

硕士论文正式答辩

嵌入TCN时序神经网络和多因子模型的 Black-Litterman量化资产配置模型

韩江浩

2024年5月



目录

contents

1 研究背景

2 本文创新点

3 模型理论基础

4 新模型构建详细步骤

5 实证检验与总结

1

研究背景



1.1 研究背景：资产配置意义重大，BL模型尚待优化

- **资产配置意义重大。**巴菲特常常强调选择正确的投资机会比频繁交易更重要，中外多名学者也已经证实资产配置对于投资组合业绩的贡献力度极大。根据Ibbotson和Kaplan（2000），资产配置对美国投资组合业绩的贡献高达90%，而根据杨朝军和周仕盈（2021），资产配置可以解释国内不同混合型基金业绩差异的40%。

资产配置不仅可以提升组合表现，也便于做业绩归因分析，值得深入挖掘。

- **BL模型尚待优化。**（1）依赖市场均衡假设，在现实中尤其是A股市场假设很难成立。（2）BL模型中投资者主观观点的设置随机性较大，不利于模型稳定性。（3）容易过拟合，对收益率均值的估计误差敏感。

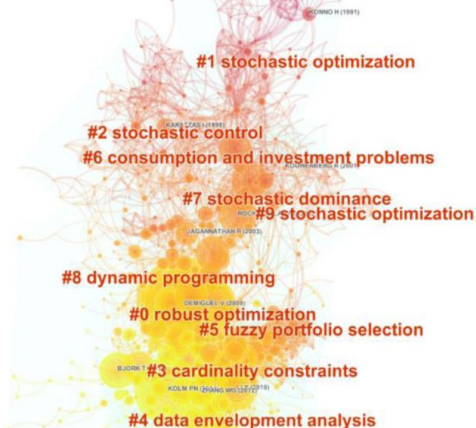
本次硕士论文是深入研究资产配置的契机，笔者希望能进一步探索BL模型在中国股票市场的应用，为沪深A股的投资者提供新的思路和新的工具。



1.2 资产配置研究现状：传统配置模型，改造目标&约束

- **研究现状：**目前3大主流资产配置模型分别为：均值方差模型、Black-Litterman模型、风险平价模型。均值方差模型、Black-Litterman模型更加关注收益与风险的平衡，而风险平价模型则主要聚焦于风险控制。
- 近年来对资产配置模型的研究，大多数是在风险平价模型、Black-Litterman模型的基础上，改善估计方法、增加更多约束、提供更多启发式优化算法、将单目标优化扩展至多目标优化、将单期配置模型拓展到多期等。

图 资产配置模型研究主要分支



资料来源：《Dynamic changes and multi-dimensional evolution of portfolio optimization》



1.3 收益预测研究现状：横截面多因子，时序神经网络

- **资产收益横截面的预测，以多因子模型为主：**起源于CAPM，利用风险因子线性预测资产收益，是目前学界和业界研究资产收益的主要手段。
- **资产收益在时间序列上的预测，多为RNN，近年TCN开始兴起：**深度学习模型凭借其对非线性关系的拟合和强大的数据特征学习能力，为提升预测精度提供了新的思路，LSTM模型在时间序列预测领域占据主导地位。
- Bai, S et al.(2018) 发现卷积架构更为简单的TCN模型在多种序列建模任务上优于典型的循环网络如LSTM。TCN可以并行处理序列信息而不是顺序处理，比循环神经网络具有更稳定的梯度，可以提高模型的计算效率，同时可以保持序列中长期历史信息记忆。TCN模型在时间序列预测上的优势被发掘之后，学者们尝试增加其特征维度，提升其数据频度。

2

本文创新点



3.1 BL模型改造动机：减小估计误差，增强稳定性

- **BL模型简介与优势：**在均值方差模型的基础上，引入主观观点，利用利用贝叶斯收缩，解决均值方差模型对输入过于敏感的问题。
- **BL模型继承了均值方差模型的弱点：**1.高度依赖历史样本导致过拟合；2.资产收益的估计误差对投资组合影响较大；3.对金融市场非线性信息的吸收不充分。
- **BL模型独有弱点：**1.对主观观点的依赖性较强，模型的有效性和准确性依赖于投资者提供的观点和信息；2.在市场均衡假设下，求出资产收益的先验分布，而现实中市场难以达到均衡。
- 本文思路：
 - 1.抛弃市场均衡假设，利用TCN时序神经网络充分吸收历史信息，对资产收益做短期预测，构建收益均值的先验分布，同时起到降低估计误差的作用；
 - 2.TCN+多因子系统性生成“主观观点”，减弱主观依赖，提升模型稳定性。



3.2 核心参考文献1：BL模型与多因子模型结合

(1) Kolm, P. N. and Ritter, G. (2020). Factor Investing with Black–Litterman–Bayes: Incorporating Factor Views and Priors in Portfolio Construction. *The Journal of Portfolio Management*, 47(2):113–126.

主要思路：基于多因子模型，利用BL模型的Bayes框架，将因子收益率的先验信息与对因子收益率的观点结合，构建Multi-Factor Risk Premium Portfolio（多因子风险溢价组合），即因子配置。

实证结果：基于常用的美国股票因子，新构建的BLB因子配置模型，优于MVO(均值方差最优)因子配置模型。

(2) Li, C. L., & Hsieh, C. H. (2023). On Adaptive Portfolio Management with Dynamic Black–Litterman Approach. arXiv preprint arXiv:2307.03391.

主要思路：基于BL模型，针对主观观点部分，嵌入改造后的多因子模型（使用弹性网回归）来系统生成投资者观点，减少估计误差。

实证结果：基于道琼斯30指数成分股和美国国债，这种自适应配置模型，具有计算优势，且投资组合表现优于原指数（市场组合）和均值方差组合（最大化效用函数）。



3.3 核心参考文献2：TCN预测股票收益

(1) Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. Z., & Chen, H. (2019, May). Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 678-685).

