积神经网络是解决机器学习相关问题的一种常用且重要的工具。在现实生活中,CNN主要应用于识别领域,比如物体识别、人脸识别。

CNN主要由三部分构成(如图2.1),分别为(1)卷积层;(2)池化层;(3)全连接层。卷积层主要是进行卷积计算。通过将输入图片的像素矩阵与不同的卷积过滤器进行连接,以用来提取输入图片的各种特征。卷积过滤器的工作原理是通过在图片上进行移动,计算点乘同时计算总和,以此方式获取输入图片的各种特征。在现实生活中的数据大多数都是非线性的,而卷积计算是一种线性计算。所以为了让神经网络更好的拟合现实数据,就需要给该计算添加非线性特征。方式就是在卷积计算之后加上一个偏置参数,将结果输入非线性函数。非线性函数在机器学习领域中有tanh函数、sigmoid函数和Relu(Rectified Linear Unit)函数。在这三种非线性函数中,Relu函数相对于其他两种非线性函数来讲,在大多数的情况下具有更好的性能。池化层也被称为下采样层,作用是压缩特征图的维度,保留特征图中关键信息,可以防止训练的模型出现过拟合的现象。最常用的池化方法有和池化(sum pooling)、平均池化(mean pooling)和最大池化(max pooling)。最后一层则为全连接层,也可以看做传统的多层感知机。通过卷积层提取特征和池化层进行压缩后,原始的图像数据就相应的映射到特定的特征空间,获得最终有效特征。最后一层全连接层将这些高层次的特征作为输入,实现对目标图片所属类别的预测。

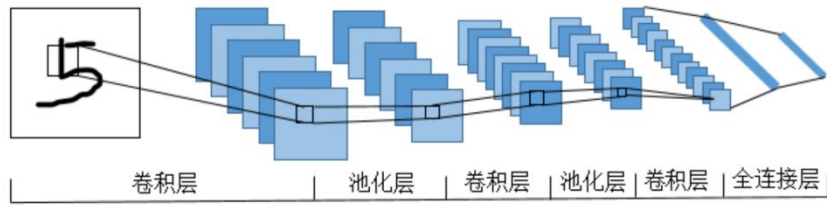


图 2.1 卷积神经网络示意图

2.1.2 循环神经网络

循环神经网络 (Recurrent Neural Networks, RNNs) 也是一种特殊的神经网络。RNN 是根据“人的认知是依据过往的经验和记忆”这一观点提出的，它与 CNN 不同之处在于，它不仅考虑前一刻时间的输入，而且还赋予了网络对于前面的记忆功能。如图 2.2 所示，RNN 在常规的前馈神经网络中加入了环状的连接，通过这样的方式，神经网络就有了“记忆”功能，可以动态捕捉到时序上的信息。因为 RNN 具有这样与众不同的功能，可以很好的分析时序数据，所以 RNN 广泛应用于时间序列数据预测领域，比如目标跟踪、动作识别、自然语言处理等任务。RNN 也存在一些问题，当 RNN 中存再多次递归会可能导致梯度爆炸、梯度消失，从而模型不能得到有效的训练。为了解决这个问题，LSTM 和 GRU 也被相继提出。这两个神经网络引入多个门操作，通过多个门分别保留重要的信息以及选择丢弃不重要的特征。

输出

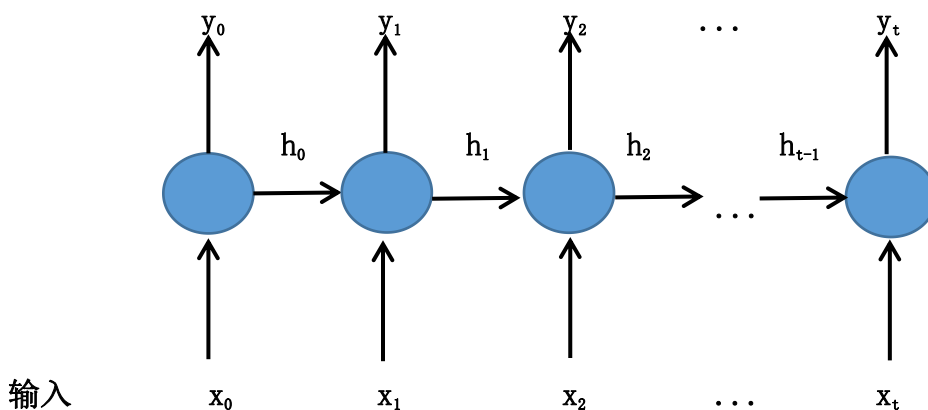


图 2.2 循环神经网络示意图

2.1.3 长短期记忆神经网络

长短期记忆网络 (Long-short Term Memory, LSTM) 是一种特殊的循环神经网络, 可以很好的处理长期依赖的问题。众多研究学者的研究结果表明, LSTM 确实解决了很多 RNN 无法解决的问题, 在时间序列预测研究中取得了进一步的成功。LSTM 神经网络通过增加一个信息存储记忆单元, 可以保持一个持续信息流, 让模型在训练的时候不会出现梯度爆炸或消失的情况。同时针对记忆单元, 还构造了三个门对其进行控制, 分别为输入门 (Input Gate), 输出门 (Output Gate) 和遗忘门 (Forget Gate)。这几个门的作用相当于滤波器, 通过输入门更新记忆单元的状态, 通过输出门控制 LSTM 的输出, 遗忘门则是对记忆单元中的信息进行判断, 选择丢弃或着是保留。LSTM 神经网络与循环神经网络有类似的结构, LSTM 与之不同是其内部各个模块是协同工作的。LSTM 神经网络结构图 2.3 如下。

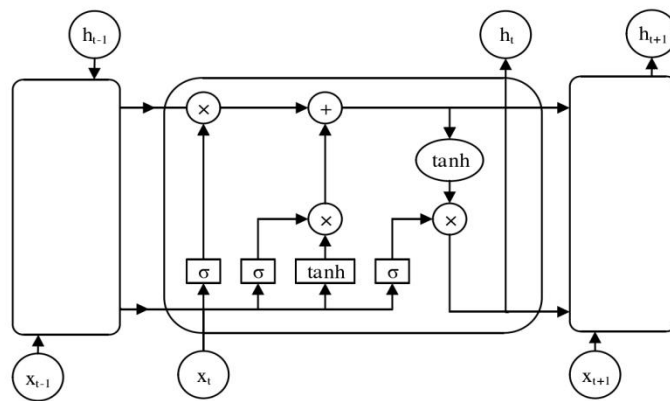


图 2.3 LSTM 内部图

2.1.4 GRU 神经网络

GRU(Gated Recurrent Unit, GRU)循环神经网络一种特殊的 LSTM。GRU 循环神经网络在应用中与 LSTM 神经网络十分相像。GRU 神经网络的目标就是构建效果与 LSTM 相同，与此同时运行效率要更快。它的工作原理就是通过降低门控数量，将三个门减掉一个余下两个门。因为门的数量减少，训练计算过程的参数减少，运行效率也因此得到提升。从整体结构看，GRU 神经网络与简单的 LSTM 模型没有任何区别，都是将基本输入与之前单元的状态通过计算传递给下一个单元，充分保留了 LSTM 的优势。GRU 的整体结构图如 2.4 所示：

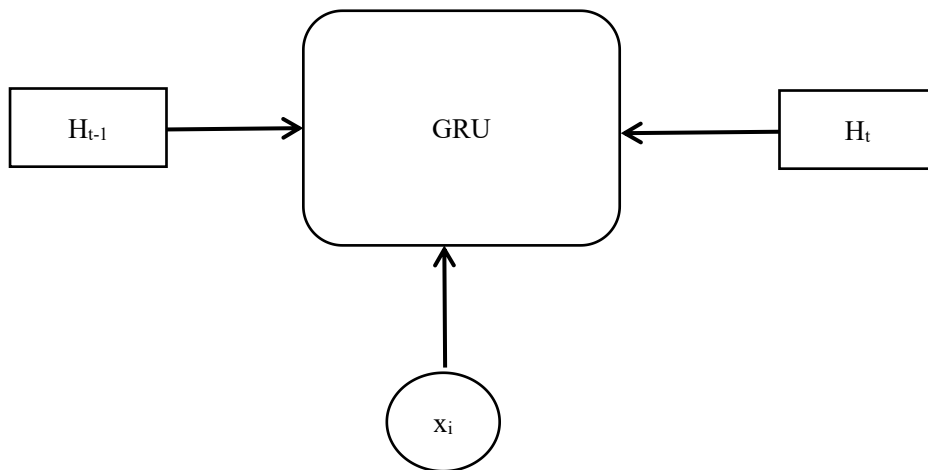


图 2.4 GRU 内部结构

LSTM 神经网络有三个门控，GRU 神经网络则是将这三个门改为两个门，分别为：更新门和重置门。这两个门的控制都是通过 sigmoid 函数，范围在 0-1 之间。更新门的计算结果越大，那么就会更多的记忆当前候选集的信息。重置门的计算结果越大，则代表上一步传入的信息就在该单元保留的更多。相比 LSTM 神经网络，GRU 神经网络几乎可以达到一样的训练结果。因为 GRU 神经网络少了一个门控，导致计算参数的降低，可以让 GRU 神经网络更高效。

2.1.5 图神经网络

图神经网络(Graph Neural Networks, GNNs)是一种重要的深度神经网络。CNN 在处理具有网状结构的数据时,通过进行卷积过滤器获取数据特征。但是在现实生活中,有很多的数据是不规则的结构,不能通过网络状结构表示。例如通信网络、骨骼架构、社交网络。这些不具备网状结构的数据也只能通过图的形式表示。此时,有的学者就提出了图神经网络,用来处理一般的图,比如:有向图、无向图。GNN 是把图与神经网络相结合,通过网络中节点的链接,实现节点间信息通信。节点间的通信由一个迭代过程实现,通过传播节点信息直到达到节点状态平衡,再利用神经网络,将节点状态转化为节点的对应输出。

2.1.6 图卷积神经网络

图卷积网络(Graph Convolutional Network, GCN)是一种特殊的深度神经网络结构,近年来非常流行。与传统的 LSTM 和 CNN 相比,GCN 拥有其独特的优势。LSTM 和 CNN 只能处理网状结构的数据(grid-based)。GCN 却不同,它可以处理具有拓扑图结构的数据,例如通信网络、蛋白质分子结构、社交网络等等。GCN 可以深入分析它们的特征^[41]。

GCN 与 CNN 在图的提取相同,在进行图的特征提取的时候,可以使用多层神经网络结构对于每一层,可以使用如下的映射函数来计算:

$$f(H^{(l)}, A) = f(H^{(l)}, A)$$

其中 $H^{(1)} \in \mathbb{R}^{N \times d(1)}$ 表示第 1 层图的节点级(node-level)表达,是一个 $N \times d(1)$ 维的矩阵, N 代表节点数量, $d(1)$ 代表第 1 层节点表达的维度(节点表达的维度可以在每一层都不一样,由 f 决定,可以灵活设置); $H(0) = X$ 表示第 0 层的初始化节点表达矩阵;假设一共有 L 层网络,则 $H(L) = Z$ 表示最后一层输出的节点表达矩阵。与 CNN 类似,GCN 也采用共享权重。与 CNN 不同的是,每个核的权重都是矩阵,按照对应的位置进行分配图卷积中的权重通常是一个集合。在对一个节点计算聚合特征值时,按一定规律将参与聚合的所有点分配为多个不同的子集,同一个子集内的节点采用相同的权重,从而实现权重共享。

概括的来说，图卷积操作就是将每个节点的特征与其邻居节点的特征加权平均后传播到下一层。这种图卷积操作称为在空域上的图卷积，有如下几个特点：

- (1) 随着层数的加深，每个节点能聚合到的特征越远。
- (2) 权重是共享的，不会具体到每个节点，这 and 传统 CNN 相同。直观的理解如果权重是因节点而不同的，那么一旦图结构发生变化，权重就会立刻失效。
- (3) 每个顶点的邻居节点数可能不同，这导致邻居节点多的顶点的特征值更显著。
- (4) 邻接矩阵在计算时无法将节点自身的特征包含到聚合特征值中。

GCN 与其他网络相比具有很多的优点。比如在图分类的时候，传统的机器学习算法需要对图的特征进行打标签，例如颜色、纹理或者更难标记的一些特征。而 GCN 只需要一张图，利用模型就可以自动提取特征，这样极大的减少的任务的工作量。其次，在生活中很多的数据都是非欧数据，使用一般的卷积网络不能精准的提取特征，比如社交网络、电商网络、人体结构等。GCN 此时就可以起到非常大的作用，GCN 能够精准的利用节点特征，非常好的提取这些网络图结构特征。其次 GCN 还具有参数共享、复杂度低、局部连接等优点。学者们将 GCN 应用到众多的领域。

2.2 投资组合理论

2.2.1 均值方差理论

投资组合理论的中心思想是在资金总量一定的情况下，通过调节不同资产的配置权重，获得组合资产最大回报。可以解释为投资者在一定风险水平下，实现投资者的期望收益最大化。投资策略有两种，可以分为主动与被动投资策略两种。主动投资策略更倾向于投机行为，通过对股票市场的信息收集，预测股价未来的变动趋势及时调仓，已获得超额收益。而被动投资策略则认为市场是有效的，不存在套利行为，也没办法通过历史信息预测未来的股价，因此他们会选择股票并进行长期持有。利用投资组合理论进行资产配置属于主动投资策略。

哈里·马科维茨在 1952 年开创了投资组合理论，提出均值方差模型。在该模型中引用期望效用理论，将一段时间内投资组合的收益的均值作为投资组合的

收益水平,用该时间内的收益率波动作为风险水平,既收益率的方差作为风险指标。在此基础上,对组合的风险水平进行规划求解,以实现最优投资组合。该模型假设如下:

(1)投资者在考虑每一次投资选择时,其依据是某一持仓时间内的证券收益的概率分布。

(2)投资者是根据证券的期望收益率的方差或标准差估测证券组合的风险。

(3)投资者的决定仅仅是依据证券的风险和收益。

(4)在一定的风险水平上,投资者期望收益最大;相对应的是在一定的收益水平上,投资者希望风险最小。

基于模型的假设可以看出,投资者都是期望收益一定的时候承担最小的风险。由此该模型可以看做一个二次规划模型。

在满足前提假设的情况下,Markowitz 均值一方差模型可以用如下公式表达:

$$\text{s. t.} \begin{cases} \sigma_p(W) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} = \theta \\ w_1 + w_2 + w_3 + \dots + w_n = 1 \end{cases}$$

或者

$$\begin{aligned} \max R_p(W) &= \sum_{i=1}^n w_i r_i \\ \begin{cases} R_p(W) = \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} r_i = r_e \\ w_1 + w_2 + w_3 + \dots + w_n = 1 \end{cases} \end{aligned}$$

通过线性规划,就可以为投资者在期望收益一定的情况下,配置风险最小投资组合。

2.2.2 最优投资组合理论

现代投资组合理论的开山之作是 Markowitz 的“均值—方差”理论。在该理论下,投资者都为理性人、市场具有流动性、投资者的资金具有完全的流动性、投资者们只会关注投资组合带来的风险与投资组合所面临的风险。他通过用资产

收益率方差定义投资风险，用资产收益率的均值代表期望回报率，通过二次规划计算出有效投资组合并得出了关键结论：可以通过多样化的投资来规避风险，尽量选择关联性比较弱的证券进行组合。在此之后，他通过计算资产收益率方差，求出了证券组合的有效边界^[42]。

虽然 Markowitz 的理论容易理解，也被投资者们接受。但是投资者们的投资组合都不太相同。对于投资者讲，股市的预期风险时刻影响着投资者的方案。在本篇文章考虑了这一点，基于均值方差理论下，提出了一种最优投资组合，即通过对资产的合理分配给不同的股票，来实现投资的风险最小进行投资，从而在股市中获利。

股票的投资组合是中含有 N 支股票，而受环境、政策、行业等因素影响，这 N 支股票的价格往往都存在一定的联系。比如有一支股票价格上涨，往往会出现其他的股票价格下跌或上涨，在统计学中，通常用相关系数或者协方差来表达这种关系。类似于单一证券，在 Markowitz 投资组合理论中，证券组合风险是通过整个组合方差进行衡量，对于投资组合中 a 股票和 b 股票，假设其预期收益率分别是 r_{ea} 和 r_{eb} ，而每期的实际收益率分别为 r_{at} 和 r_{bt} ，则 a 和 b 之间的协方差就可表示为：

$$\sigma_{ab} = \frac{1}{N} \sum_{n=1}^N (r_{at} - r_{ea})(r_{bt} - r_{eb}) \quad (3-1)$$

如公式(3-1)所示，如果 $\sigma_{ab} > 0$ ，则说明 a 股票和 b 股票的价格是正相关的，既 a 股价格上升， b 股价格也上升，二者的收益率呈同方向变动；如果 $\sigma_{ab} = 0$ ，则说明 a 股票和 b 股票的价格是不相关的，既 a 股票价格上升或下降不会影响 b 股票价格，二者的收益率无关联；如果 $\sigma_{ab} < 0$ ，则说明 a 股票和 b 股票的价格是负相关的，既 a 股价格上升， b 股价格下降，二者的收益率呈反方向变动；

