

基于股票日线数据的风险因子神经网络探索

田田
计算机系
清华大学

何牧桦
未央书院
清华大学

彭扬达
计算机系
清华大学

Abstract

通过收益率协方差矩阵预测投资组合风险和基于历史数据选择股票是实现良好投资回报的两种关键方法。在风险预测方面，传统多因子模型通常依赖于经验和先验知识，但在市场环境变化时可能面临适应性问题。本研究开发了两种基于神经网络的预测模型：一种方式先利用历史股票因子暴露和量价数据来训练网络以预测未来个股收益率的协方差矩阵，随后基于协方差矩阵最大化夏普比率进行选股；另一种方式则直接以最大化组合夏普比率为目标训练网络，跳过协方差预测步骤进行选股。这些方法旨在克服传统模型在动态市场条件下的局限性，通过更精确的风险评估和投资组合优化，提高投资决策的效果。

1 绪论

近年来，对股票市场投资组合的研究使人们越来越意识到风险管理的重要性。风险管理和资产配置决策的核心是对未来风险的准确预测，良好的风险预测可以让投资者根据对收益和不确定性的估计构建投资组合、灵活调整策略。最常见的投资组合选择方法是基于马科维兹所提出的均值方差理论，即最大化投资组合的预期收益与其风险成本之差 [9]。该选择方法的正确性基于对投资组合风险的正确预测，这正是风险模型的核心目标，即准确预测某些最优投资组合的未来风险。

在实际应用中，股票的协方差矩阵往往存在维度较高的情况，导致矩阵病态或秩亏，容易产生较大误差，影响其结果的稳定性和准确性。一种解决方式是进行矩阵收缩，使非对角项元素收缩至零 [7]。然而，该方法所基于的矩阵稀疏性的假设在股票风险预测场景下可能并不适用，因为通常来说股票的收益彼此相关。另一种方法是将高维股票回报的协方差矩阵估计分解为低维风险因子回报的协方差矩阵估计，该方法基于的假设是市场上较为类似的股票可能存在相似的市场行为，推断出具有相同因子的高维的股票回报受到低维风险因子的驱动。这种方式也被称为多因子模型，其本质思想是股票收益由一些共同的因子驱动，不能被这些因子解释的部分被称为“特质收益率”，且每支股票的特质收益率之间互不相关 [10]。

尽管多因子模型已经是在实际应用中可行的模型，其估计的结果仍存在一些局限性：

估计偏差：基于历史数据的估计假设金融市场数据在不同时间步上是独立同分布的。然而，这一假设与实际金融市场情况不符，因为金融数据往往呈现出分布漂移特性，即未来数据的分布相对于历史数据会发生变化。这导致直接使用历史统计值难以准确

预测未来协方差,尤其是在最小方差投资组合中,过去相关性较高的股票在未来的相关性可能下降,这会导致对实际风险的低估和对冲风险能力的下降。

估计方差:准确估计协方差矩阵需要足够多的样本数,但现实中可用的样本往往有限,从而增加了估计误差的方差。在传统风险估计方法中,要减小估计偏差通常只能采用近期的少量样本进行估计,但这又会因样本数量不足而导致估计方差的增大,形成了估计偏差与方差之间的矛盾性。为解决以上问题,本研究提出了一种新型的神经网络风险模型。该模型利用神经网络技术模拟样本分布随时间的漂移,以尽可能减少估计偏差。同时,通过引入基于因子模型的结构,模型减少了对大量样本的依赖,从而有效减小估计方差,实现了对风险的更精准建模。这种方法为风险管理提供了一种更为可靠且数据驱动的解决方案,旨在提升投资组合风险预测的准确性和实用性。

同时,本研究提出了另一种基于神经网络优化投资的模型,即跳过因子预测,分别通过门控循环单元 (GRU)、变换器 (Transformer)、和时间卷积网络 (TCN) 三种模型直接基于历史数据进行均值-方差优化,以最大化夏普比率。本文将对两种预测方式进行比较,探讨通过神经网络预测因子协方差矩阵进行选股是否是必要的步骤。