主要思路：构建知识驱动的TCN模型，将股价、新闻、结构化知识图谱作为特征输入，预测股价变动趋势，可以捕捉到事件驱动带来的股价突变信息，并对事件影响有一定解释效果。

实证结果：利用道琼斯30指数历史价格、财经新闻、结构化知识图谱（建立事件之间的联系），预测指数走势，通过对比Accuracy和F1 score, KDTCN的预测结果优于ARIMA、LSTM、TCN以及变种（除KDTCN外），且可以解释股价突变。



3.4 本文创新点1：全新方法获得BL模型先验分布

- **创新点1：TCN + Bootstrap**获取收益均值的先验分布。
- 本文**抛弃**传统BL模型中的**市场均衡假设**，采用时间卷积网络（Temporal Convolutional Network, TCN）来对资产收益进行**短期序列预测**。在实现序列预测之后，利用Bootstrap方法，对预测结果**重复抽样**，每次抽样可计算**预期收益率的均值**，通过**多次抽样**，构建预期收益均值的**经验分布**，可估计出预期收益均值的先验期望与先验方差，从而得到**预期收益均值的先验分布**。



3.5 本文创新点2：融合机器学习与多因子，量化观点

- **创新点2：投资者主观观点的系统化设置。**
- 本文融合非线性机器学习方法（TCN）和传统线性统计模型（多因子模型）的优势，将TCN与多因子模型的预测结果有机结合，系统化地设置主观观点，在提升预测精度的同时，增强模型的稳定性。



3.6 本文创新点3：交叉验证，增强模型稳定性与解释性

- **创新点3：双重预测，交叉验证，可解释性强。**
- 本文通过结合**TCN**的深度学习能力和**多因子模型**的经济解释力，增强模型的抗干扰能力并减少过拟合的风险。**TCN**的存在，解决了多因子模型无法充分吸收金融资产高度复杂且充满**噪音**的**非线性信息**的问题；而**多因子模型**的存在，则有效解决了纯机器学习模型可能出现的**过拟合问题**和**缺乏解释性**的问题。