在证券组合中，往往有多种可选择的证券，假设证券组合中含有 M 种证券，则需要考虑 M 种证券中任意证券之间的协方差，而已经知道证券组合的预期收益率，所以证券组合的风险可以表示如下：

$$\sigma_p^2 = D(r_p) = \sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j \text{cov}(\omega_i, \omega_j) = \sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j \sigma_i \sigma_j \rho_{ij} \quad (3-2)$$

其中 σ_p^2 —证券组合的方差，也即证券组合风险的量化； σ_{ij} ——第 i 种和第 j

种资产关于收益率的协方差; ρ_{ij} ——第 i 种和第 j 种资产关于收益率的相关系数, 且满足 $-1 < \rho_{ij} < 1$ 。

本篇文章是基于 Markowitz 的理论, 并且在此理论基础上进行研究。本篇文章的最优投资组合构建就是在 Markowitz 的理论, 对选中的股票池中的数据进行投资组合的构建, 通过最小化投资组合的收益率方差来求出投资权重。而与原理论不同的是, 本篇文章不是采用历史数据的方差来构建投资组合, 而是采用模型预测未来一天股票池中每支股票的方差, 既在 $t-1$ 日预测 t 日的股票池中的每支股票的方差。预测的方差记为 σ'_i , 最优投资组合则是基于预测方差, 对未来一日的组合方差进行最小化求解, 得到股票池中每支股票最优的投资权重。该权重记为风险最小最优投资组合权重。具体公式如下:

$$\min (\sigma'_p)^2 = D(r'_p) = \sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j \sigma'_i \sigma'_j \rho_{ij} \quad (3-3)$$

公式 (3-3) 中 σ'_i, σ'_j 为预测股票的方差, σ'_p 为基于预测方差求得最小方差, ρ_{ij} ——第 i 种和第 j 种资产关于收益率的相关系数。 ω_i, ω_j 为第 i 支, 第 j 支股票的最优投资权重。

通过最小化预测方差求得求得投资组合解, $\omega_i (i=1, 2, 3, \dots, n)$ 既为投资组合最优投资权重的解, 可以理解为在将资金按照该权重进行投资, 所面临的风险最小。对第 $i (i=1, 2, 3, \dots, n)$ 支股票投资 $\omega_i (i=1, 2, 3, \dots, n)$ 投资金额, 可以达到投资风险最小, 即为最优投资方案。该方案的股票组合即为最优投资组合。

3 最优投资组合实证研究

本章节首先是在最优投资组合理论下，对股市的数据进行整理，分别使用 GARCH 模型、LSTM 模型以及 GCN 模型对股票风险进行预测，预测未来股票池个股的风险。通过对个股未来的风险预测，进行未来一日的最低风险最优投资组合的构建。最后分别对不同模型的最优投资组合的收益进行分析。

3.1 数据整理

在目前，我国 A 股上市的股票有 4000 多支。上证 50，是根据科学客观的方法，挑选出最具有代表性的 50 支股票。这 50 支股票是具有流动性好、规模大的特点，可以很好的反应上海证券市场龙头企业的整体状况。因此我们选取的为上证 50 中所包含的股票，分别为浦发银行（600000）、中国石化（600028）、中信证券（600030）、三一重工（600031）、招商银行（600036）……。因为我国的股票市场还是处于发展期，还未成熟，所以本篇文章建立模型的时候会分成熊市、牛市、以及股市的震荡间区域来进行研究。为了研究全面，本篇文章选择的数据的时间段为 2013 年 7 月至 2021 年 6 月。股市牛市熊的划分是依据搜狐财经的报导，2013 年 7 月 1 日-2015 年 6 月 12 日为牛市，2015 年 6 月 13 日-2018 年 12 月 31 日为熊市。在 2018 年底年以后没有明确的牛熊市之分，故将 2019 年之后定义为震荡区间，并且选取震荡区间 2019 年 1 月 1 日-2021 年 6 月 30 日。在这段时间内，有上证 50 股票池中的部分股票数据不足，所以进行剔除操作，剔除的分别 601066(中信建投)、601138(工业富联)、601211(国泰君安)、601728(中国电信)、601995(中金公司)、603259(药明康德)、603288(海天味业)、603501(韦尔股份)、603986(兆易创新)这九只股票，剩余 41 支股票。

本篇文章采用的图卷积神经网络可以提取利用股票之间的相互关系，因此剩下 41 支股票中，筛选板块关联多的股票作为本篇文章投资组合构建的股票池。在 41 支股票中，选取所在地域板块，上海板块、北京板块、广东板块的二十二只股票，分别为 600000(浦发银行)、600028(中国石化)、600030(中信证券)、600031(三一重工)、600036(招商银行)、600048(保利发展)、600050(中国联通)、600104(上汽集团)、600196(复星医药)、600588(用友网络)、600837

(海通证券)、600900(长江电力)、601088(中国神华)、601288(农业银行)、601318(平安银行)、601336(新华保险)、601398(工商银行)、601601(中国太保)、601628(中国人寿)、601668(中国建筑)、601857(中国石油)、601888(中国中免)作为本篇文章投资组合的股票池。

基于2013年7月1日至2021年6月31日,股票市场分为牛市、熊市、震荡区间划分。这个对于现在来讲是已知信息,而在现实生活中投资者是不能预判牛熊市的结束时间,只能知晓市场是否处于牛熊。因此本篇文章要研究在在熊市其内部时间进行研究,而没选择整个周期。所以本篇文章熊市研究区间随机选取为2013年7月1日至2015年4月8日,熊市的时间选取为2015年6月13日的与2018年4月30日的数据,震荡区间选取2018年12月31日至2021年6月30日。

3.1.1 股价走势分析

(1)牛市股票价格走势

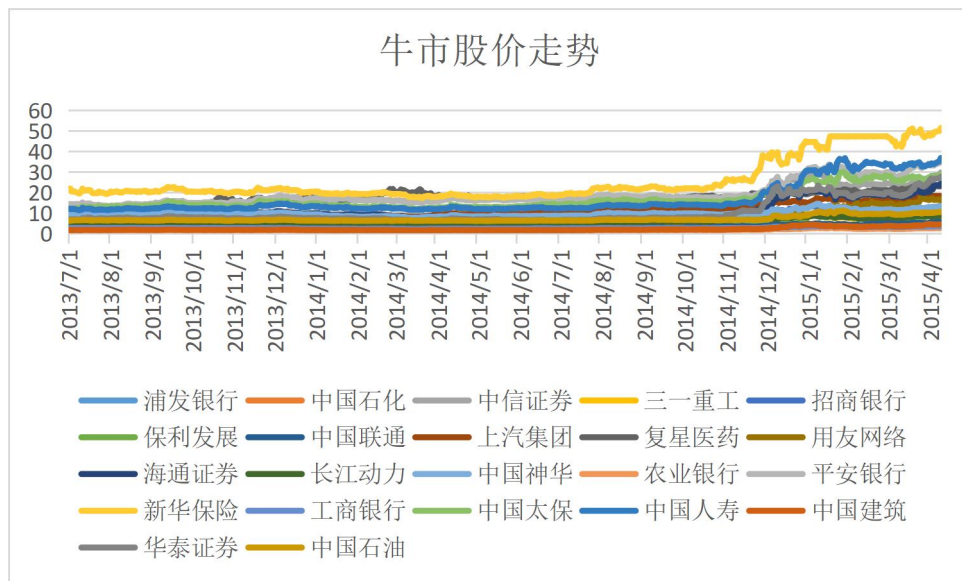


图 3.1 牛市股价走势图

从图 3.1 为股票池中 22 支股票的牛市股价走势图。从图中可以看出,在 2013 年 7 月 1 日至 2015 年 4 月 1 日,股票池中所有股票价格走势都是向上的。其中上涨趋势非常明显的有三一重工、招商银行、保利发展和中信证券,这些股票都实现了翻倍的增长。

(2)熊市股票价格走势

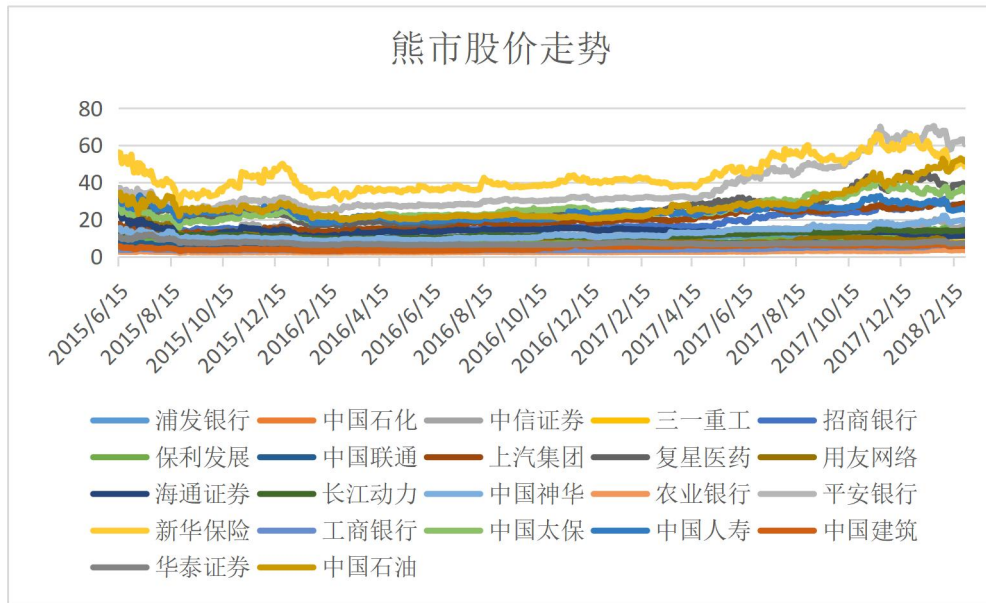


图 3.2 熊市股价走势图

从图 3.2 为股票池中 22 支股票的股价走势图。从图中可以看出，在 2015 年 6 月 15 日至 2018 年 12 月 31 日，股票池中所有股票价格总体走势都是向下的。在 2015 年 8 月 15 日，所有股票的走势是急剧下降，之后的一个月稍有反弹，之后又呈下降趋势。在 2017 年 6 月之后，股价整体走势才有向上的趋势。

(3)震荡区间股票价格走势

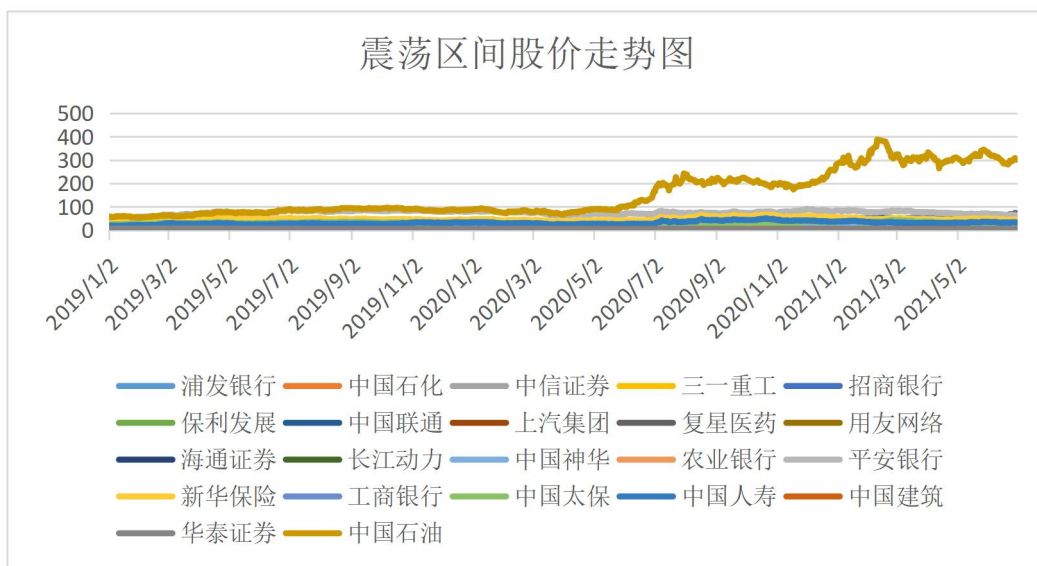


图 3.3 震荡区间股价走势图

从图 3.3 为股票池中 22 支股票的股价在股市震荡区间走势图。从图中可以看出，在 2019 年 1 月 2 日至 2021 年 6 月 30 日，股票池中所有股票价格走势都是趋于平稳的。其中中国石油走势非常好，呈向上的趋势。在 2020 年 7 月 2 日有迅速上涨的趋势。

3.1.2 股票收益率分析

(1) 牛市股票收益分布

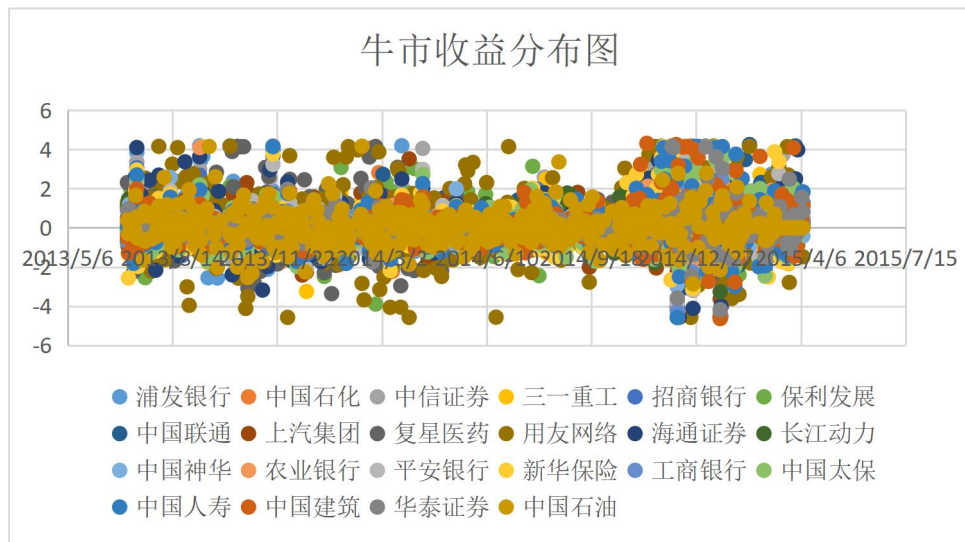


图 3.4 牛市股票收益分布

从图 3.4 为股票池中 22 支股票的牛市收益分布图。从图中可以看出，在 2013 年 7 月 1 日至 2015 年 4 月 6 日，整体市场的收益率大多数为正。其中波动范围是在 -5 至 5，可以看到市场具有很强的赚钱效应。

(2)熊市股票收益分布

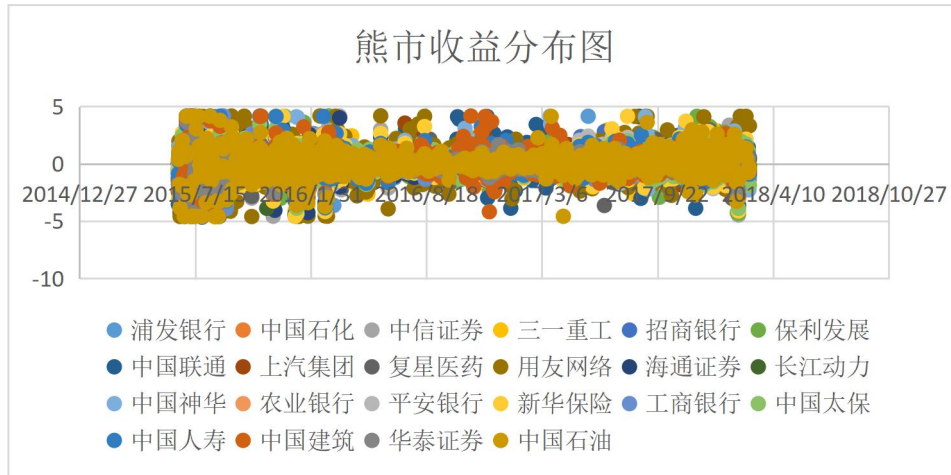


图 3.5 熊市股票收益分布

从图 3.5 为股票池中 22 支股票的熊市收益分布图。从图中可以看出，在 2015 年 6 月 17 日至 2018 年 2 月 28 日，整体市场的收益率大多数为负。其中波动范围是-1 的偏多。在市场进入熊市的时候，负收益率更密集。