2 相关研究

本研究灵感直接来源于参考文献部分的第一篇。该文献基于多因子模型,提出了一种通过神经网络获得风险因子的思路。多因子模型认为股票的收益由共同因子驱动,因此其可以表示为多变量线性模型。基于给定的因子暴露,对股票收益进行回归,可以得到因子收益和特异性收益,使得每个股票的收益 r_i 可以由以下线性关系描述:

$$r_i = \sum_{k=1}^K X_{ik} f_k + u_i$$

其中, X_{ik} 表示股票 i 在因子 k 的暴露, f_k 是因子 k 的收益率, u_i 是股票 i 的特异性收益。投资组合的收益 R_p 可以表示为其组成股票收益的加权和:

$$R_p = \sum_{i=1}^N w_i r_i$$

其中, w_i 是股票 i 在投资组合中的权重。投资组合的收益可以进一步表示为单个因子收益的加权形式:

$$R_p = \sum_{i=1}^N w_i r_i = \sum_{i=1}^N w_i \left(\sum_{k=1}^K X_{ik} f_k + u_i \right) = \sum_{k=1}^K \left(\sum_{i=1}^N w_i X_{ik} \right) f_k + \sum_{i=1}^N w_i u_i$$

考虑将整个投资组合在风险因子 k 上的暴露程度表示为:

$$X_p^k = \sum_{i=1}^N w_i X_{ik}$$

我们可以进一步简化投资组合收益的表达式:

$$R_p = \sum_{k=1}^K X_p^k f_k + \sum_{i=1}^N w_i u_i$$

另外,由于单个因子的收益与特质收益率互不相关,且不同股票的特质收益率之间也互不相关,投资组合的预期收益和方差可以通过因子暴露和特异性风险的统计特性来计算:

$$\text{var}(R_p) = \sum_{k,l} X_p^k F_{kl} X_p^l + \sum_{i=1}^N w_i^2 \text{var}(u_i)$$

其中, F_{kl} 是因子 k 和 l 之间的因子协方差矩阵。综上, 多因子模型将 N 只股票的收益分析转换为对 K 个因子的收益分析, 从而将个股协方差矩阵分解为一个低秩矩阵与一个对角矩阵之和。在实际中, 股票数量 N 要远远大于共同因子数量 K , 因此模型起到降维作用, 减少分析难度的同时提高预测准确度。此外, 多因子模型通过多种方式来调整因子协方差和特异性方差的估计值, 以减小估计偏差问题, 实现对于最优组合风险的准确预测。因子的合理选择是构建多因子分析的关键, 目前主流的因子分类有: 价值因子、成长因子、盈利因子、规模因子、动量因子等等。基本风险因子因其可解释性和有效性首先被提出, 资本资产定价模型中将其定义为系统性风险因子 [11]。在此之后, 多个因子相继被提出, 例如规模、价值、动量等 [3, 5]。这些著名的因子均是由专家基于大量历史数据和学术研究设计出来的, 所需工作量极大。尽管如此, 这些因子仍然在预测市场表现上体现出局限性。考虑到人工设计因子带来的局限性, 统计风险因子这一新的因子生成方法被提出。这通常涉及在历史股票回报上应用主成分分析或因子分析 [1, 6]。与人工因子不同的是, 这种方法在效率上较为突出, 然而其结果在训练集上表现较好, 测试集中表现往往较差, 原因是获得的风险因子往往过度拟合了历史回报。在当前两种方法都不够有效的情况下, 我们计划通过一种更有效的方式来设计风险因子。

本文在均值-方差优化中选择采用三种模型进行训练, 包括 GRU、Transformer 以及 TCN。

GRU 基于循环神经网络, 结构中包含重置门和更新门, 与长短期记忆网络不同的是采用隐藏状态而不是单元状态来传输信息, 参数减少的同时效率更高 [4]。GRU 的重置门主要负责计算参数 $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$ 和隐藏状态 $\hat{h}_t = \sigma(W_o \cdot [r_t \circ h_{t-1}, x_t] + b_o)$, 有助于捕捉时间序列中的短期依赖关系; 更新门主要用于计算参数 $z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$, 以及隐藏状态 $h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \hat{h}_t$, 有助于捕捉时间序列中的长期依赖关系。

Transformer 模型的核心是注意力机制, 即查询 (Q)、键 (K) 和值 (V) 三者之间的交互: $Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V$, 其计算效率高, 能够通过矩阵运算并行处理, 这种机制较为适合处理长距离依赖问题, 因为它可以直接计算序列中任意两个位置之间的依赖 [12]。