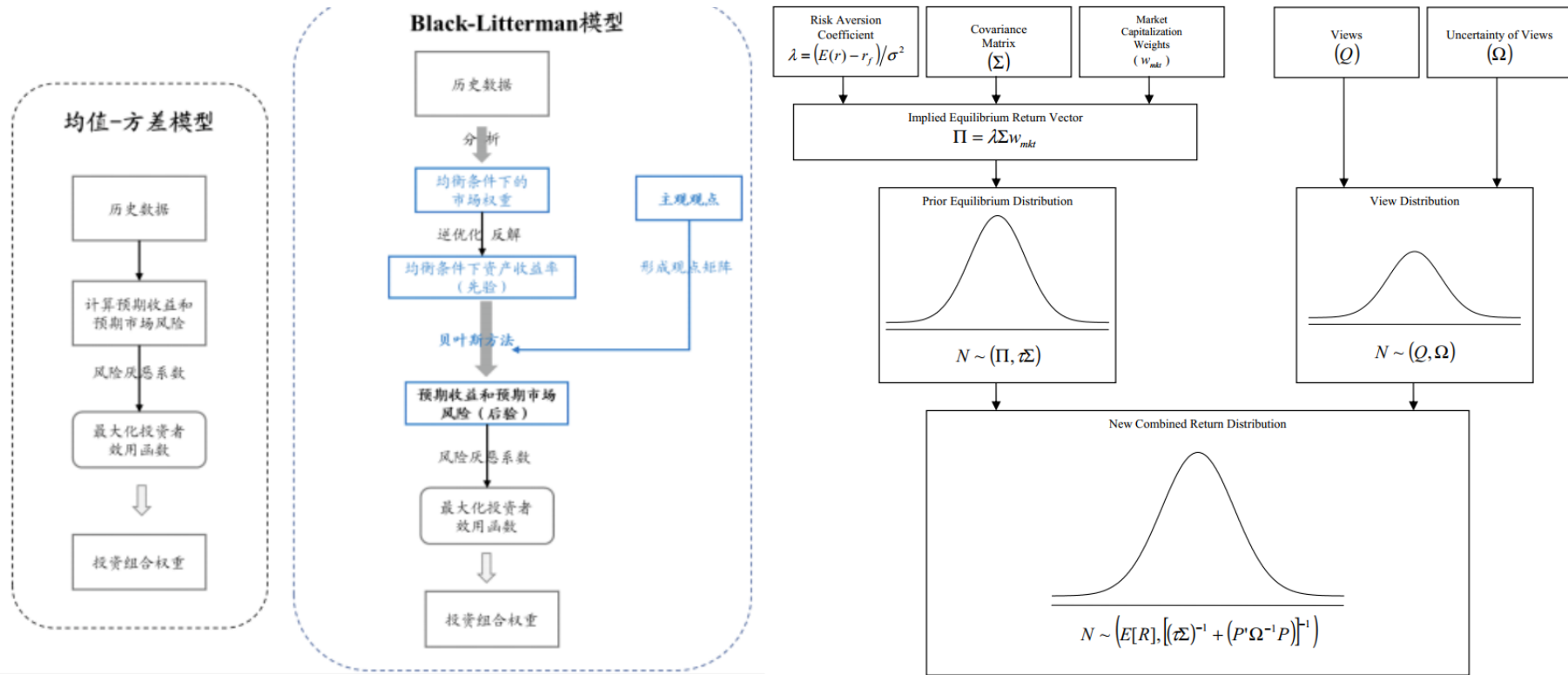
3

模型理论基础



3.1 BL模型原理简介

- 将主观观点引入均值方差模型，更新收益均值的先验分布，得到收益均值的后验分布，将后验分布作为参数输入均值方差模型，解出投资组合权重分配。





3.2 BL模型中的收益均值先验分布

- 收益均值的先验分布：假设市场均衡，根据 CAPM 模型，以实际市值 w 为权重的资产组合就是市场均衡组合。

(1) 假设收益正态分布：
$$R \sim N(\mu, \Sigma)$$

(2) 计算风险厌恶系数：
$$\delta = \frac{E(r_m) - r_f}{\sigma_m^2}$$

(3) 最大化效用函数：
$$U = w^T \Pi - \frac{\delta}{2} w^T \Sigma w$$

(4) 解出均衡收益：
$$\Pi = \delta \Sigma w_{eq}$$

(5) 得到收益均值先验分布：
$$\mu \sim N(\Pi, \tau \Sigma)$$



3.3 BL模型中的主观观点

- 收益均值的主观观点：使用贝叶斯定理，对收益率的先验均值做出更新，以期得到更准确的结果。

(1) P : 主观观点矩阵，是一个 $K \times N$ 矩阵。 K 为观点个数， N 为标的资产个数

(2) μ : 主观预期收益率向量

(3) Q : 投资者预期收益率 $K \times 1$ $P\mu = Q$

(4) Ω : 置信水平矩阵， $K \times K$ 对角矩阵，衡量观点的确信度 $\Omega = \tau(P \sum P')$

(5) 解出后验分布: $\mu_p = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\Pi + P^T\Omega^{-1}Q]$

$$\Sigma_p = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}$$



3.4 TCN(Temporal Convolutional Networks)模型原理

- 序列建模：相同长度的输入与输出，实现序列到序列的预测。 y_t 仅依赖于 x_0 至 x_t 。

$$\hat{y}_0, \dots, \hat{y}_T = f(x_0, \dots, x_T)$$

- 因果卷积：卷积层中，某层 t 时刻的输出，仅依赖上一层0至 t 时刻的元素。
- 扩张卷积：卷积时加入扩张因子 d ，扩大卷积核大小 k ，从而扩大感受野。使用更大范围的输入，计算输出。

$$F(s) = (\mathbf{x} *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot \mathbf{x}_{s-d \cdot i}$$

- 残差连接：在训练深层网络时，缓解梯度消失，使模型保持稳定。
- TCN优势：可实现并行计算，将长序列输入作为整体来处理；扩张卷积可灵活调整感受野大小；梯度稳定。



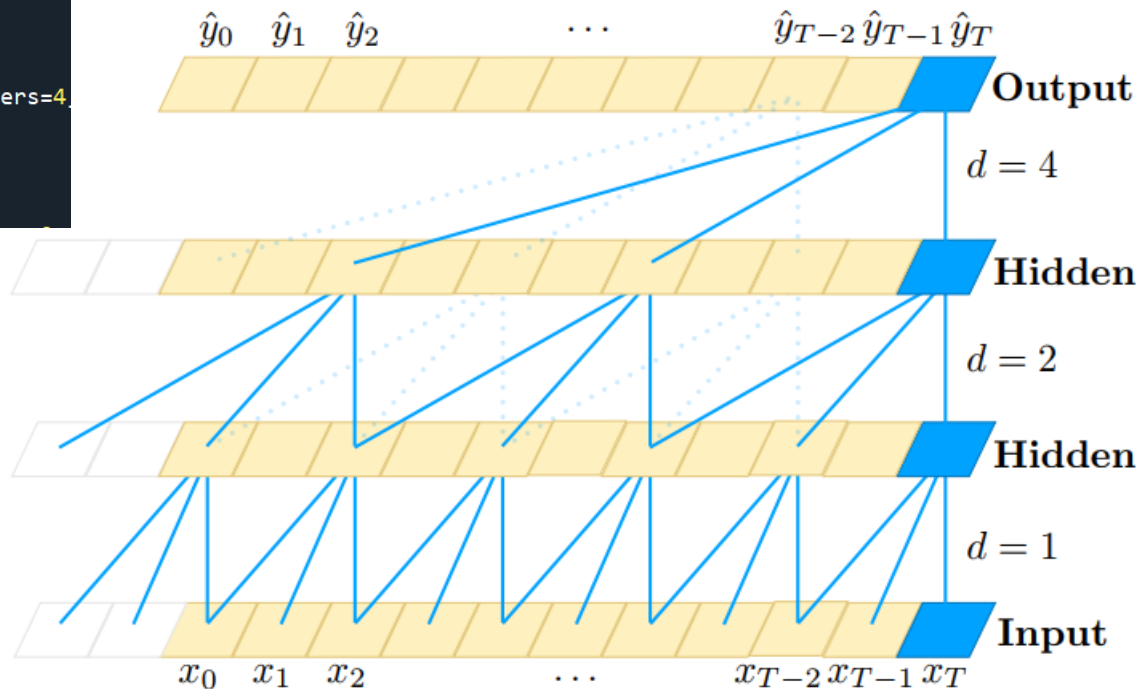
3.5 TCN重要结构：因果卷积+扩张卷积

- 因果卷积：某层t时刻的输出，仅依赖上一层0至t时刻的元素，保证时序预测。
- 扩张卷积：利用扩张因子d，卷积核大小k，扩大感受野，获取更长历史记忆。

```
# 搭建TCN神经网络
# TCN Model Definition
class TCN(nn.Module):
    def __init__(self, num_features, num_layers=4):
        super(TCN, self).__init__()
        layers = []
        for i in range(num_layers):
            dilation_size = 2 ** i
```

- $i = 0: d = 2^0 = 1$
- $i = 1: d = 2^1 = 2$
- $i = 2: d = 2^2 = 4$
- $i = 3: d = 2^3 = 8$