(3)震荡区间股票收益分布

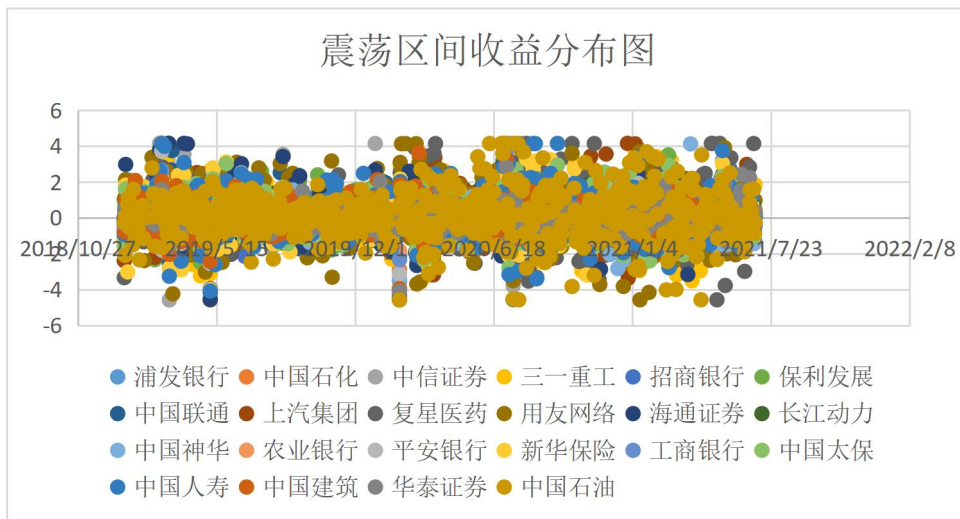


图 3.6 震荡区间股价收益分布

从图 3.6 为股票池中 22 支股票的震荡收益分布图。从图中可以看出，在 2019 年 1 月 2 日至 2021 年 6 月 30 日，市场收益率呈均分的趋势。正值与负值的数量趋于平等。

3.2 股票风险估计

3.2.1 GARCH 模型股票风险估计

3.2.1.1 数据整理

在 3.1, 已经对数据进行了初步的处理, 观察了股票价格的走势以及股票的收益率的分布。在本小节, 开始对数据进一步整理。首先分别将牛市、熊市、震荡区间研究的股票数据放入到 eviews10.0 中, 接下来对股票池中的二十二支股票收益率方差既风险进行 GARCH 建模, 分别对牛市、熊市、震荡区间的股市个股的风险进行预测。

3.2.1.2 自回归条件异方差模型建立

在本章节, 首先对 22 支股票从 2013 年 7 月 1 日至 2021 年 6 月 30 日的收益率方差数据进行单位根检验, 其次对数据进行自相关检验、JB 检验和异方差效应检验。在表 3.2 中, 为 22 支股票的收益率单位根检验结果, P 值近似于 0, 拒绝原假设。之后对数据进行自先关检验, 结果显示也都无自相关。之后进行 JB 检验, 可以检验结果的 P 值都是接近于 0 的, 说明拒绝了正态分布的假定说明。之后进行异方差效应检验, 结果也是拒绝原假设, 因此可以使用 GARCH 模型。

表 3.1 股票池部分熊市的收益率方差预测数据

ADF 检验	t-Statistic	Prob.*	t-Statistic	Prob.*	
浦发银行	-43.78431	0.0001	海通证券	-42.64723	0.0001
中国石化	-42.6005	0	长江动力	-47.5341	0.0001
中信证券	-41.91417	0	中国神华	-43.35629	0.0001
三一重工	-43.10177	0.0001	农业银行	-34.95405	0
招商银行	-44.17054	0.0001	平安银行	-44.44116	0.0001
保利发展	-43.29646	0.0001	新华保险	-43.30061	0.0001
中国联通	-43.55901	0.0001	工商银行	-33.82772	0
上汽集团	-44.18073	0.0001	中国太保	-43.92436	0.0001
复星医药	-44.723	0.0001	中国人寿	-42.69339	0.0001
用友网络	-42.11214	0.0001	中国建筑	-42.0619	0
海通证券	-42.64723	0.0001	华泰证券	-33.21853	0
长江动力	-47.5341	0.0001	中国石油	-42.57596	0.0001

3.2.1.3 模型预测

基于数据的整理，紧接着我们通过对 GARCH 模型建模，选取最优的 GARCH (1,1) 模型，进行牛市、熊市、震荡区间的收益率方差的预测。为了计算方便，在计算收益率方差的时，进行扩大 100 处理，公式如下：

$$\sigma_i = ((\log p_t - \log p_{t-1}) * 100)^2$$

(1) 牛市个股风险估计

牛市选择研究期间为 2013 年 7 月 1 日到 2015 年 3 月 31 日的数据为训练数据，对未来的 5 个交易日即 4 月 1 日、4 月 2 日、4 月 3 日、4 月 7 日、4 月 8 日，进行 Garch (1,1) 建模，从而预测未来的收益率方差，预测该风险。

表 3.2 股票池部分牛市的收益率方差预测数据

	2015/4/1	2015/4/2	2015/4/3	2015/4/7	2015/4/8
浦发银行	6.838359516	6.134370803	5.705228395	4.997045945	5.914004913
中国石化	2.873135324	2.720178422	2.510602863	2.344905677	3.201361327
中信证券	7.203829213	7.866389961	7.245319194	7.152796607	6.543283572
三一重工	10.15097105	9.500723464	8.876038101	8.252310401	8.275891347
招商银行	3.591030431	3.380147061	3.102994575	2.842584029	2.611006197

表 3.2 通过 GARCH(1,1) 对股票池中股票收益率方差进行预测，每一日的收益率方差的预测都是基于之前的历史数据。基于不断更新的已知数据，分别通过五次计算，计算出了未来五日的收益率方差。从表 3.2 中可以看到，股票池数据的收益率方差都在 2-10 之间，从部分数据中可以看到，其浦发银行的方差相对于较小，中信证券的收益率方差较大。

(2) 熊市个股风险估计

熊市研究选取的时间为 2015 年 6 月 13 日到 2018 年 2 月 21 日的数据为训练数据，来对未来的 5 个交易日进行 GARCH (1,1) 建模，预测未来的收益率方差，预测 22 支股票的风险。

表 3.3 股票池部分熊市的收益率方差预测数据

	2018/2/22	2018/2/23	2018/2/26	2018/2/27	2018/2/28
浦发银行	3.237815813	3.174211424	2.436664231	1.660822007	1.226430253
中国石化	9.337123386	9.050546064	8.190062122	7.148813964	6.343254911
中信证券	13.75066292	13.21838638	12.45571318	12.16063545	11.74788814
三一重工	5.927108953	6.276792162	5.919604305	5.835479504	5.979860172
招商银行	5.837860779	5.440937605	5.111652142	4.814816993	4.788910877

从表 3.3 可以看出,在未来的五个交易日中,整体的股票的收益方差都不大,大多数处于 3-14 附近,部分可以高达 13.7 左右。在部分展示中的数据,可以看到中信证券的方差大,风险高。

(3) 震荡区间个股风险估计

震荡区间研究选取 2019 年 1 月 2 日到 2021 年 6 月 23 日的数据为训练数据,来对未来的 5 个交易日进行 GARCH (1,1) 建模,从而预测未来的收益率方差,预测该风险。

表 3.4 股票池部分震荡区间的收益率方差预测数据

	2021/6/24	2021/6/25	2021/6/28	2021/6/29	2021/6/30
浦发银行	1.452757355	1.43222911	1.419019764	1.421973228	1.429464642
中国石化	3.876516527	3.315263248	2.863593008	2.456177667	2.235072537
中信证券	4.803305282	4.656386167	4.604189003	4.832424654	4.843754311
三一重工	5.848192032	5.53333372	5.21756083	5.504030503	5.304716424
招商银行	3.81665941	3.45319209	3.021795218	3.40218095	4.167594088

从表 3.4 可以看出,在未来的五个交易日中,股票池中的整体收益率方都比较小,大多数处在 1-6 附近。在整个收益率方差最大的为三一重工,预测方差都在 1-6 之间,最高达到 5.8。可以看到在这段时间,三一重工的投资风险是最大的。

3.2. 2LSTM 模型股票风险估计

LSTM 模型构建最优投资组合，基本流程如图 3.7 所示。

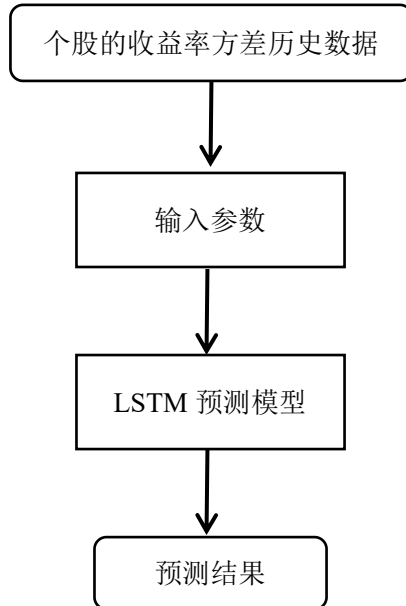


图 3.7 LSTM 模型构建最优投资组合流程图

3.2.2.1 LSTM 模型设计

LSTM 模型预测同金融时间序列的预测方式类似，需要通过股票的历史数据中获得相应的数据特征作为模型参数。在此本篇文章通过参考其他文献，选取了 7 个参数 $x=x_1, x_2, x_3, \dots, x_7$ 作为预测模型的训练数据股价收益率方差、收盘价、开盘价、最高价、最低价、成交量、成交金额。整体的思路就是通过学习股票前 N 天数据中的特征，根据学习到的规律预测第 N+1 天的数据。这七个参数中，后六个参数都可以通过调用聚宽 API，进行获取。收益率方差则是依据股票价格进行计算，为了减少计算压力，因此在计算收益率的时候进行放大 100 处理，公式如下：

$$x_{1t} = ((\log(p_t) - \log(p_{t-1})) * 100)^2$$

这里需要注意的是选取的这七个初始指标是不能直接放入神经网络函数作为参数。需要对数据格式进行转换，输入的格式利用矩阵与空间向量的形式来表示：

$$X = \begin{cases} X_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,7}) \\ X_N = (x_{N,1}, x_{N,2}, \dots, x_{N,7}) \end{cases}$$

这个矩阵的含义是每条用于训练的数据内有 N 天的股票的历史数据，并且每天的历史数据都有 7 个特征值，分别对应上文的八个参数，每天的历史数据用一个向量进行代指。最后通过学习前 N 天的历史数据，预测第 $N+1$ 天的收盘价。

在 LSTM 的模型中输入数据的方式如图 3.8:

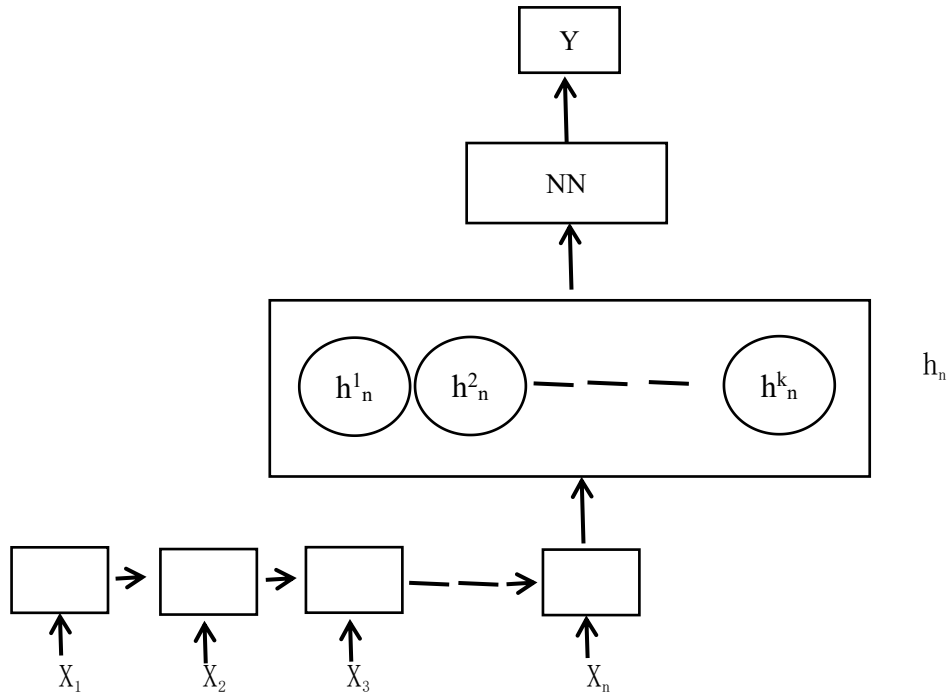


图 3.8 LSTM 模型输入数据方式图

在图中 LSTM 模型，输入的数据 N 个为一组，每一组代表一天的股票历史数据。每组共有八个参数，分别对应收益率方差、收盘价、开盘价、最高价、最低价、成交量、成交金额。每一个输入向量 X 经过神经网络的训练后都有一个输出 y ，即图中的 h_n 构成的向量几何。由于这里我们最终需要第 n 天的收益率方差最为最后的输出，因此将得到的集合经过一个单层的神经网络，即 NN 进行转换，得出第 n 天的收益率方差 y 作为预测模型输出。

本节在模型训练的过程中采用交叉验证的方式，以此来调节参数，划分出 80% 训练集，20% 验证集，在模型构建过程选取平均平方误差作为损失函数，利用梯度下降法更新权重。

3.2.2.2 模型预测

通过构造 LSTM 网络，设置好相关的参数，我们开始将整理好的数据放入到 LSTM 模型进行训练。通过 LSTM 模型来对股票池中股票收益率方差进行预测，每一日的收益率方差的预测都是基于对应日期之前的历史数据。基于不断更新的已知数据，分别通过五次计算，计算出了未来五日的收益率方差。从而预测未来的收益率方差。在下面讲分成三个市场，牛市、熊市、震荡区间分别进行收益率方差的预测。为了计算方便，在计算收益率放大 100 之后再计算收益率方差。

(1) 牛市个股风险估计

牛市选择研究期间为 2013 年 7 月 1 日到 2015 年 3 月 31 日。选择这个时间段的数据为训练数据，对未来的 5 个交易日即 4 月 1 日、4 月 2 日、4 月 3 日、4 月 7 日、4 月 8 日，进行 LSTM 建模，从而预测未来的收益率方差，预测个股风险。

表 3.5 股票池部分波动率方差数据

	2015/4/1	2015/4/2	2015/4/3	2015/4/7	2015/4/8
浦发银行	1.5142897	1.2404152	1.5367426	2.449369	2.1466305
中国石化	0.27787456	0.31300646	0.090449676	0.6273768	2.4877205
中信证券	0.43024054	1.7760105	2.059929	1.8980938	3.4977264
三一重工	0.6067964	0.061192803	0.3306683	0.4229284	0.54259145
招商银行	0.10624859	1.1145644	1.9877229	2.4018135	6.444195

从表 3.5 中可以看到，股票池数据的收益率方差都在 1-3 之间，从部分数据中可以看到，其中三一重工的方差相对于较小，更平稳一些。

(2) 熊市个股风险估计

熊市选择研究期间为 2015 年 6 月 13 日的与 2018 年 2 月 14 日。选择这个时间段的数据为训练数据，对未来的 5 个交易日即 2018 年 2 月 22 日、2 月 23 日、2 月 26 日、2 月 27 日、2 月 28 日，进行 LSTM 建模，从而预测未来的收益率方差，预测风险。

表 3.6 股票池部分熊市预测收益率方差数据

	2018/2/22	2018/2/23	2018/2/26	2018/2/27	2018/2/28
浦发银行	0.03599413	0.4382906	0.19104871	0.52598006	0.10326708
中国石化	0.1274758	0.12713435	0.141626	0.18546133	0.07824266
中信证券	0.34972608	0.35385293	0.16059865	0.47258306	0.39415067
三一重工	4.472513	8.846286	0.26835853	3.3067186	0.0985145
招商银行	0.4334404	0.5356803	0.42313227	0.41319016	0.24967784

从表 3.6 可以看出,在未来的五个交易日中,整体的股票的收益方差都不大,大多数处于 1 附近,部分达到 3。在部分展示中的数据,可以看到中信证券的预测得风险偏高,数值更大。

(3) 震荡区间个股风险估计

震荡区间选择研究期间为 2019 年 1 月 2 日到 2021 年 6 月 23 日。选择这个时间段的数据为训练数据,对未来的 5 个交易日即 2021 年 6 月 24 日、6 月 25 日、6 月 28 日、6 月 29 日、6 月 30 日,进行 LSTM 建模,从而预测未来的收益率方差,预测风险。

表 3.7 股票池部分震荡区间预测收益方差数据

	2021/6/24	2021/6/25	2021/6/28	2021/6/29	2021/6/30
浦发银行	0.25747758	0.28477782	0.2530001	0.15998608	0.3804825
中国石化	0.6938724	0.04834161	0.11743864	0.12865847	0.09527114
中信证券	0.46383303	0.53931695	0.43454045	0.1471913	1.8441491
三一重工	1.539991	2.222506	0.8624191	2.5359898	1.6941901
招商银行	0.2962794	0.40175223	0.8996085	0.31492013	0.5177361