TCN 针对时间序列数据的特点符合本研究中的研究背景。TCN 通过卷积层而非循环层来处理序列依赖关系 [2]。对于一个输入序列 x_0, x_1, \dots, x_T , 我们希望预测相应的输出: $y_0, y_1, \dots, y_T = f(x_0, x_1, \dots, x_T)$, 其中的因果约束是只能使用来自当前时间或更早时间的输入。序列建模的目标是找到一个网络模型 f 来最小化标签输出和预测之间的损失:

$$L(y_0, y_1, \dots, y_T, f(x_0, x_1, \dots, x_T))$$

此外, TCN 同时具有膨胀卷积的特性, 即在不增加参数量或计算复杂度的情况下使网络能够捕捉长范围的序列依赖关系。

3 研究主题

3.1 因子协方差矩阵预测

考虑到多因子模型中用来调整估计值的方法都是手工设计、依赖于先验知识的, 当市场环境发生改变时, 模型不一定适用当前市场。因此, 我们计划利用深度学习得到一个基于神经网络的风险模型, 它可以通过数据驱动的方式来预测因子协方差与特异性风险, 从而不需要基于专家经验, 也能实现风险建模和预测。

对于基于神经网络的因子协方差或特异性方差的预测，核心点在于设计一个数据驱动的方式使得预测的因子协方差以及特异性方差能够朝着正确的方向优化。一种直接的思路是计算预测值与真实值之间的差别，并以此作为损失函数进行优化。然而这种方法面临着两种困难。

首先，真实的未来协方差矩阵需要未来的样本数据进行估计，虽然因子协方差和特异性方差需要估计的元素数量相较个股协方差减少了很多，但是同样面临着样本数量选择的问题。假设估计区间为 $[T, T + \delta t]$ ，如果估计样本时间区间 δt 较大，那么虽然得到的因子协方差和特异性方差的估计的误差较小，但是由于因子协方差和特异性方差的是随时间变化的，因此估计值仅能代表 $[T, T + \delta t]$ 之间的真实值，而不是我们想要预测的 $T + 1$ 时间段，即得到的估计值的在时间维度的偏差较大。若是选择较小的 δt 时间区域，那么估计值本身的误差，即估计值的方差就会较大。因此该思路存在矛盾性，并不可行。

其次，假设我们可以得到真实的因子协方差和特异性方差，在考虑损失函数时，一般的损失函数，例如计算预测矩阵与真实矩阵元素之间的均方误差并不一定适用于风险模型，因此我们需要设计一种更符合风险模型的损失函数。

3.2 均值-方差优化

考虑到因子协方差预测的复杂性，我们期望尝试直接进行选股策略的预测来对比两种方法的效果，并判断因子协方差矩阵的预测是否是必须的。

在实现细节方面，我们综合比较了 GRU、Transformer、和 TCN 的模型架构，最终选定 GRU 因其在时间序列预测中的优越表现。此外，模型中还包括了一层全连接层来进行最后的输出，采用最小化负夏普比率的损失函数，确保了优化目标的一致性和直接性。模型通过 SAM 优化器改善模型在训练过程中的泛化能力，减少过拟合的风险。在模型训练过程中可以通过设置权重的上下限来满足实际投资中的各种约束条件，以此来更好地应用于现实中的风险管理和资产配置情景。

4 实验设计

4.1 因子协方差矩阵模型预测 A 股

在本研究中，我们计划使用中国 A 股市场过去十年（2012 - 2022 年）每个交易日的数据来训练我们的模型，该模型基于过去 63 天的历史数据来预测未来 21 天的股票收益率。为了获取历史数据，我们选择了 qlib 作为数据源。考虑到 qlib 数据的一些局限性，如复权方式的不一致，我们决定排除所有包含复权信息的股票。此外，我们还统一了时间跨度，并清除了部分非法数据，最终得到 639 只股票的有效数据。对于数据中的缺失值，我们采用了前一天的数据进行填充，以保持数据的连续性。

在模型的实现上，我们搭建了包含多个 GRU 层和一个全连接层的 RiskFactorExtraction 模型，旨在从历史数据中提取风险因子。模型的输入数据由两个五维的 tensor 组成，分别对应股票的历史数据和未来收益率的预测。具体来说，股票历史数据的 tensor 结构为 `torch.Size([1035, 2, 63, 639, 9])`，解释如下：1035 个 batch；每个 batch 包含 2 个元素，每个元素代表历史数据中的某一天；63 代表该天前的 63 天数据；639 是指参与分析的股票数量；9 则是每只股票的九个特征，包括最高价、最低价、股息、市盈率、收盘价、开盘价、交易量、调整收盘价和成交金额。股票未来收益率的 tensor 结构为 `torch.Size([1035, 2, 1, 639, 2])`，其维度含义与历史数据基本相同，其中第三维度代表我们预测未来某一天（本研究中为 21 天后）的收益率，第五维度表明我们存储了两次 21 天后的预测收益率，以便训练模型时重点关注两个特定的风险因子。训练得到的模型输出为一个三维 tensor，预测未来某一天每只股票的风险因