预答辩后修改





3.6 多因子模型简介：利用风险因子，解释资产收益

- 多因子模型核心理念是**将资产的预期收益率分解为一系列可预测的因子的线性组合**，以及相应的残差项。通过引入更多风险因子，来解释股票收益的更多变异性。
- 因子（Factor）：指与资产收益相关联的变量或特征。

$$R_i = a_i + \sum_{k=1}^K b_{ik} F_k + \varepsilon_i$$

- 以预测股票收益的多因子模型举例， R_i 代表第*i*只股票的预期收益率，截距项 a_i 代表第*i*只股票在不受各种风险因子影响时的预期回报水平， F_k 为因子取值， b_{ik} 为第*i*只股票在第*k*个因子上的因子暴露度，衡量股票收益率对因子值变动的敏感程度， ε_i 为均值为零的误差项，代表第*i*只股票未能被因子模型解释的收益部分。

4

新模型构建详细步骤



4.1 TCN实现收益率序列短期预测1

- TCN（时序卷积网络）模型对12只股票行业指数预期收益率序列展开短期预测。
- 标的资产为申万一级行业指数（ETF）：（1）分散风险；（2）反映行业趋势；（3）简化配置过程；（4）资产流动性好；（5）可实施行业轮动策略。
- TCN主要设置：
- 1维输入输出：对数收益率。
- 8维特征：市盈率PE(TTM)、市净率PB(LF)、股息率(近12个月)、历史PEG、涨跌幅、换手率、成交量、成交额。估值+红利+动量+市场情绪+流动性。
- 优化器：Adam优化器。
- 损失函数：均方误差损失（MSE loss）
- 设置超参数组合用于网格搜索：

超参数	含义	可能取值
num_layers	卷积层数	[2, 3, 4]
out_features	卷积层输出特征数	[16, 32]
dropout	神经元随机失活比率	[0.2, 0.3]
lr	学习率	[0.01, 0.001]
weight_decay	权重衰减	[0.001, 0.0001]



4.2 TCN实现收益率序列短期预测2

- **训练+验证数据集**：选取2000年1月4日至2023年3月7日所有交易日，按照**7：3**的比例划分为训练数据集和验证数据集。第一期60个交易日的预测完成后，将2023年3月8日至2023年6月5日共计**60个交易日**加入训练数据集和验证数据集，将数据集范围**扩充**为2000年1月4日至2023年6月5日，**重新按照7：3的比例**划分为训练数据集和验证数据集。第二期60个交易日的预测完成后，将数据集范围扩充为2000年1月4日至2023年8月30日，重新按照7：3的比例划分为训练数据集和验证数据集。
- **测试/预测数据集**：本文为了验证模型的有效性和稳定性，进行**三期滚动测试**，单期测试天数为60个交易日，共计180个交易日。连续的3个预测期分别为：2023年3月8日至2023年6月5日、2023年6月6日至2023年8月30日、2023年8月31日至2023年11月30日。**每次单期预测完成后**，将该期60个交易日的数据，添加到训练、验证数据集中，**重新对TCN进行训练**，并预测下一期的收益序列。
- **使用最优模型预测**：在验证集上遍历所有超参数组合，找出最优模型预测下一期。



4.2 TCN实现收益率序列短期预测3: TCN优劣势

- **参数量少，训练数据充足，可充分训练TCN模型：**TCN模型的初始总参数量为1105个，此参数规模在深度学习模型中属于轻量级。本研究采用了长达23年的股票日度数据，以确保充分训练TCN模型。
- **优势：**训练效率高、过拟合风险低。
- **劣势：**特征学习能力不够、泛化能力受限。

预答辩后修改

```
'''
计算TCN模型参数量
'''

num_features = 8 # 输入特征数
num_layers = 4 # 层数
out_features = 16 # 每层输出特征数
dropout = 0.3 # 丢弃率 (这里不影响参数计算)
kernel_size = 1 # 核大小, 因为1维卷积

# 卷积层和批量归一化层的参数计算
conv_params_per_layer = (num_features * kernel_size * out_features) + out_features if num_layers > 0 else 0
batch_norm_params_per_layer = out_features * 2 # gamma和beta

# 第一层之外的卷积层和批量归一化层的参数
additional_layers_params = ((out_features * kernel_size * out_features) + out_features + out_features * 2) * (num_layers - 1)

# 全连接层的参数
fc_params = (out_features * 1) + 1

# 总参数量
total_params = conv_params_per_layer + batch_norm_params_per_layer + additional_layers_params + fc_params
total_params
```

```
Help Variable Explorer Plots Files
Console 1/A X
- 1)
...: # 全连接层的参数
...: fc_params =
(out_features * 1) + 1
...: # 总参数量
...: total_params =
conv_params_per_layer +
batch_norm_params_per_layer +
additional_layers_params +
fc_params
...: total_params
Out[3]: 1105
```




4.2 Bootstrap得到预期收益率均值的先验分布

- **Bootstrap方法：**非参数统计方法，不依赖于数据的具体分布来估计统计量的分布。通过重复抽样，能够模拟出收益率的不确定性和变异性。本文将Bootstrap抽样的次数设置为**1000次**。这个数字足够大，能够提供稳健的统计估计。
- **正态假设：**针对每期60个交易日的预测，使用Bootstrap方法计算出预期收益率均值的期望与方差，在**预期收益率均值服从正态分布的假定**之下，确定预期收益率均值的先验分布。每次单期预测为60个交易日，拥有大于30的原始样本量，可以近似认为抽样形成的分布是正态分布。
- **构建经验分布：**每次从预测出的60个交易日收益中随机抽取**60个样本**（允许重复），模拟实际的收益率分布，对这60个样本计算均值，得到**预期收益率的均值**，并储存起来。将此过程**重复1000次**，于是得到1000个预期收益率的均值样本，基于1000个样本，可以计算出预期收益率均值的**期望与方差**。



4.3 多因子模型对收益率均值的预测

- **FF5模型：**使用Fama-French五因子模型对股票行业指数的预期收益率均值展开短期预测。因子数据来自国泰安CSMAR数据库中的沪深A股Fama-French五因子。