表 3.7 显示在未来的五个交易日中,预测整体的方差还是在 0-2 之间,中信证券会稍微高一些在 2-3 之间。从整体上看,中信证券的收益率方差是稍高一些,具有更大的风险。

3.2. 3GCN 模型股票风险估计

GCN 模型构建最优投资组合，基本流程如图 3.9 所示。

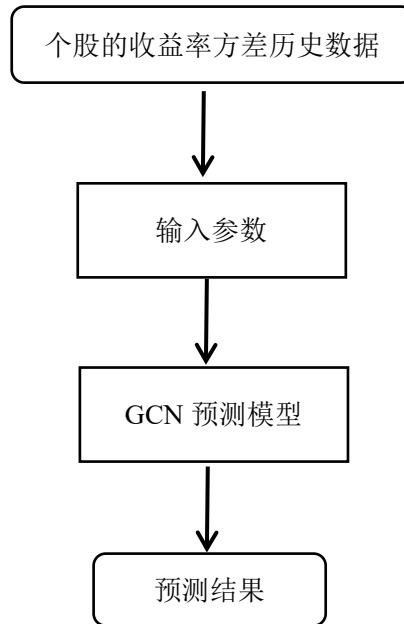


图 3.9 GCN 模型构建最优投资组合流程图

3.2.3.1 图的构建

基于数据整理，开始进行图卷积神经网络中的知识图的构建。在数据处理的时候，选出了北京、上海、广东三个板块中的 22 支股票进行投资组合的构建。通过东方财富网，提取所选股票池个股相关的信息，最终筛选 22 支股票的板块信息，如下表 3.8 所示。

表 3.8 股票池板块相关信息

股票代码	股票名称	所属板块
600000	浦发银行	上海板块；银行；
600028	中国石化	北京板块；石油行业；天然气；
600030	中信证券	广东板块；华为概念；证券；
600031	三一重工	北京板块；工程机械；工业互联；军工；新能源车；
600036	招商银行	广东板块；银行；
600048	保利发展	广东板块；房地产开发；养老概念；
600050	中国联通	北京板块；华为概念；区块链；
600104	上汽集团	上海板块；汽车整车；燃料电池；无人驾驶；新能源车；
600196	复星医药	上海板块；互联医疗；化学制药；中药概念；
600588	用友网络	北京板块；工业互联；互联医疗；华为概念；区块链；
600837	海通证券	上海板块；证券；
600900	长江动力	北京板块；电力行业；风能；核能核电；绿色电力；
601088	中国神华	北京板块；煤化工；煤炭行业；
601288	农业银行	北京板块；银行；
601318	平安银行	广东板块；保险；银行；互联医疗；区块链；养老概念；
601336	新华保险	北京板块；保险；养老概念；
601398	工商银行	北京板块；区块链；银行；
601601	中国太保	上海板块；保险；养老概念；
601628	中国人寿	北京板块；保险；
601668	中国建筑	北京板块；工程建设；
601857	中国石油	北京板块；石油行业；天然气；
601888	中国中免	北京板块；旅游酒店；免税概念；

数据来源：东方财富网

筛选好股票池所有股票相关的板块信息。定义股票相关联系网 G 图, $G = (V, E)$, 将每个板块视为一个节点, 其中 V 为板块节点的集合, $V = \{v_1, v_2, \dots, v_N\}$, N 是节点数, E 是边的集合。邻接矩阵 A 用于表示道路之间的连接, $A \in R^{N \times N}$ 。邻接矩阵只包含 0 和 1。如果板块有交集, 则元素为 1, 0 表示有无连接。定义 2: 特征矩阵 $X^{N \times P}$ 。紧接着将股票的收益率方差作为属性特征网络中的节点, 表示为 $X \in R^{N \times T}$, 其中 P 表示节点属性特征的个数 (长度的历史时间序列) 和 $X_t \in R^{N \times 1}$ 习惯于表示时间 i 时每支股票收益率的方差。所以依据 22 支股票板块的信息, 构建股票池的图, 既矩阵 $R^{22 \times 22}$, 板块有交集股票则元素为 1 否则为 0。

3.2.3.2 模型设计

GCN 模型在傅里叶域中构造一个 filter，filter 作用于图的节点及其 first-order 邻域捕获节点之间的空间特征，然后通过堆叠多个卷积层来构建 GCN 模型。假设节点 1 为中心股票，GCN 模型可以得到中心股票与其周边相关联系股票的拓扑关系，对股票网的拓扑结构和股票市场上的属性进行编码，进而得到空间依赖。总之，我们使用 GCN 模型从股票数据中学习空间特征。一个 2 层 GCN 模型可以表示为：

$$f(X, A) = \sigma(\hat{A} \text{Relu}(\hat{A} X W_0) W_1)$$

其中 X 表示特征矩阵， A 表示邻接矩阵， $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ 表示数据的预处理， $\tilde{A} = A + I_N$ 是具有自连接结构的矩阵， \tilde{D} 是度矩阵， $\tilde{D} = \sum_j \tilde{A}_{ij}$ 。 W_0 和 W_1 代表第一层和第二层的权重矩阵， $\sigma(\cdot)$ 和 $\text{Relu}()$ 表示激活函数。

本篇文章为了更好研究股票收益率方差，在图卷积神经网络后加入了门控循环单元，可以更好的捕捉时间依赖。完整图卷积网络单元如图 3.10。

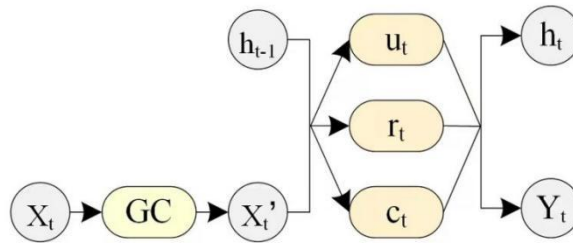


图 3.10 GCN 模型单元图

$$\begin{aligned} u_t &= \sigma(W_u[f(A, X_t), h_{t-1}] + b_u) \\ r_t &= \sigma(W_r[f(A, X_t), h_{t-1}] + b_r) \\ c_t &= \tanh(W_c[f(A, X_t), (r_t * h_{t-1})] + b_c) \\ h_t &= u_t * h_{t-1} + (1 - u_t) * c_t \end{aligned}$$

式中 h_{t-1} 表示 $t-1$ 时刻的输出，GC 是图卷积过程， u_t , r_t 是时间 t 的更新门和复位门， h_t 表示时间 t 的输出。具体计算过程如下图。 $f(A, X_t)$ 表示图卷积过程，定义在等式 2 中。 W 和 b 表示训练过程中的权重和偏差。

整个的股票收益率方差预测流程如图 3.11

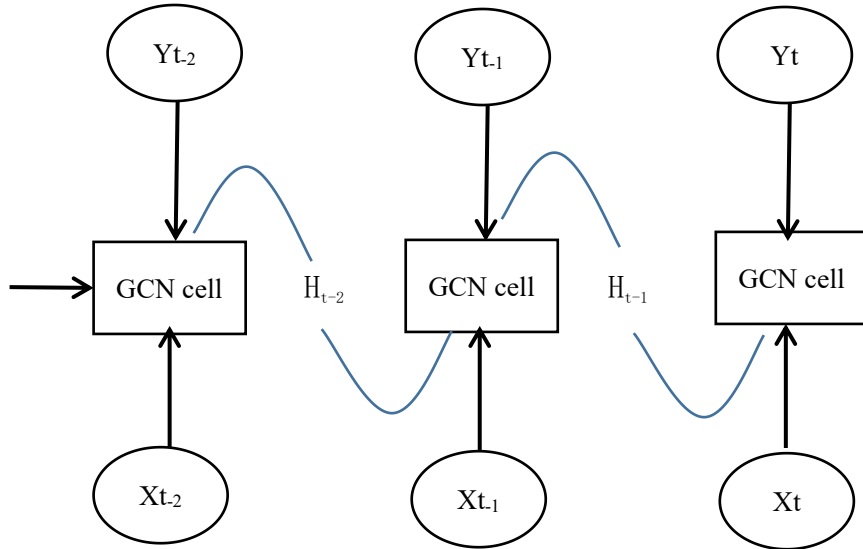


图 3.11 GCN 模型流程图

模型的超参数主要包括：学习率、batch size、训练 epoch、隐藏层数。在模型中，设置两个卷积层，用来提取特征，其次手动设置学习率为 0.001，batchsize=106, epoch=1000 次，隐藏单元设置为 100，。损失函数设置为均方误差 (mse) 与 lamda 倍的 L2 范数。这个设定可以避免参数的无限增长。公式如下：

$$\text{LOSS} = \text{MSE} + \lambda \frac{1}{2} L_2$$

3.2.3.3 模型预测

通过构造图卷积神经网络，同时设置好相关的参数，我们开始将整理好的数据放入到图卷积神经网络进行训练，从而预测未来的收益率方差。在下面讲分成三个市场，牛市、熊市、震荡区间分别进行收益率方差的预测。

(1) 牛市个股风险估计

接着通过对牛市的数据进行图神经网络建模。分别导出 2013 年 7 月 2 日的与 2015 年 3 月 31 日的数据，分别对 22 支个股进行建模预测，得到预测收益率

波动方差，进行保留 5 位小数处理，部分结果方差如下表所示：

表 3.9 股票池部分牛市预测收益率方差数据

	2015/4/1	2015/4/2	2015/4/3	2015/4/7	2015/4/8
浦发银行	1.0116011	0.8152768	0.872535	0.9693155	0.9823642
中国石化	0.70343566	0.8414222	0.85067934	0.7313978	0.8237315
中信证券	1.3890178	0.9964632	1.1183975	1.2182759	1.2395753
三一重工	0.7562981	0.86432934	0.8894632	0.8191281	0.8761205
招商银行	0.7581033	0.826618	0.8437043	0.526317	0.76816505

表 3.9 通过图卷积神经网络来对股票池中股票收益率方差进行预测，每一日的收益率方差的预测都是基于对应日期之前的历史数据。基于不断更新的已知数据，分别通过五次计算，计算出了未来五日的收益率方差。从表 4.3 中可以看到，股票池数据的收益率方差都在 1-3 之间，从部分数据中可以看到，其中招商银行的方差相对于较小，更平稳一些。

(2)熊市个股风险估计

接着通过对牛市的数据进行图神经网络建模。2015 年 6 月 13 日的与 2018 年 2 月 14 日的数据，分别对 22 支个股进行建模预测，得到预测收益率波动方差，进行保留 5 位小数处理，部分结果方差如下表所示：

表 3.10 股票池部分熊市预测收益率方差数据

	2018/2/22	2018/2/23	2018/2/26	2018/2/27	2018/2/28
浦发银行	1.764915	1.4756866	2.4316325	1.8065538	1.7039078
中国石化	1.339892	1.1496847	1.6525192	1.2409096	1.3782758
中信证券	2.217832	1.6627032	3.1599503	2.1959627	1.9895133
三一重工	1.373777	1.2412912	1.7507789	1.2995471	1.424789
招商银行	1.192901	1.0015525	1.571555	1.1915553	1.1864103

从表 3.10 可以看出，在未来的五个交易日中，整体的股票的收益方差都不打，大多数处于 1 附近，部分可以高达 3 左右。在部分展示中的数据，可以看到中信证券的预测得风险更高，数值更大。

(3) 震荡区间个股风险估计

表 4.10 显示, 收益率的总体正向负向分布均等, 可以看出整个市场价格呈现震荡的状态。通过对熊市的数据进行图卷积神经网络建模。分别导出 2018 年 12 月 31 日的与 2021 年 6 月 23 日的的数据, 分别对 20 支个股进行建模预测, 得到预测收益率波动方差, 进行保留五位小数处理, 最终如下表所示:

表 3.11 股票池部分震荡区间预测收益方差数据

	2021/6/24	2021/6/25	2021/6/28	2021/6/29	2021/6/30
浦发银行	1.764915	1.4756866	2.4316325	1.8065538	1.7039078
中国石化	1.339892	1.1496847	1.6525192	1.2409096	1.3782758
中信证券	2.217832	1.6627032	3.1599503	2.1959627	1.9895133
三一重工	1.373777	1.2412912	1.7507789	1.2995471	1.424789
招商银行	1.192901	1.0015525	1.571555	1.1915553	1.1864103

表 3.11 显示在未来的五个交易日中, 预测整体的方差还是在 1-2 之间, 中信证券会稍微高一些在 2-3 之间。从整体上看, 中信证券的收益率方差是稍高一些, 具有更大的风险。

3.3 最优投资组合的构建

在本小节在基于最优投资理论上, 应用模型预测的风险进行对最优投资组合的构建并且进行分析, 流程如下:

(1) 首先通过公式计算股票池中 22 支股票之间的相关系数, 并且汇总成为相关系数矩阵。分别依据研究的时间段, 计算牛市、熊市、震荡区间的股票池的相关系数矩阵。两支股票的相关系数计算公式如下:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{VAR[X]*VAR[Y]}}$$

股票池的相关系数矩阵是通过两两计算的出, 然后汇总成整个股票池的相关系数矩阵。计算好的相关系数矩阵展示于附录。

(2) 在计算股票池的相关系数矩阵之后, 引用模型预估的风险。通过预测, 在第 t 天的第 i 支收益率方差为 $\sigma_{t,i}(i=1,2,3,\dots,22)$, 其次初始赋予股票池中 22 支股票的投资权重总和为 1, 既 $\sum_i^{22} \omega_i = 1$ 。之后通过组合风险计算公式, 对第 t 天的组合

方差进行最小化求解。公式如下：

$$\left\{ \begin{array}{l} \min \quad \sigma_t^2 = \sum_{i=1}^{22} \sum_{j=1}^{22} \omega_{t,i} \omega_{t,j} \sigma_{t,i} \sigma_{t,j} \rho_{ij} \quad (i=1,2,3,\dots,22; j=1,2,3,\dots,22) \\ \text{s.t.} \quad \omega_{t,i} \geq 0 \\ \sum_{i=1}^{22} \omega_{t,i} = 1 \end{array} \right.$$

通过线性规划，对投资组合最小风险方案进行求解 $\omega_i(i=1,2,3,\dots,22)$ 。通过对 t-1 天的股票池中股票的收益率方差进行预测，然后计算第 t 天的组合收益率方差，通过最小化求解解得在 t 天，股票池中 22 支股票最优的投资权重分配 $\omega_i(i=1,2,3,\dots,22)$ 。

(3) 经过计算得到股票池中 22 支股票的最优投资权重，按照该权重进行的股票投资搭配即为模型构建的最优投资组合。确定在 t 天的最优投资组合权重，结合 t 日的个股的收益率，进行计算该组合在 t 日的组合收益率，计算公式如下：

$$R_t = \sum_1^i \omega_i * r_i \quad (i=1,2,3,\dots,22)$$

通过公式可以计算模型构建的最优投资组合的日收益率。通过分别汇总不同研究区间的五个交易日的最优投资组合的收益，进行加和处理，可以算出周收益率，即 $R_{\text{周}} = R_t + R_{t+1} + R_{t+2} + R_{t+3} + R_{t+4}$ 。之后对 $R_{\text{周}}$ 进行分析。

3.3. 1GARCH 构建最优投资组合

(1) 牛市投资组合

表 3.12 股票池牛市投资权重

	复星医药	长江动力	工商银行	中国石油
2015/4/1	46.34%	29.05%	0.00%	24.61%
2015/4/2	50.22%	47.35%	0.00%	2.43%

2015/4/3	56.35%	38.34%	0.00%	5.31%
2015/4/7	62.47%	0.00%	37.53%	0.00%
2015/4/8	28.44%	67.13%	0.00%	4.43%

表 3.12 为在牛市研究区间, 结合 GARCH(1, 1) 模型对 22 支股票预测的收益率方差, 使用组合方差计算公式对组合方差进行最小化计算, 求得 22 支股票在未来五个交易日的最优投资权重, 其中表中未出现的股票投资权重为 0%。在 4 月 1 日至 4 月 8 日, 通过组合方差最小化计算, 涉及到投资权重大于 0 的股票为复星医药、长江动力、工商银行、中国石油。因为其他股票投资权重增加组合会导致组合整体风险变高, 所以投资占比为 0%。

经过计算 GARCH(1, 1) 模型在 2015 年 4 月 1 日的投资组合为持仓 46.34% 复星医药、29.05% 长江动力、24.61% 中国石油。在 2015 年 4 月 2 日投资组合为持仓 50.22% 复星医药、47.35% 长江动力、2.43% 中国石油, 在 2015 年 4 月 3 日投资组合为持仓 56.35% 复星医药、38.34% 工商银行、5.31% 中国石油。在 2015 年 4 月 7 日的投资组合为持仓为 62.47% 复星医药、37.53% 工商银行。在 2015 年 4 月 8 日的投资组合为持仓 28.44% 复星医药、67.13% 的长江动力、4.43% 的中国石油。