子值。我们根据论文中的数学公式，使用因子的协方差矩阵进行变换，主要通过正则化方法实现降维，从而得到股票的协方差矩阵。接着，基于这些预测的未来协方差矩阵以及当前所有股票的收益率，我们运用投资学中的 Markowitz 模型构建了旨在最大化夏普比率的投资组合。此外，我们实施了动态学习率调整的训练策略，总共进行了 60 轮的训练，并按照 8:2 的比例划分训练集和验证集。考虑到时间序列数据的连续性特点，为了在可视化和分析过程中获得更直观的理解，我们选择不打乱数据顺序。这种方法不仅保持了数据的时序关系，也有助于更精确地评估模型对时间依赖性的捕捉能力。

最终使用的超参数如下， γ 代表损失函数对方差和期望的权重值：

$$num_layers = 4, dropout = 0.5, lr = 0.0005, \gamma = 0.01$$

在训练中的 loss 值如图 1 所示，在大约 25 轮训练之后基本就保持了稳定的收敛。

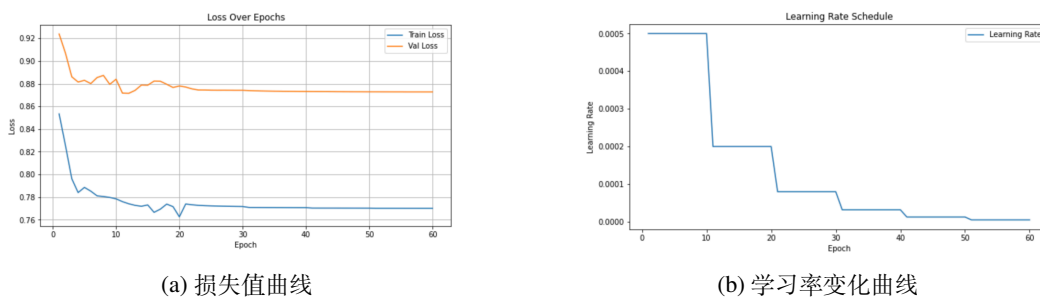


图 1: 因子协方差矩阵预测模型训练结果

在验证集上，我们使用训练好的模型预测协方差矩阵并构建投资组合，组合的每日收益率和累计收益率如图 2 所示。

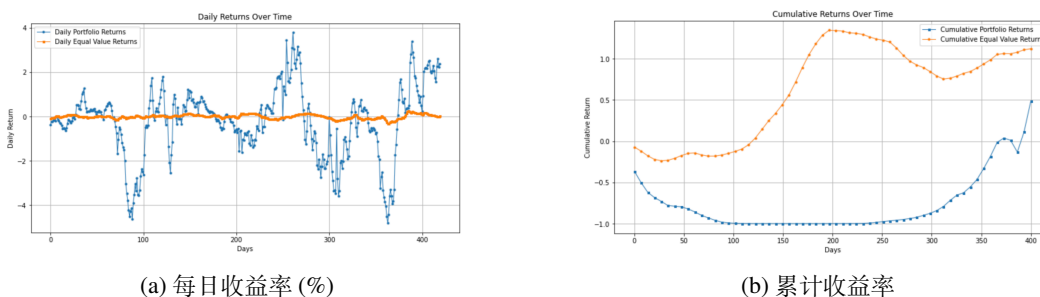


图 2: 因子协方差矩阵预测模型结果

从实验结果可以看出，与等权重投资组合相比，我们的策略收益率表现并不理想，且伴随有更大的波动和回撤。我们认为这一结果由两个主要因素造成。

首先，我们的方法通过使用股票的历史数据来预测未来的风险因子，并进一步利用这些因子来估计未来的协方差矩阵，最后依据这些协方差矩阵构建投资策略。这种方法虽然理论上可行，但在实际操作中，直接使用历史数据调整投资组合的权重可能更为直接和有效。