TradingDate	RiskPremium	SMB	HML	RMW	CMA
2000/1/4	0.030801	-0.004573	-0.00497	0.000279	0.002189
2000/1/5	0.003616	0.005409	0.000409	-0.008477	0.003345
2000/1/6	0.04094	-0.010635	0.001252	0.011655	-0.006575
2000/1/7	0.037513	-0.003423	0.001015	0.003647	-0.002661
2000/1/10	0.020559	0.002515	-0.000892	0.001481	-0.003595
...
2023/8/28	0.012068	0.001861	0.011136	-0.000153	0.004417
2023/8/29	0.009163	0.017074	-0.015442	0.001511	0.001651
2023/8/30	-0.001488	0.003512	-0.011618	0.005378	-0.000891

- **回归+预测：**连续的3个预测期分别为：2023年3月8日至2023年6月5日、2023年6月6日至2023年8月30日、2023年8月31日至2023年11月30日。对于每个预测期，仅使用往前**3年**的历史数据构建回归模型并预测，目的是**捕获快速变化的短期趋势**。



4.4 量化生成投资者主观观点

- **主观观点系统生成规则：**融合两种模型的预测结果形成主观观点，此投资观点即是两种模型的预测集成。

预期收益率均值的方向	细分预测情况	观点设置规则
同向	TCN 正收益, FF5 正收益	最终预测结果 = $0.8 * TCN + 0.2 * FF5$
	TCN 负收益, FF5 负收益	
反向	TCN 正收益, FF5 负收益	最终预测结果 = TCN
	TCN 负收益, FF5 正收益	

规则设置详细解释：

- （1）预测方向同向：TCN预测方向得到FF5验证，观点在一定程度上支持先验信息。继续使用**FF5调整TCN的预测幅度**。如果FF5收益高于TCN，则提升预测收益，小幅增加权重分配；如果FF5收益低于TCN，则降低预测收益，小幅减少权重分配。目的是利用多因子模型较强的经济学解释能力来有效**防止TCN过拟合**。
- （2）预测方向反向：FF5仅作点估计，无法刻画收益均值的分布，无法预测长序列，选择相信TCN，维持先验分布。目的是当主观观点与先验信息出现较大冲突时，屏蔽主观观点，维持先验均值，**增强模型稳定性**。此思路受到原始BL模型利用先验分布增强模型稳定性的启发。



4.4 加权生成主观观点的理论依据

(1) Deng, S., Zhang, N., Zhang, W., Chen, J., Pan, J. Z., & Chen, H. (2019, May). Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 678-685).

实证结果：通过对比Accuracy和F1 score, TCN的预测结果显著优于线性模型ARIMA。

4.2.1 Basic Evaluation for TCN. In order to demonstrate that generic TCN architecture can outperform some traditional prediction models, we make comparisons with them, shown in Table 2. Note that all experiments reported in this part are only input with price values.

预答辩后修改

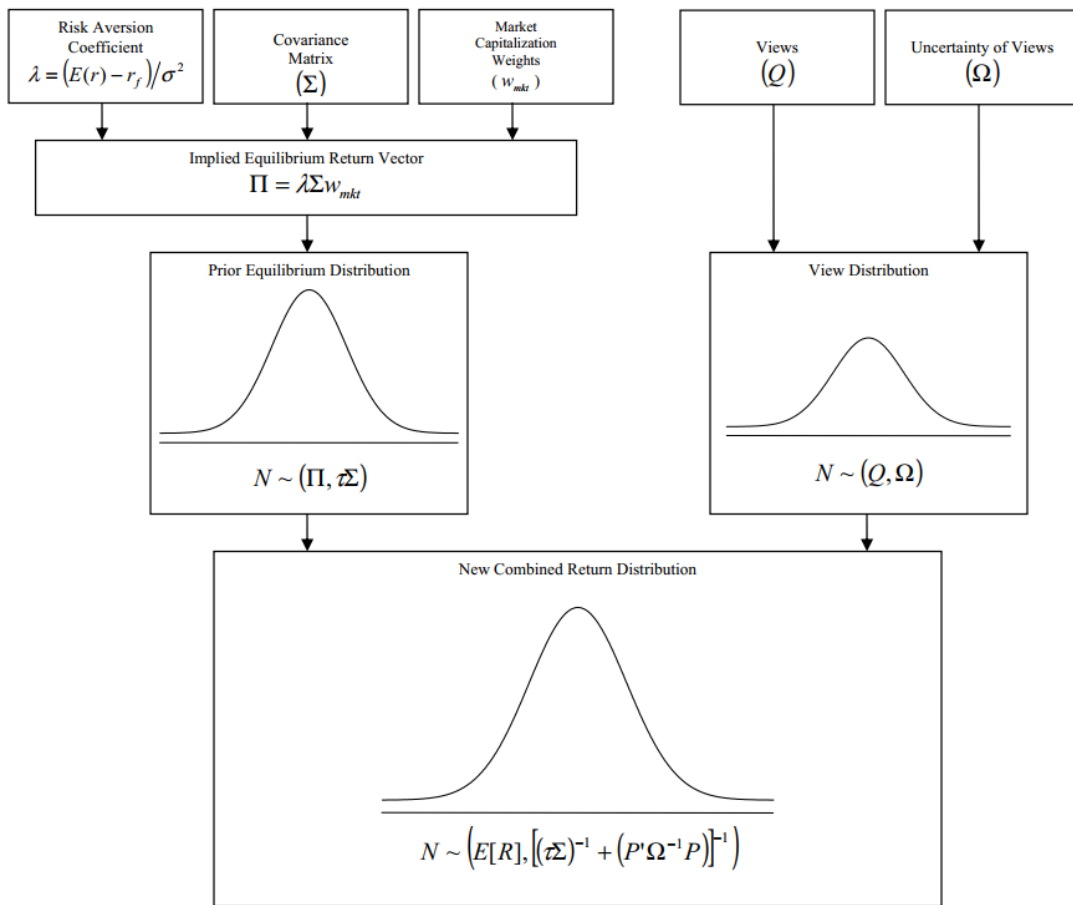
Model	Accuracy	F1 Score
ARIMA	52.40% \pm 0.25%	0.3844 \pm 0.0052
LSTM	56.17% \pm 0.26%	0.5348 \pm 0.0039
CNN	55.41% \pm 0.25%	0.5280 \pm 0.0021
TCN	60.96% \pm 0.26%	0.5317 \pm 0.0036