(2) 熊市投资组合

表 3.13 股票池熊市投资权重

	浦发银行	用友网络	长江动力	华泰证券
2018/2/22	68.88%	1.50%	17.86%	11.75%
2018/2/23	66.94%	0.00%	17.68%	15.39%
2018/2/26	80.56%	0.01%	5.48%	13.94%
2018/2/27	89.12%	0.18%	0.00%	10.71%
2018/2/28	91.78%	0.37%	0.00%	7.85%

表 3.13 为在熊市研究区间, 结合 GARCH(1, 1) 模型对 22 支股票预测的收益率方差, 使用组合方差计算公式对组合方差进行最小化计算, 求得 22 支股票在未来五个交易日的最优投资权重, 其中表中未出现的股票投资权重为 0%。在 2018 年 2 月 22 日至 2 月 28 日, 通过组合方差最小化计算, 涉及到投资权重大于 0 的股票为浦发银行、用友网络、长江动力、华泰证券。因为其他股票投资权重增加组合会导致组合整体风险变高, 所以投资占比为 0%。

经过计算 GARCH(1, 1) 模型在 2018 年 2 月 22 日的投资组合为持仓 68.88% 浦发银行、1.50% 用友网络、17.86% 长江动力、11.75% 华泰证券。在 2018 年 2 月 23 日的投资组合持仓为 66.94% 浦发银行、17.68% 长江动力、15.39% 华泰证券。在 2018 年 2 月 26 日的投资组合持仓为 80.56% 浦发银行、0.01% 用友网络、10.71% 华泰证券。在 2018 年 2 月 27 日的投资组合持仓为 89.12% 浦发银行、0.18% 用友网络、10.71% 华泰证券。在 2018 年 2 月 28 日的投资组合持仓为 91.78% 浦发银行、0.37% 用友网络、7.85% 华泰证券。

(3) 振荡区间投资组合

表 3.14 股票池震荡区间投资权重

	保利发展	用友网络	长江动力
2021/6/24	6.90%	0.56%	0.00%
2021/6/25	0.99%	4.33%	6.31%
2021/6/28	1.34%	4.48%	3.84%
2021/6/29	0.31%	4.84%	3.86%
2021/6/30	0.31%	5.11%	2.08%
	农业银行	华泰证券	中国石油
2021/6/24	41.24%	27.34%	23.96%
2021/6/25	45.12%	34.19%	9.05%
2021/6/28	47.83%	32.26%	10.25%
2021/6/29	54.76%	27.96%	8.28%
2021/6/30	52.58%	29.68%	10.24%

表 3.14 为在震荡研究区间，结合 GARCH(1, 1) 模型对 22 支股票预测的收益率方差，使用组合方差计算公式对组合方差进行最小化计算，求得 22 支股票在未来五个交易日的最优投资权重，其中表中未出现的股票投资权重为 0%。在 2021 年 6 月 24 日至 6 月 30 日，通过组合方差最小化计算，涉及到投资权重大于 0 的股票为保利发展、用友网络、长江动力、农业银行、华泰证券、中国石油。其他股票因为加入组合会导致组合风险变高，因此投资占比为 0%。因为其他股票投资权重增加组合会导致组合整体风险变高，所以投资占比为 0%。

经过计算 GARCH(1, 1) 模型在 2021 年 6 月 24 日的投资组合为持仓 6.90% 保利发展、0.56% 用友网络、41.24% 农业银行、27.34% 华泰证券、23.96% 中国石油。在 2021 年 6 月 25 日的投资组合持仓为 0.99% 保利发展、4.33% 用友网络、6.31%

长江动力、45.12%农业银行、34.19%华泰证券、9.05%中国石油。在2021年6月28日的投资组合持仓为1.34%保利发展、4.48%用友网络、3.84%长江动力、47.83%农业银行、32.26%华泰证券、10.25%中国石油。在2021年6月29日的投资组合持仓为0.31%保利发展、4.84%用友网络、3.86%长江动力、54.76%农业银行、27.96%华泰证券、8.28%中国石油。在2021年6月30日的投资组合持仓为0.31%保利发展、5.11%用友网络、2.08%长江动力、52.58%农业银行、29.68%华泰证券、10.24%中国石油。

3.3. 2LSTM 构建最优投资组合

(1) 牛市投资组合

表 3.15 股票池牛市投资权重

	中国联通	中国神华	工商银行	中国石油
2015/4/1	3.13%	0.00%	0.00%	96.87%
2015/4/2	0.00%	0.00%	100.00%	0.00%
2015/4/3	0.00%	100.00%	0.00%	0.00%
2015/4/7	0.00%	0.00%	100.00%	0.00%
2015/4/8	0.00%	0.00%	0.00%	100.00%

表 3.15 为在牛市研究区间,结合 LSTM 模型对 22 支股票预测的收益率方差,使用组合方差计算公式对组合方差进行最小化计算,求得 22 支股票在未来五个交易日的最优投资权重,其中表中未出现的股票投资权重为 0%。在 2015 年 4 月 1 日至 4 月 8 日,通过组合方差最小化计算,涉及到投资权重大于 0 的股票为中国联通、中国神华、工商银行、中国石油。因为其他股票投资权重增加组合会导致组合整体风险变高,所以投资占比为 0%。

经过计算 LSTM 模型在 2015 年 4 月 1 日的投资组合为持仓 3.13%中国联通、96.87%中国石油。在 2015 年 4 月 2 日投资组合为持仓 100.00%工商银行,在 2015 年 4 月 3 日投资组合为持仓 100%中国神华。在 2015 年 4 月 7 日的投资组合为持仓为 100%的工商银行。在 2015 年 4 月 8 日的投资组合为持仓 100%中国石油。

(2)熊市投资组合

表 3.16 股票池熊市投资权重

	浦发银行	中国联通	复星医药	用友网络	海通证券
2018/2/22	77.30%	0.00%	0.45%	0.17%	0.34%
2018/2/23	0.00%	0.00%	0.00%	0.00%	0.15%
2018/2/26	0.00%	39.94%	0.00%	0.00%	1.12%
2018/2/27	0.08%	15.94%	0.00%	0.00%	15.88%
2018/2/28	0.10%	0.00%	0.01%	0.00%	85.63%
	长江动力	工商银行	中国太保	中国人寿	华泰证券
2018/2/22	0.00%	0.00%	16.79%	0.00%	4.96%
2018/2/23	0.00%	99.85%	0.00%	0.00%	0.00%
2018/2/26	0.00%	58.94%	0.00%	0.00%	0.00%
2018/2/27	0.00%	30.47%	0.00%	0.00%	37.64%
2018/2/28	0.00%	0.00%	0.00%	14.26%	0.00%

表 3.16 为在熊市研究区间, 结合 LSTM 模型对 22 支股票预测的收益率方差, 使用组合方差计算公式对组合方差进行最小化计算, 求得 22 支股票在未来五个交易日的最优投资权重, 其中表中未出现的股票投资权重为 0%。在 2018 年 2 月 22 日至 2 月 28 日, 通过组合方差最小化计算, 涉及到投资权重大于 0 的股票为浦发银行、中国联通、复星医药、用友网络、海通证券、长江动力、工商银行、中国太保、中国人寿、华泰证券。因为其他股票投资权重增加组合会导致组合整体风险变高, 所以投资占比为 0%。

经过计算 LSTM 模型在 2018 年 2 月 22 日的投资组合为持仓 77.30%浦发银行、0.45%复星医药、0.17%用友网络、0.34%海通证券、16.79%中国太保、4.96%华泰证券。在 2018 年 2 月 23 日的投资组合持仓为 0.15%海通证券、99.85%工商银行。在 2018 年 2 月 26 日的投资组合持仓为 39.94%中国联通、1.12%海通证券、58.94%工商银行。在 2018 年 2 月 27 日的投资组合持仓为 0.08%浦发银行、15.94%中国联通、15.88%海通证券、30.47%工商银行、37.64%华泰证券。在 2018 年 2 月 28 日的投资组合持仓为 0.10%浦发银行、0.01%复星医药、85.63%海通证券、14.26%中国人寿。

(3) 振荡区间投资组合

表 3.17 股票池震荡区间投资权重

	浦发银行	招商银行	中国联通
2021/6/24	0.00%	0.00%	100.00%
2021/6/25	0.00%	100.00%	0.00%
2021/6/28	0.00%	100.00%	0.00%
2021/6/29	0.00%	100.00%	0.00%
2021/6/30	35.87%	0.00%	64.13%

表 3.17 为在股市震荡研究区间，结合 LSTM 模型对 22 支股票预测的收益率方差，使用组合方差计算公式对组合方差进行最小化计算，求得 22 支股票在未来五个交易日的最优投资权重，其中表中未出现的股票投资权重为 0%。在 2021 年 6 月 24 日至 6 月 30 日，通过组合方差最小化计算，涉及到投资权重大于 0 的股票为浦发银行、招商银行、中国联通。因为其他股票投资权重增加组合会导致组合整体风险变高，所以投资占比为 0%。

经过计算 LSTM 模型在 2021 年 6 月 24 日的投资组合为持仓 100%中国联通。在 2021 年 6 月 25 日的投资组合持仓为 100%的招商银行。在 2021 年 6 月 28 日的投资组合持仓为 100%招商银行。在 2021 年 6 月 29 日的投资组合持仓为 100%招商银行。在 2021 年 6 月 30 日的投资组合持仓为 35.87%浦发银行、64.13%中国联通。

3.3. 3GCN 构建最优投资组合

(1) 牛市投资组合

表 3.18 股票池牛市投资权重

	三一重工	复星医药	中国人寿
2015/4/1	0.00%	0.00%	100.00%
2015/4/2	0.00%	24.33%	75.67%
2015/4/3	0.38%	23.14%	76.48%
2015/4/7	0.00%	0.00%	100.00%
2015/4/8	0.00%	10.45%	89.55%

表 3.18 为在股市牛市研究区间，结合 GCN 模型对 22 支股票预测的收益率方

差，使用组合方差计算公式对组合方差进行最小化计算，求得 22 支股票在未来五个交易日的最优投资权重，其中表中未出现的股票投资权重为 0%。在 2018 年 2 月 22 日至 2 月 28 日，通过组合方差最小化计算，涉及到投资权重大于 0 的股票为三一重工、复星医药、中国人寿。因为其他股票投资权重增加组合会导致组合整体风险变高，所以投资占比为 0%。

经过计算 GCN 模型在 2015 年 4 月 1 日的投资组合为持仓 100%中国人寿。在 2015 年 4 月 2 日投资组合为持仓 24.33%复星医药、75.67%中国人寿。在 2015 年 4 月 3 日投资组合为持仓 0.38%三一重工、23.14%复星医药、76.48%中国人寿。在 2015 年 4 月 7 日的投资组合为持仓 100%中国人寿。在 2015 年 4 月 8 日的投资组合为持仓 10.45%复星医药、89.55%中国人寿。

(2) 熊市投资组合

表 3.18 股票池熊市投资权重

	浦发银行	招商银行	中国联通		
2018/2/22	20.37%	27.05%	0.00%		
2018/2/23	21.07%	27.96%	0.06%		
2018/2/26	16.60%	18.47%	0.05%		
2018/2/27	16.69%	19.86%	0.05%		
2018/2/28	23.24%	29.65%	0.05%		
	用友网络	中国人寿	中国建筑	华泰证券	
2018/2/22	41.79%	0.00%	5.42%	5.37%	
2018/2/23	40.16%	0.00%	5.42%	5.32%	
2018/2/26	44.17%	10.50%	10.21%	0.00%	
2018/2/27	50.03%	0.00%	13.37%	0.00%	
2018/2/28	37.22%	0.00%	1.14%	8.71%	

表 3.19 为在熊市研究区间，结合 GCN 模型对 22 支股票预测的收益率方差，使用组合方差计算公式对组合方差进行最小化计算，求得 22 支股票在未来五个交易日的最优投资权重，其中表中未出现的股票投资权重为 0%。在 2018 年 2 月 22 日至 2 月 28 日，通过组合方差最小化计算，涉及到投资权重大于 0 的股票为浦发银行、招商银行、中国联通、用友网络、中国人寿、中国建筑、华泰证券。因为其他股票投资权重增加组合会导致组合整体风险变高，所以投资占比为 0%。

经过计算 GCN 模型在 2018 年 2 月 22 日的投资组合为持仓 20.37%浦发银行、

27.05%招商银行、41.79%用友网络、5.42%中国建筑、5.37%华泰证券。在2018年2月23日的投资组合持仓为21.07%浦发银行、27.96%招商银行、0.06%中国联通、40.16%用友网络、5.42%中国建筑。在2018年2月26日的投资组合持仓为16.60%浦发银行、18.47%招商银行、0.05%中国联通、44.17%用友网络、10.50%中国人寿、10.21%中国建筑。在2018年2月27日的投资组合持仓为16.69%浦发银行、19.86%招商银行、0.05%中国联通、37.22%用友网络、13.37%中国建筑。在2018年2月28日的投资组合持仓为23.24%浦发银行、29.65%招商银行、0.05%中国联通、37.22%用友网络、13.37%中国建筑。

(3) 振荡区间投资组合

表 3.20 股票池震荡区间投资权重

	中国石化	三一重工	招商银行	保利发展	中国联通
2021/6/24	1.72%	0.00%	10.31%	12.88%	0.06%
2021/6/25	0.00%	6.12%	2.99%	13.94%	0.07%
2021/6/28	0.00%	5.14%	4.82%	15.07%	0.07%
2021/6/29	0.00%	5.14%	4.82%	15.09%	0.07%
2021/6/30	0.00%	7.39%	2.17%	14.19%	0.06%
	用友网络	长江动力	农业银行	中国建筑	华泰证券
2021/6/24	27.29%	2.42%	0.60%	1.21%	43.51%
2021/6/25	23.96%	5.43%	0.04%	0.00%	47.46%
2021/6/28	23.27%	4.51%	0.04%	0.00%	47.09%
2021/6/29	23.30%	4.51%	0.04%	0.00%	47.03%
2021/6/30	23.18%	5.47%	0.06%	0.00%	47.54%

表 3.20 为在股市震荡研究区间，结合 GCN 模型对 22 支股票预测的收益率方差，使用组合方差计算公式对组合方差进行最小化计算，求得 22 支股票在未来五个交易日的最优投资权重，其中表中未出现的股票投资权重为 0%。在 2021 年 6 月 24 日至 6 月 30 日，通过组合方差最小化计算，涉及到投资权重大于 0 的股票为中国石化、三一重工、招商银行、保利发展、中国联通、用友网络、长江动力、农业银行。因为其他股票投资权重增加组合会导致组合整体风险变高，所以投资占比为 0%。

经过计算 GCN 模型在 2021 年 6 月 24 日的投资组合为持仓 1.72%中国石化、10.31%招商银行、12.88%保利发展、0.06%中国联通、27.29%用友网络、2.42%

长江动力、0.60%农业银行、1.21%中国建筑、47.54%华泰证券。在2021年6月25日的投资组合持仓为6.12%三一重工、2.99%招商银行、13.94%保利发展、0.07%中国联通、23.96%用友网络、5.43%长江动力、0.04%农业银行、47.46%华泰证券。在2021年6月28日的投资组合持仓为5.14%三一重工、4.82%招商银行、15.07%保利发展、0.07%中国联通、23.27%用友网络、4.51%长江动力、0.04%农业银行、47.09%华泰证券。在2021年6月29日的投资组合持仓为5.14%三一重工、4.82%招商银行、15.09%保利发展、0.07%中国联通、23.30%用友网络、4.51%长江动力、0.04%农业银行、47.03%华泰证券。在2021年6月30日的投资组合持仓为7.3928%三一重工、2.17%招商银行、14.19%保利发展、0.06%中国联通、23.18%用友网络、5.47%长江动力、0.06%农业银行、47.54%华泰证券。

3.4 最优投资组合收益的计算

在第三章的时候，分别介绍了通过第 t 天的历史数据，对 $t+1$ 天股票池个股的风险进行预测，通过计算求出 $t+1$ 天的最优投资组合 22 支股票的投资权重 $\omega_{t+1, i}$ ($i=1,2,3,\dots,22$)。投资组合的收益率则是通过最优投资组合的股票持仓权重乘以未来一日的个股收益率，在进行求和，如公式：

$$R_{t+1} = \sum_{i=1}^{22} \omega_i * R_i (i = 1,2,3,\dots,22)$$

通过公式可以计算 GCN 模型、GARCH 模型、LSTM 模型在不同环境的股票市场构建的最优投资组合日收益率如表 3.21 所示。通过分别汇总不同研究区间的五个交易日的最优投资组合的收益，进行加和处理，可以算出周收益率，即

$R_{\text{周}} = R_t + R_{t+1} + R_{t+2} + R_{t+3} + R_{t+4}$ 。之后对 $R_{\text{周}}$ 进行分析。

表 3.21 GCN 模型、GARCH 模型、LSTM 模型构建投资组合的日收益

	GCN 模型	GARCH 模型	LSTM 模型
2015/4/1	0.88%	0.00%	0.13%
2015/4/2	-0.21%	-0.37%	-0.37%
2015/4/3	0.05%	0.22%	-0.24%
2015/4/7	1.66%	0.51%	1.23%