其次，我们使用的是通过 qlib 获取的日线数据来提取风险信息，而原论文使用的则基于 Barra 风格因子来提取风险信息。相比较而言，日线数据噪声太大且 qlib 数据质量不高，原论文基于经验因子的基础上用神经网络做进一步提取或许更为有效。此外，我们的风险因

子预测网络大体借鉴的是原论文的网络结构，虽然尽量调整了参数，但因为数据源头的不一致，所以可能还是不适配。

4.2 模型切换

因此，我们决定针对这两个问题作出调整。在模型结构方面，我们计划增加一组新模型，该模型将直接利用股票的历史数据来实施均值-方差选股策略，选择预测中夏普比率最高的投资组合。我们将使用这一新模型的结果与我们按照论文实现的模型结果进行对比分析。在数据来源方面，我们决定转而使用更为稳定的美股数据，具体为标普 500 指数中表现较好的 50 支股票，覆盖 2001 年到 2020 年这 20 年的日线数据，数据通过 `yfinance` 获取。在计算夏普比率时，我们选定的无风险利率为 0.02。

基于这些调整，我们新增了两组实验。第一组实验使用新选定的美股数据来训练之前的因子预测协方差模型，旨在检验模型在不同数据类别下的表现，并作为对照组。第二组实验作为实验组，将使用我们新构建的直接预测夏普比率的模型，在美股数据上进行训练并进行效果检验，与第一组实验进行对比分析。由于我们缺乏在此类预测中使用 RNN 模型的经验，我们分别对 GRU、TCN 和 Transformer 这三个适用于处理序列数据的模型进行了调参训练。在比较各模型的表现后，我们选择了表现最佳的 GRU 模型作为最终结果。

新获取的美股历史数据包括六个特征：开盘价、收盘价、最高价、最低价、复权后的收盘价和交易量。

4.3 因子协方差矩阵模型预测 A 股票

对于原来的模型，我们并不改变训练框架和数据格式，输入给模型的历史数据的格式为：`torch.Size([2547, 2, 63, 50, 6])`，输入给模型的未来收益率数据格式为：`torch.Size([2547, 2, 1, 50, 2])`，具体含义和之前的对应格式相同，这里不再赘述。考虑到数据更长的时间跨度，我们增加了模型的复杂度，最终选择的超参数如下：

$$hidden_size = 64, num_layers = 6, dropout = 0.5, lr = 0.0001, gamma = 0.01$$

我们共训练了 60 轮，根据图 3，从损失值的曲线中可以看出，大约在 35 轮后验证集上的损失值不再提升。我们使用从 2017 年 4 月 11 日到 2020 年 2 月 8 日的数据作为验证集，在验

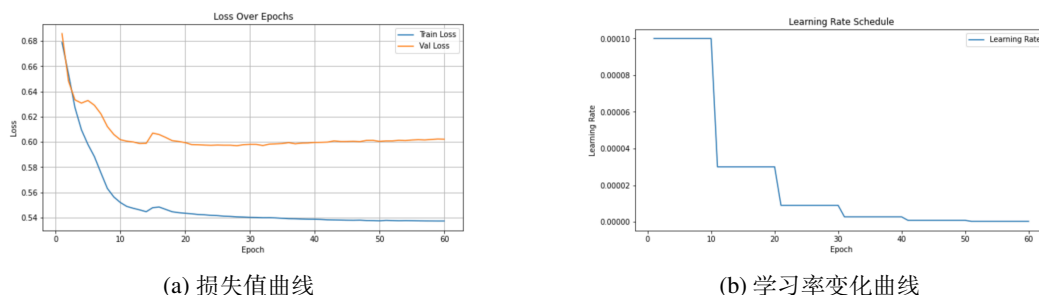


图 3: 因子协方差矩阵预测模型训练结果 (美股)

证集上，模型预测的协方差矩阵构建的策略的表现如图 4 所示。

蓝线是这 50 支股票等权重构成的投资组合，橙色的线是根据我们每天预测出来的协方差矩阵动态调整得到的投资组合。可以发现，我们构建的投资组合的表现并不好，甚至没有超过