(2) **TCN**在验证数据集上**最小化MSE**，而多因子模型基于**OLS最小化残差平方和**，损失函数类型不同，很难直言孰优孰劣。**OLS**需要满足较多假设：线性、同方差、误差独立&正态，而最小化**MSE**适用于任何损失最小化问题，无需假设线性/非线性、数据及误差结构，通用性更强。



4.5 求出预期收益率均值的后验分布

- **先验均值：**先验预期收益率均值，本文利用TCN与Bootstrap求出。
- **先验协方差：**资产收益率均值的先验协方差矩阵，用于估计不同资产之间的风险关系，本文利用TCN与Bootstrap求出。
- **主观观点：**本文全部使用绝对观点，即给某资产预期收益的均值赋值，赋值结果由TCN和FF5的预测结果量化生成。





4.6 根据收益收益率均值的后验分布解出权重分配

- **目标函数：**最大化投资组合的夏普比率。

约束条件：

- (1) 所有资产权重的总和必须等于1，确保了投资组合的完整性。
 - (2) **行业集中度限制：**每个资产的权重被限制在0到最大权重（35%）之间，这有助于防止对单一行业过度投资，降低特定行业的系统性风险。
 - (3) **纯多头约束：**限制每个资产的权重大于零，适应中国A股市场的特性，因为其做空操作较为困难。
-
- **具体操作：**在得出12只申万一级行业指数预期收益率均值的后验分布之后，针对每一个包含60个交易日的预测期，利用最大化组合夏普比率的优化算法，得出每个资产在该测试期的最优权重，构建用于该测试期的投资组合。

5

实证检验与总结



5.1 数据来源及解释

- **数据来源：**国泰安CSMAR金融经济研究数据库和Wind数据库。12只申万一级行业指数的历史价格、成交量、行业盈利指标、行业估值指标等数据，针对中国沪深A股市场的Fama-French五因子数据。

序号	数据	数据来源	数据解释
1	Fama French五因子	CSMAR	1、时间范围：2000年1月4日-2023年11月30日；数据频率：日度数据 2、详细内容：RiskPremium [市场风险溢价因子(流通市值加权)]、SMB [市值因子(流通市值加权)]、HML [帐面市值比因子(流通市值加权)]、RMW [盈利能力因子(流通市值加权)]、CMA [投资模式因子(流通市值加权)]
2	农林牧渔(申万)指数	Wind	1、时间范围：2000年1月4日-2023年11月30日 2、数据频率：日度数据 3、详细内容：'收盘价\n[单位] 元', '市盈率PE(TTM)\n[单位] 倍', '市净率PB(LF)\n[单位] 倍', '股息率(近12个月)\n[单位] %', '历史PEG\n[增长率(G值)匹配方法] 当年年报增长率', '涨跌幅\n[单位] %', '换手率\n[单位] %', '成交量\n[单位] 万手', '成交额\n[单位] 万元'
3	有色金属(申万)指数	Wind	
4	食品饮料(申万)指数	Wind	
5	纺织服饰(申万)指数	Wind	
6	医药生物(申万)指数	Wind	
7	交通运输(申万)指数	Wind	
8	房地产(申万)指数	Wind	
9	商贸零售(申万)指数	Wind	
10	社会服务(申万)指数	Wind	
11	非银金融(申万)指数	Wind	
12	汽车(申万)指数	Wind	
13	机械设备(申万)指数	Wind	
14	沪深300指数	Wind	1、时间范围：2000年1月4日-2023年11月30日；数据频率：日度数据
15	中证500指数	Wind	3、详细内容：'收盘价\n[单位] 元'



5.2 样本外测试

- **对照模型：** 中证500指数（ZZ500）、等权重组合（Equal_Weights）、嵌入了TCN时序神经网络的均值方差模型（TCN_MVO）以及嵌入了多因子的BL模型（FF_BL）共计4个对照组合。用于验证新模型（TCN_FF_BL）的有效性。
- **区间业绩指标：**
 - 1、年化收益率： $\text{Retrun} = (1 + r)^{\frac{365}{T}} - 1$
 - 2、年化波动率： $\sigma_p = \sigma_0 \cdot \sqrt{\frac{365}{T}}$
 - 3、年化夏普比率： $\text{Sharpe Ratio} = \frac{E(R_p) - R_f}{\sigma_p}$
 - 4、完整测试期最大回撤： $\text{Max Drawdown} = \frac{\max(P_x - P_y)}{P_x}$
- **区间净值曲线：** 可视化不同投资策略在特定期间的绩效表现。
- **每个测试期中各组合的持仓情况：** 业绩归因，分析收益来源与策略改进。

5.3 测试期各组合的持仓情况

- 第一测试期（2023.03.08-2023.06.05）各投资组合持仓情况。

[illegible]

- 第二测试期（2023.06.06-2023.08.30）各投资组合持仓情况。

[illegible]