2015/4/8	1.64%	0.27%	0.00%
2018/2/22	1.98%	0.60%	0.77%
2018/2/23	0.13%	0.30%	0.00%
2018/2/26	0.59%	0.11%	0.63%
2018/2/27	-0.45%	-0.18%	-0.84%
2018/2/28	0.74%	-0.77%	-0.43%
2021/6/24	0.47%	0.41%	0.08%
2021/6/25	0.20%	0.10%	0.30%
2021/6/28	0.41%	0.18%	0.40%
2021/6/29	-0.26%	-0.14%	-0.60%
2021/6/30	-0.27%	-0.30%	0.25%

通过表 3.21，可以看到三种模型构建的最优投资组合的收益率都在-0.5%至 2%左右，说明通过投资组合获取的整体收益是不高的。从整体数据分布看，负值收益率的数量是非常低的，可以看出，三个模型建立的最优投资组合在大多数交易日都有正向收益。

表 3.22 GCN 模型、GARCH 模型、LSTM 模型构建投资组合的周收益

	GCN 模型	GARCH 模型	LSTM 模型
牛市研究时间段	4.01%	0.63%	0.75%
熊市研究时间段	2.98%	0.05%	0.13%
震荡区间研究时间段	0.55%	0.24%	0.44%

表 3.22 为三个模型在研究时间段的周收益率。通过对研究时间段的五个交易日的组合的日收益率进行加总，求得组合的周收益率。从表 4.2 中可以看到，GCN 模型构建的投资组合的收益情况，明显高于另外两个模型。

4 三种模型构建组合对比分析

本篇文章分别通过对股市中三个时间段，牛市熊市震荡区间进行划分研究。牛市研究选取的时间为 2013 年 7 月 1 日至 2015 年 4 月 8 日，熊市研究选取的时间为 2015 年 6 月 13 日至 2018 年 2 月 28 日，震荡区间研究选取的时间为 2019 年 1 月 2 日至 2021 年 6 月 30 日，分别使用 GARCH 模型、LSTM 模型、图卷积神经网络对股票的收益率方差进行预测，通过最小化 22 支股票投资组合的方差来确定对 22 支股票投资的权重，构建最优投资组合。最后根据最优投资组合的个股投资权重以及当日个股的收益率计算求和得到组合的当日收益率，进行对比研究。

4.1 组合结构对比分析

GCN 模型、GARCH 模型、LSTM 模型在不同研究时间段构建的最优投资组合都有差异。

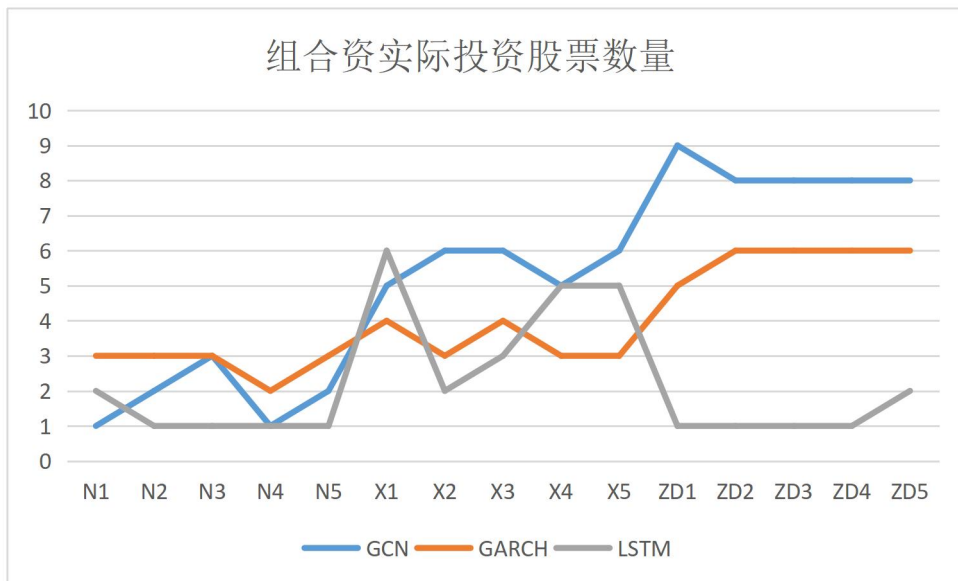


图 4.1 GCN 模型、GARCH 模型、LSTM 模型构建投资组合实际投资的股票数量图

图 4.1 为 GCN 模型、GARCH 模型、LSTM 模型分别在牛市熊市以及震荡区间研究时间段对股票实际投资个数。其中图中 N_i ($i=1, 2, 3, 4, 5$) 为牛市的第 i 个交易日， X_i ($i=1, 2, 3, 4, 5$) 为熊市的第 i 个交易日， ZD_i ($i=1, 2, 3, 4, 5$) 为震荡区间的第

i 个交易日。从整体上看，GCN 模型构建的最优投资组合的个股数量大于另外两个模型构建。LSTM 模型构建的最优投资组合的个股数量大多数为 1，在牛牛市研究时间段的后四个交易日投资股票的个数为 1，在震荡区间研究时间段前四个交易日投资股票数量为 1。GARCH 模型构建的最优投资组合在牛市研究时间段，大于另外两个模型投资的股票数量，组合涉及个股的数量在 2-3 个，而在熊市以及震荡区间研究时间段，投资个股的数量居中。从整体分析，GCN 构建的投资组合涵盖个股的数量最多，投资股票更加分散。如果按照 Markowitz 曾经的结论，投资更多的股票可以更好的降低风险，那么可以看出 GCN 模型构建的投资组合涵盖的股票数量多，风险更低。

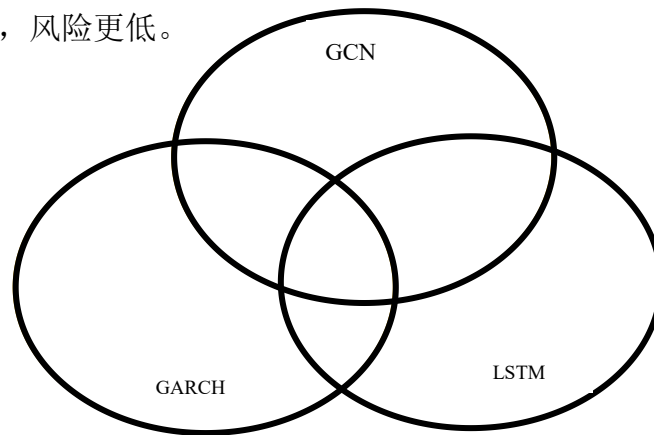


图 4.2 GCN 模型、GARCH 模型、LSTM 模型构建投资组合涵盖股票相关示意图

如图 4.2，对 GCN 模型、GARCH 模型、LSTM 模型构建最优投资组合涉略股票进行对比分析示意图。分别对牛市、熊市、震荡区间三个研究时间段，将三个模型构建的投资组合成分股进行对比分析。

在牛市研究时间段，GCN 模型构建最优投资组合涉及的股票有三一重工、复星医药、中国人寿。GARCH 模型构建最优投资组合涉及的股票有复星医药、长江动力、工商银行、中国石油。LSTM 模型构建最优投资组合涉及的股票有中国联通、中国神华、工商银行、中国石油。在牛市，GCN 模型与 GARCH 模型构建的组合都投资了复星医药，GARCH 模型与 LSTM 模型构建的最优投资组合同时投资了工商银行、中国石油。综合看，在牛市的时候，通过三个模型构建最优投资组合对比分析，可以观测复星医药、工商银行、中国石油是共同认为是值得投资的，用来构建投资组合可以降低风险。

在熊市研究时间段, GCN 模型构建最优投资组合涉及的股票有浦发银行、招商银行、中国联通、用友网络、中国人寿、中国建筑、华泰证券。GARCH 模型构建最优投资组合涉及的股票有浦发银行、用友网络、长江动力、华泰证券。LSTM 模型构建最优投资组合涉及的股票有浦发银行、中国联通、复星医药、用友网络、海通证券、长江动力、工商银行、中国太保、中国人寿、华泰证券。从覆盖股票来看, GCN 与 GARCH 都选取了浦发银行、用友网络、华泰证券。GCN 与 LSTM 模型都选取了浦发银行、中国联通、用友网络、华泰证券。LSTM 与 GARCH 模型都选取了浦发银行、用友网络、长江动力。通过分析, 可以看到在熊市的时候, 三个模型的选择都涵盖了浦发银行、用友网络、华泰证券, 说明这三支股票在该研究时间段是具有投资价值的, 可以用来构建投资组合可以降低风险。

在震荡区间研究时间段, GCN 模型构建最优投资组合涉及的股票有中国石化、三一重工、招商银行、保利发展、中国联通、用友网络、长江动力、农业银行、中国建筑、华泰证券。GARCH 模型构建最优投资组合涉及的股票有保利发展、用友网络、长江动力、农业银行、华泰证券、中国石油。LSTM 模型构建最优投资组合涉及的股票有浦发银行、招商银行、中国联通。在该研究时间段从覆盖股票来看, GCN 与 GARCH 模型都选取了保利发展、用友网络、长江动力、农业银行、华泰证券。GCN 与 LSTM 模型都选取了招商银行、中国联通。LSTM 与 GARCH 模型无交集。通过分析, 可以看到在震荡区间该研究时间段, GCN 模型覆盖股票的数量比较多, 保利发展、用友网络、长江动力、农业银行、华泰证券、招商银行、中国联通具有投资价值, 可以用来构建投资组合可以降低风险。

总体分析, GCN 模型构建的投资组合涵盖股票的数量相对于其他的模型要多, LSTM 模型构建的投资组合一般情况下最少。同时, 在不同时间段, 三者大多数都会有共同的预期风险低的股票, 该类股票值得关注, 可以用来构建投资组合降低风险。

4.2 组合收益对比分析

4.2.1 牛市收益率对比分析

应用 GARCH 模型、LSTM 模型、GCN 模型对牛市研究区间进行预测，进行投资组合构建，日收益率如下图 4.3。

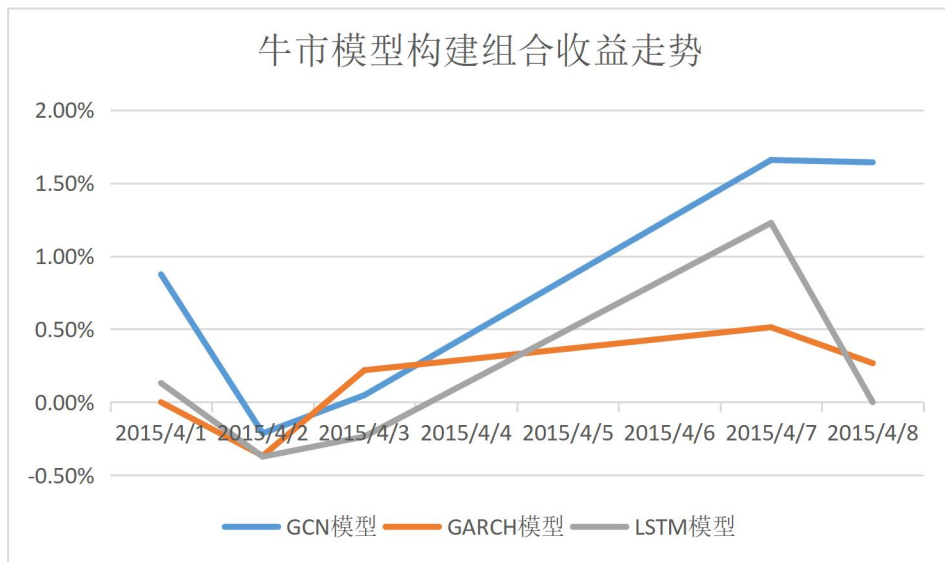


图 4.3 GCN 模型、GARCH 模型、LSTM 模型构建投资组合收益率走势图

图 4.3 展示了三个模型构造的投资组合的收益率。用图卷积神经网络构建的投资组合，在未来五个交易日 2015 年 4 月 1 到 4 月 8 日的收益率分别为 0.88%，-0.21%，0.05%，1.66%，1.64%，整个一周的收益率为 4.00%。用 GARCH 构造的投资组合，在未来五个交易日的收益率分别为 0%，-0.37%，0.22%，0.51%，0.27%，整个一周的收益率为 0.63%。用 LSTM 模型的最小方差投资组合，可以看到未来五个交易日的收益率分别为 0.13%，-0.37%，-0.24%，1.23%，0%，整个一周的收益率为 0.75%。从三个模型构建的投资组合看，三个模型在 4 月 2 日都有小幅的下跌，其中 GCN 模型构建的组合相对于跌幅最少，抗风险强。其次可以看到，LSTM 模型构建的投资组合更敏感，波动更大，GCN 模型构建的投资组合相对于稳定。从整体上看，图卷积神经网络构造的最优投资组合的收益好于 GARCH 模型、LSTM 模型。

4.2.2 熊市收益率对比分析

应用 GARCH 模型、LSTM 模型、GCN 模型对熊市研究区间进行预测，进行投资组合构建，收益率走势图如下 4.4。

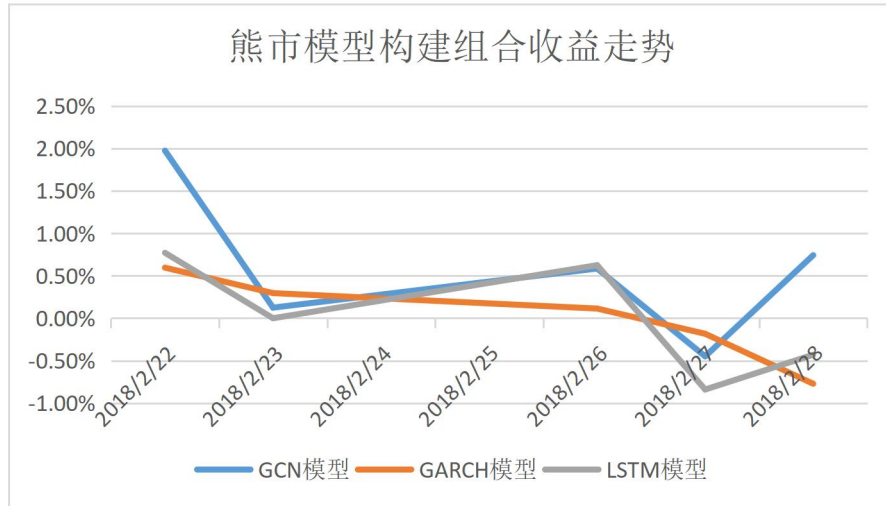


图 4.4 GCN 模型、GARCH 模型、LSTM 模型构建投资组合收益率走势图

图 4.4 展示了三个模型构造的投资组合的收益率。用图卷积神经网络构建的投资组合，在未来五个交易日 2018 年 2 月 22 到 2 月 28 日的收益率分别为 1.98%，0.13%，0.59%，-0.45%，0.74%，整个一周的收益率为 3.00%。用 GARCH 构造的投资组合，在未来五个交易日的收益率分别为 0.6%，0.3%，0.11%，-0.18%，-0.77%，整个一周的收益率为 0.05%。用 LSTM 模型的最小方差投资组合，可以看到未来五个交易日的收益率分别为 0.77%，0%，0.63%，-0.84%，-0.43%，整个一周的收益率为 0.13%。整个市场处于熊市，股市整体价格的方向是向下的。而三个模型通过预测最小方差构造的最优投资组合，在下行的市场上还可以取得正向的收益，可以看出三个方法构建的最优投资组合具有很强的抵抗风险能力。从三者的收益对比可以看到，GARCH 模型构建的投资组合的收益波动幅度是三者最小，整个走势曲线是相对于最平稳的，LSTM 模型则是波动最大的。从整体看，虽然 GCN 模型波动是趋于中间的，波动幅度居中，但 GCN 模型构建的最优投资组合收益最高。