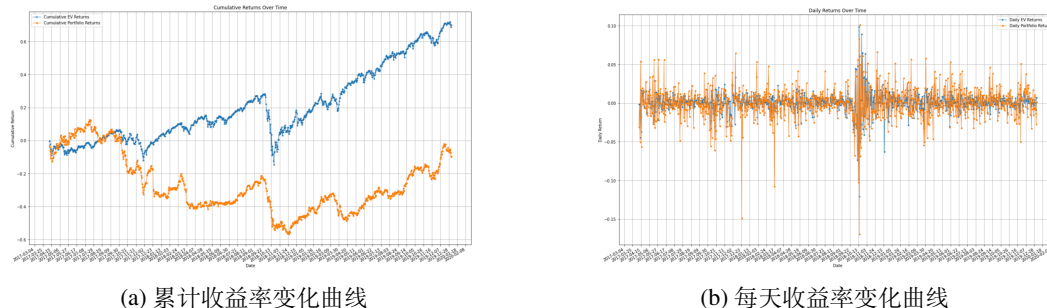


图 4: 因子协方差矩阵预测模型投资组合结果

整体的指数, 并且也有着更大的回撤。从每天的收益率变化曲线中也可以看出, 我们的策略相比指数来讲有着明显更加剧烈的波动, 整体的风险很大。

在更换了数据之后, 因子预测协方差矩阵预测模型的表现依然不好, 说明并不是数据的问题, 而是该模型的方式不够直接。

193

194

195 4.4 基于夏普比率的 GRU 选股模型

因此, 我们采用最大化夏普比率的方式直接预测投资组合。为了最大化夏普比率, 我们在训练过程中采用夏普比率的相反数作为损失值, 目标是寻找具有最高夏普比率的投资组合比重。在咨询了具有深度学习经验的同行后, 我们选择使用 SAM 优化器来进行模型训练。首先, 我们分别对 GRU、TCN 和 Transformer 这三种模型进行了参数调整, 并详细观察了它们的性能表现, 以便选出表现最优的模型。这一参数调整过程的详细记录已包含在附件中。经过比较, 我们发现 GRU 模型在夏普比率方面表现最佳, 因此我们选择它作为我们的最终模型。在最终的训练中, 我们使用了 dataloader 来批量装填数据, 设置不打乱数据保留时间序

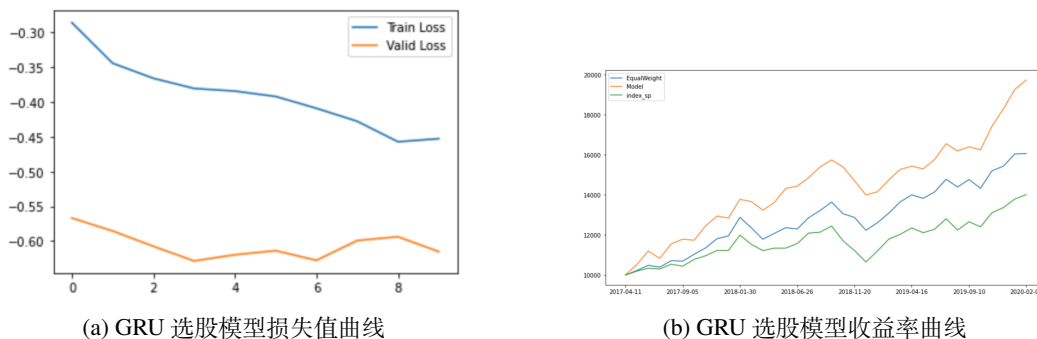


图 5: GRU 选股模型结果

202

列的顺序关系, 对于最后不足一个 batch 的数据直接删除, 最终采用的超参数如下:

$EPOCHS = 10, EARLY_STOP = 20, LR = 0.025, MOMENTUM = 0.95,$

204

$N_LAYER = 3, HIDDEN_DIM = 128, DROPOUT = 0.3$

我们总共训练了 10 轮, 如图 5 所示, 从损失值的变化曲线中可以看出, 在 5 轮之后基本上验证集的表现就不提升了, 收敛速度较快。

使用同样的均值-方差策略进行测试, 采用业界更加常用的、设定初始资产为 10000 点的模

208 拟交易回测框架衡量策略表现，验证集的时间跨度同样是从 2017 年 4 月 11 日到 2020 年 2
209 月 8 日，每 21 天调整一次仓位，策略表现如图 6 所示。