- **第三测试期（2023.08.31-2023.11.30）各投资组合持仓情况。**

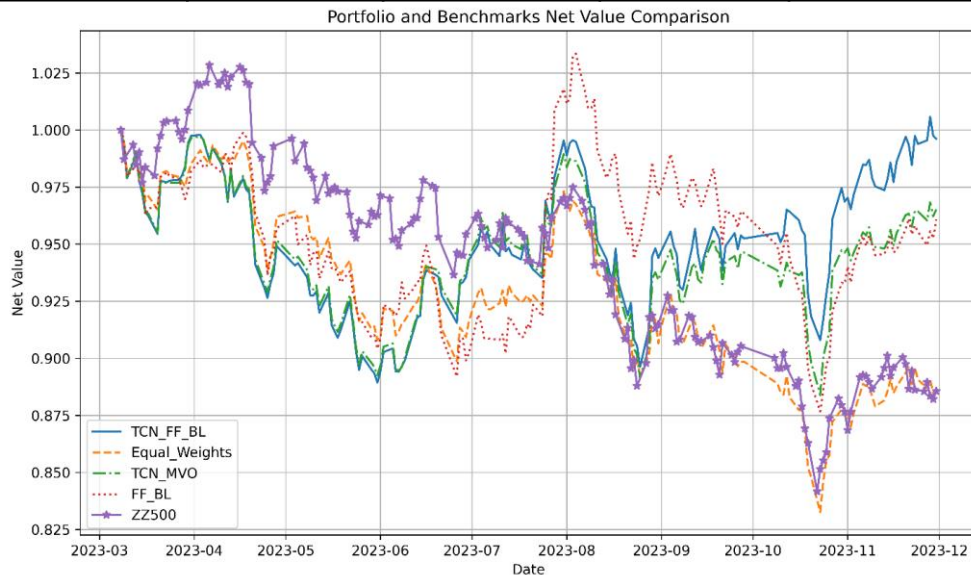
[illegible]



5.4 样本外测试结果

- 原始模型：TCN模型训练集:验证集=7:3，多因子模型使用Fama-French五因子。

投资组合	年化收益	年化波动	最大回撤	年化夏普
TCN_FF_BL	-0.54%	15.64%	-11.06%	-0.03
Equal_Weights	-15.60%	12.93%	-16.74%	-1.21
TCN_MVO	-4.84%	14.90%	-11.61%	-0.32
FF_BL	-5.44%	16.65%	-15.16%	-0.33
ZZ500	-15.64%	13.22%	-18.16%	-1.18

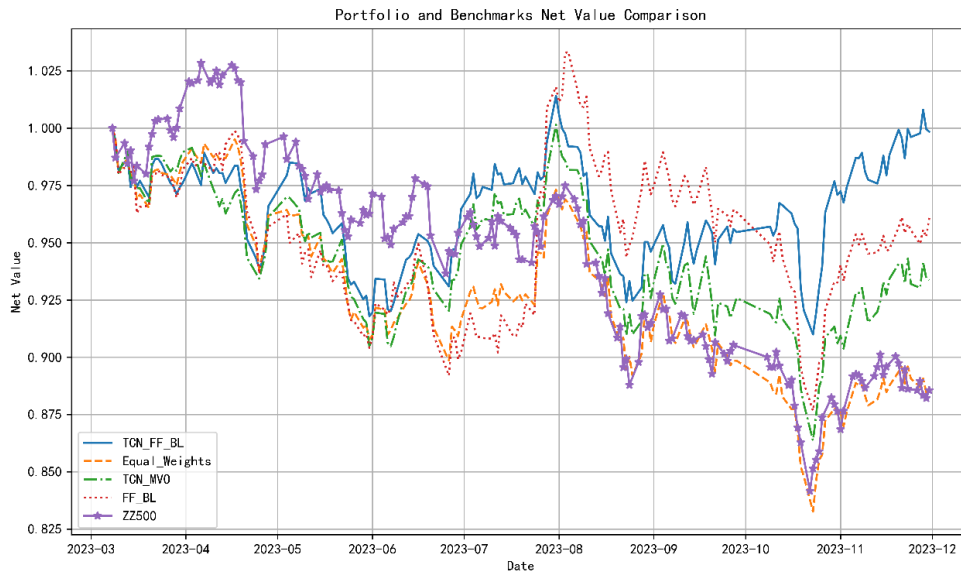




5.4 模型稳定性检验1

- **变体模型1：** 将TCN训练数据集与验证数据集的划分比例从7：3调整为6：4。

投资组合	年化收益	年化波动	最大回撤	年化夏普
TCN_FF_BL	-0.22%	13.72%	-10.26%	-0.02
Equal_Weights	-15.60%	12.93%	-16.74%	-1.21
TCN_MVO	-9.15%	14.16%	-13.77%	-0.65
FF_BL	-5.44%	16.65%	-15.16%	-0.33
ZZ500	-15.64%	13.22%	-18.16%	-1.18



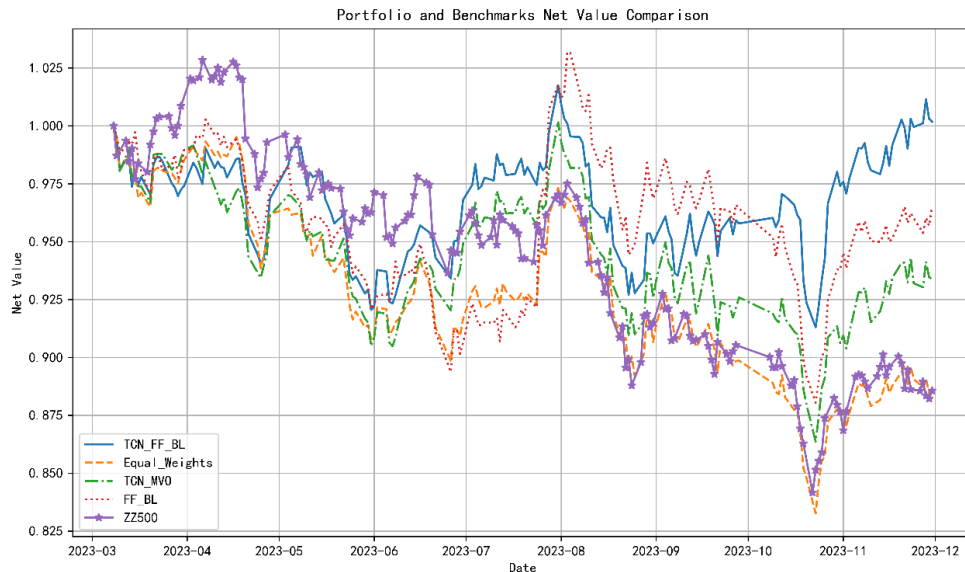
预答辩后修改



5.4 模型稳定性检验2

- 变体模型2：训练集:验证集=6:4，仅使用RiskPremium、HML和CMA三因子。

投资组合	年化收益	年化波动	最大回撤	年化夏普
TCN_FF_BL	0.25%	13.87%	-10.26%	0.02
Equal_Weights	-15.60%	12.93%	-16.74%	-1.21
TCN_MVO	-9.15%	14.16%	-13.77%	-0.65
FF_BL	-4.86%	15.83%	-14.68%	-0.31
ZZ500	-15.64%	13.22%	-18.16%	-1.18