4.2.3 震荡区间收益率对比分析

应用 GARCH 模型、LSTM 模型、GCN 模型对震荡区间研究区间进行预测，进行投资组合构建，收益率走势图如下。

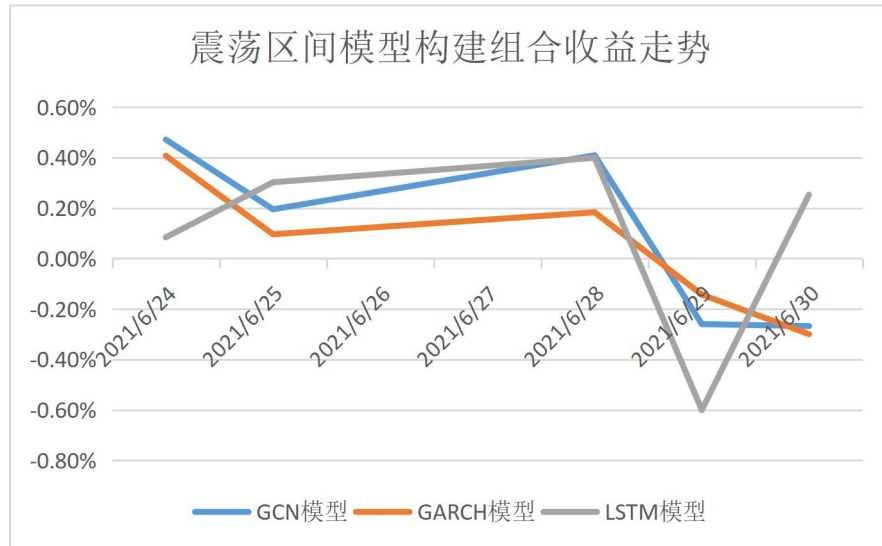


图 4.5 GCN 模型、GARCH 模型、LSTM 模型构建投资组合收益率走势图

图 4.5 展示了三个模型构造的投资组合的收益率。用图卷积神经网络构建的投资组合，在未来五个交易日 2021 年 6 月 24 到 6 月 30 日的收益率分别为 0.47%，0.2%，0.41%，-0.26%，-0.27%，整个一周的收益率为 0.55%。用 GARCH 构造的投资组合，在未来五个交易日的收益率分别为 0.41%，0.1%，0.18%，-0.14%，-0.3%，整个一周的收益率为 0.24%。用 LSTM 模型的最小方差投资组合，可以看到未来五个交易日的收益率分别为 0.08%，0.3%，0.4%，-0.6%，0.25%，整个一周的收益率为 0.44%。整个市场处于震荡区间，股市整体价格的方向不明朗。而三个模型通过预测最小方差构造的最优投资组合，在不明朗的股票市场上还可以取得正向的收益，可以看出通过三个方法构建的最优投资组合具有很好的收益能力。从三者的收益对比可以看到，GARCH 模型构建的投资组合的收益波动幅度是三者最小，GCN 整个走势曲线是相对于最平稳的，LSTM 模型则是波动最大的。从整体看，GCN 模型波动是趋于中间的，收益率最高。

5 总结与展望

5.1 总结

随着投资需求的增加以及计算机算法的发展,机器学习应用在金融市场中越来越广泛。在前辈们丰硕的理论和实证的研究成果上,本文介绍了 garch 模型、LSTM 模型、图卷积神经网络模型构造最小方差最优投资组合,重点对比了这三方法的优缺点。首先是通过介绍 GARCH 模型、LSTM 模型、GCN 模型的概念进行分析,说了三种模型在金融领域的应用。在此基础上使用这三个模型分别对牛市、熊市、震荡区间三个区间建模,对未来的收益率方差进行个预测。在预测结果之后,对所构造的股票池中进行最小方差计算,得到每只股票相应的投资权重,构造方差最小最优投资组合。通过实证,可以初步得到一下结论:

5.1.1 收益方面

基于最优投资理论,分别使用 GARCH 模型、LSTM 模型、GCN 模型对股市牛市、熊市和震荡区间进行建模预测构建最优投资组合。在整体上来看,三个模型构建的最优投资组合都可以取得相对不错的收益。在牛市的时候三者都具有可观的收益,在熊市以及震荡区间也同样都有正向的收益。尤其在熊市的时候,GCN 模型在五个交易日内取得了接近于 3%的收益。在三个模型收益进行对比分析,可以看到 GCN 模型无论是在牛市、熊市、震荡区间的收益都是相对于 GARCH 模型、LSTM 模型更好的,都具有更高的收益。

5.1.2 抗风险能力方面

在本篇文章分别选取了牛市、熊市、震荡区间三个时间段进行研究。在不同的时间段内,通过 GARCH 模型、LSTM 模型、GCN 模型构建的投资组合的整体波动都是相对较小的,与同期的大盘相比,都是小于大盘的波动的。可以明显的看出最优投资组合可以很好的规避市场的风险,可以避免股市下行的时候有重大经济损失以及在股市上行的时候具有稳定的收益。同时三个模型构建的组合收益率对比,GARCH 模型收益的走势是更趋于稳定的,波动幅度更小。LSTM 模型收益的走势则是大起大落,波动幅度很大。GCN 模型收益走势波动则是介于两者之间。从大多数的交易日看,GCN 模型相对于 GARCH 模型、LSTM 模型的投资组合向下波动

更小一些。说明 GCN 模型构建的投资组合具有更好的抗风险能力。

5.1.3 股市波动方面

通过对文献的梳理，本篇文章是分成了三个时间段进行研究。同时在这三个时间段内，三个模型投建的最优投资组合收益是有很大差异的。在牛市的时候，三个模型构建的投资组合都取得不错的正向收益。在熊市上升区间也是具有正向的收益。而在不明朗的震荡区间，GCN 模型、LSTM 模型、GARCH 模型构建的最优投资组合都取得比较小的正向收益。可以看到市场的变化会极大的影响组合的收益波动。但是总体上来看，通过最优投资组合还是可以具有很好的收益。

综上，实证研究表明投资者的通过 GARCH 模型、LSTM 模型、GCN 模型构造的最优投资组合会不同程度的受到所处市场阶段影响。同时还可以看到在不同的市场中，GARCH 模型、LSTM 模型、GCN 模型构造的最优投资组合，都是可以很好的抵抗风险，获取收益。而在三个模型中，GCN 模型构造的最优投资组合是更优的，收益高波动小。

5.2 展望

证券市场的震荡会影响国民经济，而且极有可能会产生巨大的经济损失。分析金融市场的稳定问题，宏观上主要受到国家的政策与方针影响，微观上看投资者的行为才算风险最根本的源头，因此投资者如何构建最优投资组合最关键。本文分别从三个方法来对风险最小化最优投资组合进行研究，在一定程度上可以帮助投资者降低投资风险，增强股票投资的抗风险能力。不过限于笔者的水平，本文仍存在一些学术或实践上的不足的问题需要进一步地改进与解决。

(1)在图卷积神经网络构图中，添加的只有行业、地理板块信息，其他的股票之间的联系并没有完全融入到构图中，这个如何添加更多的图元素，仍值得更深一步探索。

(2)本文研究的时间段分成牛市熊市以及股市的震荡区间，这个都是基于先验条件。在已知的数据中，随机筛选各个时期的后期的一段时间。并没有考虑到月底月初等因素，这个时间因素还是值得探讨。

(3)本文的 GCN 构建的最优投资组合并未设置过多的限制条件。在现实生活中，股票上有几千只股票，投资者们进行构造投资组合可以从更大的股票池中筛

选，而本品文章则是优中选优的原则，所以后期更多的对股票进行研究也是值得讨论的。

参考文献

- [1]Varfis A , Versino C . Univariate economic time series forecasting by connectionist methods[C]// 1990.35(1).
- [2]Hammad A A A, Ali S M A, Hall E L. Forecasting the jordanian stock prices using artificial neural networks[M]// Intelligent engineering systems through artificial neural networks.[S.l.]: ASME Press, 2007: 502-505.
- [3]Bertail P, Cléménçon S, Tressou J. Extreme Values Statistics for Markov Chains via the (Pseudo-) Regenerative Method[J]. Extremes, 2009, 12(4) : 327.
- [4]张伟. 基于 GA 优化的 SVM 算法的股票趋势预测[D]. 长春: 吉林大学, 2010.
- [5]李巍. 基于宏观经济指标和人工智能方法的上证综合指数预测[D]. 成都: 西南财经大学, 2012.
- [6]宋敏晶. 基于情感分析的股票预测模型研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [7]张希影. 基于遗传算法优化的 BP 神经网络股票价格预测[D]. 青岛: 青岛科技大学, 2014.
- [8]高天. 基于最优小波包变换、ARIMA 与 SVR 的股票价格预测研究[J]. 贵州财经大学学报, 2015(06): 57-69.
- [9]张贵勇. 改进的卷积神经网络在金融预测中的应用研究[D]. 郑州: 郑州大学, 2016.
- [10]吕涛, 郝泳涛. 基于 K 线序列相似性搜索的股票价格预测[J]. 计算机应用, 2017, 37(S2): 229-235.
- [11]余传明, 龚雨田, 王峰, 安璐. 基于文本价格融合模型的股票趋势预测[J]. 数据分析与知识发现, 2018, 2(12): 33-42.
- [12]冉杨帆, 蒋洪迅. 基于 BPNN 和 SVR 的股票价格预测研究[J]. 山西大学学报(自

- 然科学版),2018,41(01):1-14.DOI:10.13451/j.cnki.shanxi.univ(nat.sci.).2018.01.001.
- [13]宋晨光.基于长短期记忆神经网络模型的股票价格走势预测[D].南昌:江西财经大学,2019.
- [14]谢琪,程耕国,徐旭.基于神经网络集成学习股票预测模型的研究[J].计算机工程与应用,2019,55(08):238-243.
- [15]宋刚,张云峰,包芳勋,秦超.基于粒子群优化 LSTM 的股票预测模型[J].北京航空航天大学学报,2019,45(12):2533-2542.
- [16]王禹,陈德运,唐远新.基于 Cart 决策树与 boosting 方法的股票预测[J].哈尔滨理工大学学报,2019,24(06):98-103.
- [17]罗鑫,张金林.基于多时间尺度复合深度神经网络的股票价格预测[J].武汉金融,2020(09):32-40.
- [18]肖雯.基于机器学习的中国股票市场收益率预测研究[D].南京:南京大学,2020.
- [19]骆桦,秦艳艳.中国股市动量与反转效应模型的研究[J].浙江理工大学学报,2011,28(04):643-646.
- [20]陈健,胡文伟,李湛.不同策略条件下的投资组合平均风险比较与分散[J].上海交通大学学报,2011,45(12):1824-1828+1835.DOI:10.16183/j.cnki.jsjtu.2011.12.016.
- [21]林德发,杨潇宇.跑赢沪深 300 指数的成分股组合构建——基于多因素模型的实证分析[J].中国商贸,2014(02):83-84.
- [22]柯原,郑双阳.价值投资与行业轮动相结合的量化择股策略研究[J].福建金融管理干部学院学报,2014(01):3-10.
- [23]巨红岩,李俊林,董安强.股票资金流强度的实证研究及评测[J].科技和产业,2015,15(11):133-136.
- [24]黄宏运,王梅,朱家明.基于多元回归分析的多因子选股模型[J].通化师范学院学报,2016,37(08):44-46.

- [25]吴瑜琪.因子分析法在上市公司投资价值分析的实用性探究——基于 TMT 板块[J].全国流通经济,2017(15):74-75.
- [26]Zhong X, Enke D. Forecasting daily stock market return using dimensionality reduction[J]. Expert Systems with Applications, 2017, 67: 126-139.
- [27]刘述忠.基于 GM-RBF 神经网络的股票价格预测分析[J].计算机与现代化, 2018(08): 8-11.
- [28]Hu H, Tang L, Zhang S, et al. Predicting the direction of stock markets using optimized neural networks with Google Trends[J]. Neurocomputing, 2018, 285: 188-195.
- [29]周亮.基于分位数回归的多因子选股策略研究[J].西南大学学报(自然科学版),2019,41(01):89-96.
- [30]王慎敏.随机森林模型在选股策略中的应用[J].科技经济导刊,2020,28(33):188-190.
- [31]徐冰冰,岑科廷,黄俊杰,沈华伟,程学旗.图卷积神经网络综述[J].计算机学报, 2020, 43 (05) :755-780.
- [32] Gori M , Monfardini G , F Scarselli. A new model for learning in graph domains[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.
- [33]Kipf T N , Welling M . Semi-Supervised Classification with Graph Convolutional Networks[J]. 2016.
- [34]Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs[J]. Advances in neural information processing systems, 2017, 30.
- [35]Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
- [36]Xu D, Zhu Y, Choy C B, et al. Scene graph generation by iterative message passing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5410-5419.

- [37] Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Thirty-second AAAI conference on artificial intelligence. 2018.
- [38] 李志帅, 吕宜生, 熊刚. 基于图卷积神经网络和注意力机制的短时交通流量预测[J]. 交通工程, 2019, 19(04): 15-19+28. DOI:10.13986/j.cnki.jote.2019.04.003.
- [39] 闫旭, 范晓亮, 郑传潘, 臧彧, 王程, 程明, 陈龙彪. 基于图卷积神经网络的城市交通态势预测算法[J]. 浙江大学学报(工学版), 2020, 54(06): 1147-1155.
- [40] 王天保, 刘昱, 郭继昌, 晋玮佩. 图卷积神经网络行人轨迹预测算法[J]. 哈尔滨工业大学学报, 2021, 53(02): 53-60.
- [41] 李小妍. 基于图神经网络的交通流量预测[D]. 电子科技大学, 2020. DOI:10.27005/d.cnki.gdzku.2020.001107.
- [42] 刘腾娇. 多目标投资组合均值方差模型的改进及应用研究[D]. 哈尔滨工业大学, 2016.

致谢

在兰财的三年硕士学习生活转瞬即逝。回顾在兰财的三年的校园生活，有难过、有开心、有辛酸也有感动。在此，我要向所有在硕士生涯帮助过我的人，发自内心的对他们说一声感谢！

首先感谢我的导师，韩老师。韩老师是一位学识渊博、学风严谨、对待工作一丝不苟的老师。还记得当第一次见韩老师，还是考研复试的时候。但是他作为我们专业复试的主考官的，穿着正装，非常帅气，给我们所有的考生都留下了非常深刻的印象。兰财选导师是学生与导师双选，非常感谢韩老师给个机会让我可以加入韩门学习。在校的生活，他时常督促我们要以学业为重，抓紧学习课程，同时多涉猎机器学习领域，说改领域前景非常好。同时他还经常带着师门的学生参加比赛，斩获诸多奖项，比如全国 SAS 大赛三等奖、全统计建模三等奖、全国高校计算机能力大赛挑战赛甘肃赛区竞赛省奖等等。同时在毕业论文的撰写的时候，老师也对我的指导非常的多。

其次感谢我师兄师姐师弟师妹们，他们不仅在平日的学科研道路中给予了无限帮助，同时在平日日常生活中也给了我很多的关怀。还有统计学院的老师们，感谢他们传授的高质量课程，让我在专业知识上有了进一步的提升。还有要感谢我的父母，感谢他们一直以来的付出、对我的支持与关爱。是他们给予了我强大的后盾，让我可以舒心的完成三年的硕士学业任务。还要感谢 19 级统计学院的班委们，在我研究生当班长的这三年时间、感谢各位班委们的大力支持与拥护，才能让我可以更好的为统计学院服务，尤其要感谢女班长，协助我分担了很多院内的的工作。最后感谢 19 级统计学院男生，他们让我在研究生期间的业余生活更加丰富，比如一起校外健身、一起打羽毛球、一起打篮球、一起研究机器学习算法等。

最后，还要感谢那些对我的论文评审的老师们，因为他们的修改意见，让我认识到论文的不足，可以从更多的方面去改进。再次向各位老师表达真诚的感谢！

附录

附录一：牛市股票池的相关系数矩阵

中国石油	0.919452	0.806831	0.904814	0.82147	0.891702	0.855241	0.83535	0.892704	0.659762
华泰证券	0.960911	0.888333	0.966512	0.858097	0.970537	0.957177	0.911679	0.916271	0.636925
中国建筑	0.965066	0.889333	0.969693	0.894877	0.98035	0.95403	0.954298	0.921167	0.640644
中国人寿	0.959329	0.861685	0.969157	0.826592	0.954135	0.924347	0.865146	0.924287	0.662981
中国太保	0.968203	0.883292	0.969516	0.82084	0.962416	0.922353	0.875143	0.955566	0.660511
工商银行	0.941653	0.835495	0.944839	0.909868	0.977453	0.979677	0.903652	0.903083	0.497185
新华保险	0.958392	0.848268	0.968548	0.842414	0.960771	0.927041	0.857204	0.935721	0.627254
平安银行	0.98267	0.922356	0.981497	0.816712	0.970791	0.912815	0.902927	0.949254	0.732808
农业银行	0.9782	0.905368	0.981041	0.862035	0.985448	0.949246	0.899897	0.947596	0.635164
中国神华	0.91203	0.795729	0.917776	0.929347	0.954169	0.962452	0.881061	0.872197	0.44057
长江动力	0.940744	0.883429	0.942253	0.838529	0.961841	0.923365	0.898178	0.962304	0.588991
海通证券	0.962381	0.832579	0.9821	0.926918	0.977109	0.955799	0.904452	0.894131	0.593851
用友网络	0.930201	0.885063	0.91948	0.722788	0.894274	0.845457	0.867927	0.922699	0.775697
复星医药	0.70791	0.805043	0.664596	0.35723	0.59839	0.454201	0.648188	0.716167	1
上汽集团	0.950502	0.931805	0.936582	0.739044	0.930853	0.857069	0.871212	1	0.716167
中国联通	0.906717	0.883137	0.882312	0.820721	0.91611	0.902021	1	0.871212	0.648188
保利发展	0.916898	0.788793	0.924865	0.924733	0.959651	1	0.902021	0.857069	0.454201
招商银行	0.97034	0.877738	0.977844	0.892939	1	0.959651	0.91611	0.930853	0.59839
三一重工	0.844062	0.64543	0.86459	1	0.892939	0.924733	0.820721	0.739044	0.35723
中信证券	0.975159	0.885453	1	0.86459	0.977844	0.924865	0.882312	0.936582	0.664596
中国石化	0.912719	1	0.885453	0.64543	0.877738	0.788793	0.883137	0.931805	0.805043
浦发银行	1	0.912719	0.975159	0.844062	0.97034	0.916898	0.906717	0.950502	0.70791
	浦发银行	中国石化	中信证券	三一重工	招商银行	保利发展	中国联通	上汽集团	复星医药