210 可以看出，策略有着比较好的表现，优于等权重的股票和标普 500 指数（因为这 50 支股票
211 本身就是选择的标普 500 中表现比较好的），同时也没有出现比较大的回撤。并且，策略的
212 收益率曲线、等权重组合的曲线和标普 500 指数的变化整体趋势比较接近，说明无论如何调
213 整仓位，还是比较难规避掉市场的影响。同时，我们增加了调仓的限制，规定每次调整单支
214 股票的变化范围不得大于整体资金池的 10%，整体策略比较保守。如图 6 所示，在这段时
215 间中，我们的投资组合中各支股票的仓位变化如下，可以发现整体上占比的变化并不剧烈：
相比于对照组的实验，我们使用了相同的策略，使用了相同的数据，调整了预测方式后就

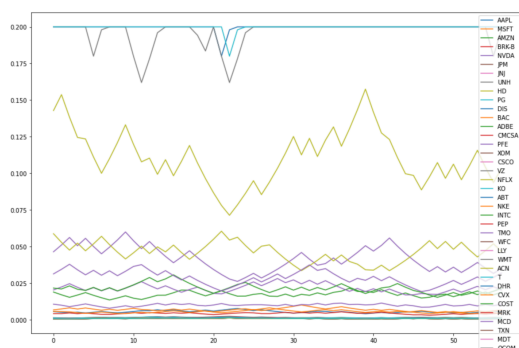


图 6: GRU 选股模型测试结果股票仓位变化

216 得到了比较大的性能提升，说明用历史数据去预测协方差矩阵的方式有些过于不直接，对
217 模型预测能力的影响很大。
218

219 5 总结

222 经过我们的探索，基于股票日线数据预测协方差矩阵的策略并不那么有吸引力，在我们的测
223 试中，借助 GRU 神经网络跳过风险因子预测步骤直接以最大化夏普比率为目标的选股策略
224 在表现上优于风险因子预测策略。或许在这个问题上，减少中间的变量数，直接关注最终变
225 量更加有效。

226 当然，或许是我们对风险因子策略的实现并不完美，我们还可以从数据处理和模型设计两个
227 部分来进一步进行优化提升。

228 从数据处理的角度，我们期望从 A 股获取和提炼出更加高质量的数据。我们认为现在模型
229 训练效果在 A 股上的表现不佳与我们使用的数据质量不高有着较大的关系。这也是我们在
230 对 qlib 获取的原始数据集进行蒸馏时所认识到的，比如复权、拆股等信息并没有在 qlib 股
231 票数据中得到妥善处理。此外我们还对最终股票集合做了若干人为的遴选，我们只留下了具
232 有指定时段、指定特征的股票族，或许这种丢弃并不是合理的。比如，我们可以利用我们丢
233 弃的拆股信息，在那些因拆股而未使用的股票数据上再训练一个模型，对比两个模型的性能
234 来判断我们的模型效果是否可以因拆股特征的加入而有所变化。此外，可以尝试切换数据时
235 段，观察我们的模型是否会在特定市场风格下有着更好的表现。总而言之，在得到或者筛选
236 出更高质量的数据后，我们可以尝试在新的数据集上训练这两个模型，观察模型训练效果是
237 否会发生改变。

从模型设计的角度，风险因子模型应有着提升的空间，我们可以参考一些学界提出的前沿同类技术，例如使用 AutoEncoder、VAE 等等，尝试通过新技术的引入增加整体因子的质量，提高其预测的准确性和稳定性。对于直接通过优化夏普比率调仓的模型。

参考文献

- [1] Lin, H., Zhou, D., Liu, W., & Bian, J. (2021, November). Deep risk model: a deep learning solution for mining latent risk factors to improve covariance matrix estimation. In *Proceedings of the Second ACM International Conference on AI in Finance* (pp. 1-8).
- [2] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- [3] Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), 57–82. <https://doi.org/10.2307/2329556>
- [4] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [5] Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427–465. <https://doi.org/10.2307/2329112>
- [6] Harman, H. H. (1976). *Modern factor analysis*. University of Chicago press.
- [7] Ledoit, O., & Wolf, M. (2017). Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies*, 30(12), 4349–4388. <https://www.jstor.org/stable/48616725>
- [8] Avellaneda, M., & Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7), 761–782.
- [9] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.2307/2975974>
- [10] Rosenberg, B. (1974). Extra-Market Components of Covariance in Security Returns. *Journal of Financial and Quantitative Analysis*, 9(2), 263–274. doi:10.2307/2330104
- [11] Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3), 425–442. <https://doi.org/10.2307/2977928>
- [12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