预答辩后修改



5.4 多因子模型优化理论依据（论文P20）

- **优化思路：**Fama-French五因子模型主要针对美国股市构建，尽管国泰安CSMAR金融经济研究数据库为中国沪深A股市场计算了相应的Fama-French五因子，但是这套因子体系能否有效阐释及预测A股市场的收益率仍需检验。对各因子进行信息系数（Information Coefficient，IC）检验，以评估因子与未来收益率之间的相关性，并判定其对未来收益率预测的有效性。 **预答辩后修改**
- **测试数据：**采用2020年1月3日至2023年3月7日的12只申万一级行业指数收益率代表资产收益率，而因子时间序列的范围为2020年1月2日至2023年3月6日。
- **检验结果：**EP不显著，小市值公司因子（SMB）和盈利能力因子（RMW）亦显示出不显著的结果，因此在后续的实证分析中，可以仅考虑三个因子：市场风险溢价（RiskPremium）、高减低账面市值比（HML）和投资风格（CMA）。

Factor	EP	RiskPremium	SMB	HML	RMW	CMA
IC	-0.0436	0.7164	0.0345	-0.2709	0.0308	-0.1794
Significance	No	Yes	No	Yes	No	Yes



5.4 因子IC检验修正时间范围

```
# 读取EP因子
EP_factor = pd.read_csv('EP.csv', index_col=0)
EP_factor.index = pd.to_datetime(EP_factor.index)
EP_factor.index = EP_factor.index.strftime('%Y-%m-%d')
# 筛选时间范围
EP_factor = EP_factor.loc['2020-01-02': '2023-03-06']
EP_factor = EP_factor["EP"]

# 读取FF5因子
FF_factor = pd.read_csv('STK_MKT_FIVEFACDAY_cleaned.csv', index_col=0)
FF_factor.index = pd.to_datetime(FF_factor.index)
# 然后再将索引格式化为 '%Y-%m-%d'
FF_factor.index = FF_factor.index.strftime('%Y-%m-%d')
# 筛选时间范围
FF_factor = FF_factor.loc['2020-01-02': '2023-03-06']
```

```
def factor_ic_test(stock_price_file, factor_series):
    # 读取股票价格序列
    stock_price = pd.read_excel(stock_price_file, index_col=0)

    # 计算股票收益率
    stock_returns = stock_price.pct_change()
    stock_returns.index = stock_returns.index.strftime('%Y-%m-%d')

    # 筛选股票收益率数据范围
    stock_returns = stock_returns.loc['2020-01-03': '2023-03-07']
```

预答辩后修改

```
...: # 调用函数比较因子的IC
...: ic_comparison_table = compare_factor_ic('申万一级指数行情序列.xlsx',
EP_factor, FF_factor)
...: print(ic_comparison_table.T)
```

	EP	RiskPremium1	SMB1	HML1	RMW1	CMA1
IC	-0.043551	0.716452	0.034526	-0.270905	0.030843	-0.17939



5.4 多因子模型优化理论依据2（论文P20）

- **EP因子简介：**Liu等（2019）在研究中国A股市场股票收益率预测时指出，相比于欧美国家常用的账面市值比（Book-to-market ratio, BM），采用盈利价格比（Earnings-price ratio, EP）作为价值因子，能更准确地反映价值投资效应，具有更优的预测表现。
- **计算方法：**根据JFE论文《size and value in China》附录，计算EP [盈利价格比因子(流通市值加权)]。实操： $EP = \text{扣非净利润} / (\text{收盘价} * \text{自由流通份额})$ 。

(b) Value

i. Earnings-price ratio (EP) is calculated as

- Net profit excluding non-recurring gains and losses
(NET_PROFIT_AFTER_DED_NRLP)
divided by
- Daily close price (S_DQ_CLOSE) multiplied by total shares outstanding
(TOT_SHR)

预答辩后修改

- **测试效果：**使用2020.01.01至2023.11.30股票指数收益测试，EP因子IC仅-0.04，不如HML（-0.27）。



5.7 总结与展望

- **总结：**本研究构建了一种创新的资产配置模型，融合了TCN和多因子模型的关键思想及优点。通过深入的研究和实证分析，本研究成功提出并验证了这一全新资产配置框架，该框架在理论和实践应用上都具备显著的创新性和实用价值。

主要亮点包括：

- （1）采用TCN进行资产收益预测，构建先验分布，摒弃了市场均衡假设。
 - （2）结合TCN和多因子模型，系统化设置主观观点和置信水平，提高了模型对市场变动的适应性和稳定性。
 - （3）采用双重预测和交叉验证策略，增强了模型的可解释性，减少了过拟合风险。有效整合TCN的非线性预测和多因子模型的解释能力。
- 预答辩后修改**
- **展望：**模型稳定性检验为优化路径提供了明确的指引：（1）对TCN训练架构进行进一步的优化；（2）探索和引入更多的高质量因子。
 - 同时未来也需要进一步提高新模型在不同市场环境下的泛化能力，以及如何更加有效地、合理地结合多因子模型和TCN的优势去量化主观观点。

谢谢

请各位老师提出问题和建议

汇报人：韩江浩

2024年5月7日