0.913414	0.9042	0.889614	0.84818	0.886987	0.897999	0.920489	0.857451	0.919015	0.914951	0.888809	0.89394
0.924438	0.961264	0.931825	0.929918	0.980502	0.974816	0.962112	0.958642	0.970723	0.978377	0.966167	1
0.909141	0.975816	0.955749	0.935183	0.972288	0.9663	0.943498	0.96276	0.94573	0.943381	1	0.966167
0.948416	0.953688	0.919249	0.888317	0.964783	0.97641	0.986981	0.928709	0.988449	1	0.943381	0.978377
0.938932	0.949368	0.94489	0.910009	0.972937	0.977163	0.985795	0.93972	1	0.988449	0.94573	0.970723
0.850627	0.959184	0.95561	0.974893	0.976223	0.93288	0.940072	1	0.93972	0.928709	0.96276	0.958642
0.941824	0.957364	0.946474	0.901226	0.965743	0.964908	1	0.940072	0.985795	0.986981	0.943498	0.962112
0.946236	0.958085	0.933314	0.895267	0.980186	1	0.964908	0.93288	0.977163	0.97641	0.9663	0.974816
0.908594	0.965573	0.961339	0.948797	1	0.980186	0.965743	0.976223	0.972937	0.964783	0.972288	0.980502
0.786957	0.93888	0.929894	1	0.948797	0.895267	0.901226	0.974893	0.910009	0.888317	0.935183	0.929918
0.893397	0.931111	1	0.929894	0.961339	0.933314	0.946474	0.95561	0.94489	0.919249	0.955749	0.931825
0.893752	1	0.931111	0.93888	0.965573	0.958085	0.957364	0.959184	0.949368	0.953688	0.975816	0.961264
1	0.893752	0.893397	0.786957	0.908594	0.946236	0.941824	0.850627	0.938932	0.948416	0.909141	0.924438
0.775697	0.593851	0.588991	0.44057	0.635164	0.732808	0.627254	0.497185	0.660511	0.662981	0.640644	0.636925
0.922699	0.894131	0.962304	0.872197	0.947596	0.949254	0.935721	0.903083	0.955566	0.924287	0.921167	0.916271
0.867927	0.904452	0.898178	0.881061	0.899897	0.902927	0.857204	0.903652	0.875143	0.865146	0.954298	0.911679
0.845457	0.955799	0.923365	0.962452	0.949246	0.912815	0.927041	0.979677	0.922353	0.924347	0.95403	0.957177
0.894274	0.977109	0.961841	0.954169	0.985448	0.970791	0.960771	0.977453	0.962416	0.954135	0.98035	0.970537
0.722788	0.926918	0.838529	0.929347	0.862035	0.816712	0.842414	0.909868	0.82084	0.826592	0.894877	0.858097
0.91948	0.9821	0.942253	0.917776	0.981041	0.981497	0.968548	0.944839	0.969516	0.969157	0.969693	0.966512
0.885063	0.832579	0.883429	0.795729	0.905368	0.922356	0.848268	0.835495	0.883292	0.861685	0.889333	0.888333
0.930201	0.962381	0.940744	0.91203	0.9782	0.98267	0.958392	0.941653	0.968203	0.959329	0.965066	0.960911
用友网络	海通证券	长江动力	中国神华	农业银行	平安银行	新华保险	工商银行	中国太保	中国人寿	中国建筑	华泰证券

1
0.89394
0.888809
0.914951
0.919015
0.857151
0.920489
0.897999
0.886987
0.84818
0.889614
0.9042
0.913414
0.659762
0.892704
0.83535
0.855241
0.891702
0.82147
0.904814
0.806831
0.919452
中国石油

附录二：熊市股票池的相关系数矩阵

中国石油	0.275506	0.774738	0.531737	0.775963	0.867816	0.864791	0.496841	0.712472	0.908806
华泰证券	-0.33265	0.516611	0.771051	0.565818	0.073548	0.200183	0.44279	-0.0387	0.237314
中国建筑	0.533137	0.843259	0.346345	0.738447	0.7664	0.592095	0.835082	0.87958	0.770047
中国人寿	0.236356	0.821242	0.674792	0.8679	0.685992	0.610165	0.797885	0.610587	0.799378
中国太保	0.639948	0.746724	0.393363	0.660057	0.945636	0.782688	0.505169	0.90702	0.884678
工商银行	0.506806	0.84992	0.449918	0.739889	0.9733	0.858042	0.579509	0.89542	0.904561
新华保险	0.5477103	0.809743	0.574885	0.794697	0.875484	0.746694	0.644255	0.827349	0.867727
平安银行	0.534814	0.800873	0.420947	0.730538	0.976668	0.8625	0.526867	0.892206	0.935258
农业银行	0.518837	0.875867	0.501958	0.749987	0.955444	0.86953	0.572671	0.894271	0.872351
中国神华	0.44831	0.904271	0.504227	0.827257	0.94257	0.833295	0.669572	0.907078	0.932631
长江动力	0.359531	0.78609	0.406972	0.771157	0.925729	0.725382	0.640726	0.838007	0.920425
海通证券	0.049287	0.291225	0.693488	0.291088	-0.13506	-0.03373	0.253534	-0.04197	-0.0207
用友网络	-0.37287	0.198099	0.741132	0.367726	-0.1381	0.064473	0.172139	-0.2772	0.033408
复星医药	0.397013	0.836912	0.453018	0.806051	0.903232	0.79622	0.655463	0.854387	1
上汽集团	0.667877	0.795597	0.273964	0.640707	0.92202	0.736678	0.597907	1	0.854387
中国联通	0.174466	0.754608	0.412262	0.788189	0.539693	0.357441	1	0.597907	0.655463
保利发展	0.525409	0.761513	0.536753	0.679745	0.843033	1	0.357441	0.736678	0.79622
招商银行	0.561173	0.79954	0.348115	0.707932	1	0.843033	0.539693	0.92202	0.903232
三一重工	0.204513	0.870431	0.694821	1	0.707932	0.679745	0.788189	0.640707	0.806051
中信证券	0.120114	0.658718	1	0.694821	0.348115	0.536753	0.412262	0.273964	0.453018
中国石化	0.359962	1	0.658718	0.870431	0.79954	0.761513	0.754608	0.795597	0.836912
浦发银行	1	0.359962	0.120114	0.204513	0.561173	0.525409	0.174466	0.667877	0.397013
浦发银行		中国石化	中信证券	三一重工	招商银行	保利发展	中国联通	上汽集团	复星医药

0.193724	-0.10401	0.871069	0.888293	0.865083	0.91021	0.796965	0.892613	0.813005	0.760636	0.573497	0.346043
0.810431	0.553658	0.237202	0.262509	0.239357	0.135863	0.272635	0.201711	0.04628	0.612709	0.162702	1
-0.1267	0.166602	0.735058	0.834413	0.783622	0.728521	0.751102	0.770577	0.748439	0.686648	1	0.162702
0.37085	0.271369	0.787435	0.777658	0.71868	0.739129	0.848868	0.727483	0.691798	1	0.686648	0.612709
-0.09821	-0.0123	0.883335	0.898766	0.891344	0.962407	0.913297	0.912129	1	0.691798	0.748439	0.04628
-0.01668	-0.05701	0.915674	0.952394	0.988114	0.953329	0.866141	1	0.912129	0.727483	0.770577	0.201711
0.110977	0.165269	0.880293	0.874627	0.855955	0.90881	1	0.866141	0.913297	0.848868	0.751102	0.272635
-0.06652	-0.10929	0.921312	0.942701	0.930337	1	0.90881	0.953329	0.962407	0.739129	0.728521	0.135863
0.011196	0.011732	0.888709	0.952481	1	0.930337	0.855955	0.988114	0.891344	0.71868	0.783622	0.239357
0.024469	0.023437	0.915492	1	0.952481	0.942701	0.874627	0.952394	0.898766	0.777658	0.834413	0.262509
0.061725	-0.04659	1	0.915492	0.888709	0.921312	0.880293	0.915674	0.883335	0.787435	0.735058	0.237202
0.641626	1	-0.04659	0.023437	0.011732	-0.10929	0.165269	-0.05701	-0.0123	0.271369	0.166602	0.553658
1	0.641626	0.061725	0.024469	0.011196	-0.06652	0.110977	-0.01668	-0.09821	0.37085	-0.1267	0.810431
0.033408	-0.0207	0.920425	0.932631	0.872351	0.935258	0.867727	0.904561	0.884678	0.799378	0.770047	0.237314
-0.2772	-0.04197	0.838007	0.907078	0.894271	0.892206	0.827349	0.89542	0.90702	0.610587	0.87958	-0.0387
0.172139	0.253534	0.640726	0.669572	0.572671	0.526867	0.644255	0.579509	0.505169	0.797885	0.835082	0.44279
0.064473	-0.03373	0.725382	0.833295	0.86953	0.8625	0.746694	0.858042	0.782688	0.610165	0.592095	0.200183
-0.1381	-0.13506	0.925729	0.94257	0.955444	0.976668	0.875484	0.9733	0.945636	0.685992	0.7664	0.073548
0.367726	0.291088	0.771157	0.827257	0.749987	0.730538	0.794697	0.739889	0.660057	0.8679	0.738447	0.565818
0.741132	0.693488	0.406972	0.504227	0.501958	0.420947	0.574885	0.449918	0.393363	0.674792	0.346345	0.771051
0.198099	0.291225	0.78609	0.904271	0.875867	0.800873	0.809743	0.84992	0.746724	0.821242	0.843259	0.516611
-0.37287	0.049287	0.359531	0.44831	0.518837	0.534814	0.547103	0.506806	0.639948	0.236356	0.533137	-0.33265
用友网络	海通证券	长江动力	中国神华	农业银行	平安银行	新华保险	工商银行	中国太保	中国人寿	中国建筑	华泰证券

1
0.346043
0.573497
0.760636
0.813005
0.892613
0.796965
0.91021
0.865083
0.888293
0.871069
-0.10401
0.193724
0.908806
0.712472
0.496841
0.864791
0.867816
0.775963
0.531737
0.774738
0.275506
中国石油

附录三：震荡区间股票池的相关系数矩阵

中国石油	0.906105	0.771051	0.883654	0.781148	0.873953	0.829613	0.908207	0.866696	0.594061
华泰证券	0.948166	0.847831	0.962862	0.808052	0.961402	0.942824	0.943184	0.898622	0.547023
中国建筑	0.962286	0.843298	0.987189	0.860138	0.980142	0.933865	0.970084	0.92042	0.544954
中国人寿	0.954593	0.840284	0.962067	0.783863	0.953031	0.918177	0.950303	0.907835	0.59602
中国太保	0.965994	0.866471	0.962399	0.775627	0.96293	0.914937	0.961885	0.945871	0.591798
工商银行	0.918859	0.776245	0.93547	0.8775	0.968574	0.973146	0.930391	0.88202	0.362892
新华保险	0.958094	0.829808	0.962995	0.810432	0.967205	0.929233	0.955941	0.923598	0.557307
平安银行	0.976365	0.905733	0.979559	0.751544	0.961924	0.879681	0.955178	0.939541	0.672643
农业银行	0.971365	0.881533	0.977556	0.8162	0.983152	0.935844	0.955336	0.935803	0.549338
中国神华	0.883403	0.719506	0.899219	0.906785	0.940313	0.953505	0.8893	0.840735	0.295473
长江动力	0.921753	0.847081	0.928965	0.783651	0.950198	0.898162	0.938358	0.955322	0.488348
海通证券	0.948884	0.775482	0.981232	0.902258	0.971162	0.939952	0.943178	0.868668	0.492045
用友网络	0.910935	0.858899	0.898348	0.637218	0.865206	0.797046	0.917143	0.904659	0.72853
复星医药	0.640313	0.767266	0.59311	0.188585	0.49627	0.29651	0.585501	0.656595	1
上汽集团	0.942071	0.927599	0.921695	0.66581	0.918497	0.823255	0.944171	1	0.656595
中国联通	0.955069	0.864336	0.959148	0.802782	0.954367	0.905227	1	0.944171	0.585501
保利发展	0.884167	0.702895	0.912206	0.898851	0.945236	1	0.905227	0.823255	0.29651
招商银行	0.960204	0.836604	0.980084	0.854502	1	0.945236	0.954367	0.918497	0.49627
三一重工	0.787837	0.519233	0.827705	1	0.854502	0.898851	0.802782	0.66581	0.188585
中信证券	0.972085	0.864985	1	0.827705	0.980084	0.912206	0.959148	0.921695	0.59311
中国石化	0.89046	1	0.864985	0.519233	0.836604	0.702895	0.864336	0.927599	0.767266
浦发银行	1	0.89046	0.972085	0.787837	0.960204	0.884167	0.955069	0.942071	0.640313
浦发银行		中国石化	中信证券	三一重工	招商银行	保利发展	中国联通	上汽集团	复星医药

0.894924	0.887009	0.867593	0.817229	0.862255	0.87757	0.904238	0.828327	0.901481	0.896455	0.886409	0.875377
0.904298	0.94817	0.909074	0.904709	0.975903	0.968286	0.966279	0.943979	0.971503	0.982658	0.95993	1
0.894307	0.977165	0.949082	0.915551	0.977202	0.966101	0.967009	0.953367	0.95978	0.958365	1	0.95993
0.937201	0.947936	0.903342	0.866241	0.959028	0.975757	0.984575	0.918738	0.985961	1	0.958365	0.982658
0.925168	0.941838	0.93511	0.892748	0.969059	0.97655	0.983355	0.932735	1	0.985961	0.95978	0.971503
0.805517	0.944946	0.942316	0.969656	0.970135	0.907917	0.939978	1	0.932735	0.918738	0.953367	0.943979
0.931562	0.957499	0.942227	0.888319	0.963761	0.964936	1	0.939978	0.983355	0.984575	0.967009	0.966279
0.931906	0.943356	0.911931	0.861449	0.974088	1	0.964936	0.907917	0.97655	0.975757	0.966101	0.968286
0.882187	0.954928	0.949648	0.934228	1	0.974088	0.963761	0.970135	0.969059	0.959028	0.977202	0.975903
0.722973	0.919087	0.906428	1	0.934228	0.861449	0.888319	0.969656	0.892748	0.866241	0.915551	0.904709
0.86282	0.908239	1	0.906428	0.949648	0.911931	0.942227	0.942316	0.93511	0.903342	0.949082	0.909074
0.86162	1	0.908239	0.919087	0.954928	0.943356	0.957499	0.944946	0.941838	0.947936	0.977165	0.94817
1	0.86162	0.86282	0.722973	0.882187	0.931906	0.931562	0.805517	0.925168	0.937201	0.894307	0.904298
0.72853	0.492045	0.488348	0.295473	0.549338	0.672643	0.557307	0.362892	0.591798	0.59602	0.544954	0.547023
0.904659	0.868668	0.955322	0.840735	0.935803	0.939541	0.923598	0.88202	0.945871	0.907835	0.92042	0.898622
0.917143	0.943178	0.938358	0.8893	0.955336	0.955178	0.955941	0.930391	0.961885	0.950303	0.970084	0.943184
0.797046	0.939952	0.898162	0.953505	0.935844	0.879681	0.929233	0.973146	0.914937	0.918177	0.933865	0.942824
0.865206	0.971162	0.950198	0.940313	0.983152	0.961924	0.967205	0.968574	0.96293	0.953031	0.980142	0.961402
0.637218	0.902258	0.783651	0.906785	0.8162	0.751544	0.810432	0.8775	0.775627	0.783863	0.860138	0.808052
0.898348	0.981232	0.928965	0.899219	0.977556	0.979559	0.962995	0.93547	0.962399	0.962067	0.987189	0.962862
0.858899	0.775482	0.847081	0.719506	0.881533	0.905733	0.829808	0.776245	0.866471	0.840284	0.843298	0.847831
0.910935	0.948884	0.921753	0.883403	0.971365	0.976365	0.958094	0.918859	0.965994	0.954593	0.962286	0.948166
用友网络	海通证券	长江动力	中国神华	农业银行	平安银行	新华保险	工商银行	中国太保	中国人寿	中国建筑	华泰证券

1
0.875377
0.886409
0.896455
0.901481
0.828327
0.904238
0.87757
0.862255
0.817229
0.867593
0.887009
0.894924
0.594061
0.869696
0.908207
0.829613
0.873953
0.781148
0.883654
0.771051
0.906105
中国石油