A 基于协方差矩阵的预测模型框架数学原理

线性的多因子风险模型形式可以表示为：

$$V = XF X^T + \Delta$$

其中 V 表示个股协方差， F 表示因子协方差， Δ 表示特异性方差。我们将对于 V 的预测拆分为对于 F 和 Δ 的预测，具体如下：

1. 数据准备：输入历史先验因子暴露度 X_{prior} 、量价数据 S 、未来个股收益 r 。

273 2. 隐式因子：通过多层感知机将先验因子映射为隐式因子 X 。该步骤通过对先验因子的组
274 合映射起到降维的作用，并且要求对于学习到的隐式因子最好是相互正交的。

275 3. 线性回归：基于因子暴露，对股票收益进行线性回归：

$$r = Xf + u$$

276 考虑到数据的异方差性，回归方式采用市值加权的最小二乘法：

$$f = (X^T W X)^{-1} X^T W r$$

277 得到因子组合权重 W_f 、纯因子收益 f 、以及特异性收益 u 。

278 4. 特征提取：利用循环神经网络等方式，从量价序列数据中提取出个股特征 e_s 。

279 5. 因子协方差风险预测：

280 a. 利用纯因子组合权重对个股特征进行聚合操作，得到纯因子特征：

$$e_f = W_f \cdot e$$

281 b. 计算纯因子特征的内积，将其作为因子协方差矩阵（为了满足协方差矩阵的半正定
282 性质，此处采用 Gram 矩阵的方式进行构造）：

$$F_{ij} = e_{f_i}^T \cdot e_{f_j}$$

283 c. 对因子协方差矩阵做特征值分解（可以采用反向传播的方式）：

$$F = U U^T$$

284 d. 根据得到的特征值和特征向量，计算白化矩阵

$$P = U^{-\frac{1}{2}} U^T$$

285 然后对纯因子收益做 ZCA 正则化：

$$f' = U^{-\frac{1}{2}} U^T f$$

286 ZCA 正则化的目的是将纯因子收益转变为正则因子收益，其协方差应该是单
287 位对角矩阵。此处不能使用 PCA 正则化，否则得到的基底会随着因子协方差矩
288 阵的变化而变动，计算损失值时无法在一个批量上直接计算协方差。

289 e. 计算正则因子收益的协方差矩阵 $F' = f' \cdot f'^T$ 与单位对角矩阵之间的差别，将其作
290 为因子协方差风险预测的损失值。此处损失函数的设计应当保证 F' 的对角元
291 素接近 1，并且让非对角元素趋于 0。

292 一种设计是保证一个批量中的样本是时间连续的，这样可以得到连续多天的特
293 征组合收益 f' ，并以此计算 F' ；在此之上分别计算对角元素与非对角元素的损
294 失值，对角元素采用 $Q = f_i'^2 - \ln(f_i'^2)$ 作为损失函数来对低估风险施加更大的
295 惩罚，对非对角元素采用 $\frac{1}{k(k-1)} \sum_{i \neq j} |f_i' f_j'|$ 作为损失函数来使非对角线趋于 0。

296 f. 还可以进行一些改进使得因子协方差矩阵的真值非对角线元素是 0：

297 i. 在 4. 特征提取的步骤中，可以加入 Deep Risk Model(DRM) 的注意力机制来
298 降低不同股票间的协方差，以此来发现类似股票的不同性，例如一个行
299 业中的某些股票可能主要被市值因子影响，另一些可能被负债率影响。

300 ii. 计算因子协方差矩阵时，可以采用

$$F_{ij} = k(e_i, e_j) = \exp(\|e_i - e_j\|_2)$$

301 来得到非线性关系，以此来降低非线性关系产生的协方差。此方法可能
302 使指数梯度函数衰减较快，导致反向传播计算的效率较低以及优化较慢。

303 6. 特异性风险的预测：

304 a. 计算个股特征与因子特征的残差，以此作为特异性特征：

$$e_u = e_s - W_f^T \cdot e_f$$

305 b. 根据特异性特征，通过内积，多层感知机等方式计算出特异性方差 Δ_{ii} 。

306 c. 计算特异性收益 u 的协方差矩阵与预测的特异性方差 Δ 的差别，将其作为特异性
307 风险预测的损失值。由于 Δ 的维度较高，可能难以优化，因此只计算对角元素
308 的 Q 统计量作为损失值。

309 7. 个股协方差预测：

$$V = XFX^T + \Delta$$

310 此处两个半正定矩阵之和仍为半正定矩阵，因此满足协方差半正定性质的要求。

311

312