

HRP-FC/GMSC: 一种使用基于相关性的最小生成树进行选股并采用分层风险平价模型进行资产配置的策略¹

--一篇论文复现

李思远²

摘要

2016 年提出的分层风险平价算法 (Hierarchical Risk Parity, HRP) 结合了资产组合的相关系数矩阵的层次结构来考虑资产组合内权重分配的问题, 许多学者回测检验后认为其在风险的处理上有良好的表现. Cho(2023)提出了一种采用基于减去全局运动的相关性矩阵的最小生成树进行选股并采用分层风险平价算法为每个资产赋予权重的股票交易策略, 并发现其有不错的表现. 本研究复刻了这一策略, 并增加考虑了交易费用. 由于具体参数和实验方法上存在的不确定性, 我们采用简单的最小生成树的度替代了原文中的混合排名策略, 并实现了类似的效果. 我们发现在 0.02 无风险利率下, 这个策略的夏普比率并不显著优于基准算法. 此外, 我们认为为相关性矩阵减去全局运动的做法在组内资产数较小 (5~20) 的组合上可能能够有效提升策略表现.

¹关于 HRP 算法的实现部分复用了笔者的另一项研究(资产组合层次结构与 HRP 算法表现)的代码.

² 李思远是中山大学 20 级金融学本科生, 香港科技大学(广州)24 级金融科技方向硕士研究生(Mphil). 他的研究方向包括资产配置算法等. lisiyuansven@foxmail.com

目录

摘要	1
1. 项目目录及文件的说明.....	3
1.1 项目概览	3
1.2 项目文件导览.....	3
2. 涉及的模型/工具/算法.....	5
2.1 相关系数矩阵及其全局运动.....	5
2.2 基于最小生成树的选股策略.....	6
2.3 分层风险平价算法	8
2.4 基准算法	9
2.5 整个策略的流程	9
3. 数据和回测	11
3.1 数据	11
3.2 回测方法	11
3.3 回测表现	13
3.3.1 采用原始相关系数矩阵进行聚类.....	13
3.3.2 采用 FC/GMSC 处理的相关系数矩阵进行聚	15
结论	19
参考文献.....	20

1. 项目目录及文件的说明

1.1 项目概览

本文复现了 Cho(2023)提出的一种交易策略。我们将此策略称为 HRP-FC/GMSC，其流程是，首先基于若干资产日收益率的时间序列数据生成其相关系数矩阵，将该矩阵标准化处理并减去全局运动，基于处理后的矩阵绘制最小生成树，取最外围的 N 个资产构建资产组合，并以 HRP 算法回测生成其权重。最后评估组合的表现。

进一步地，考虑到文献中的回测窗口固定在 120 个交易日，我们尝试检验该策略在不同回测窗口下的表现情况。我们计划在后续研究中探究这一部分。

1.2 项目文件导览

本项目的根目录为 HRP-MST，但代码的工作路径设在 HRP-MST/Scripts。在 HRP-MST 目录下，Reference 路径放置了本研究的核心参考文献；WashedData 路径放置了用于回测的数据；Scripts 路径放置了核心的代码和运行输出的结果。

特别地，HRP-MST/Scripts/demo_single_portfolio.ipynb 是对原文基于相关性的 MST 进行选股的一个基本演示，包含了中间数据的可视化等内容。同目录下 original_method_paper.ipynb 是本文的核心代码，包含了所含资产数不同&采用不同相关系数矩阵处理方式的 18 种组合的回测过程与结果。每个回测的表现在 HRP-MST/Scripts/original_alpha/performance 路径下，对于所有结果整体的评估在 HRP-MST\Scripts\原文策略复现评估.xlsx 文件中。

进一步地，我们尝试以更多回测窗口长度和资产数量的组合进行回测。考虑到这一部分研究的运行时间较长，代码均采用多进程进行加速。首先采用 multi_select_stock.py 进行选股，结果保存在 HRP-MST/Scripts/selected_stocks 文件夹下。接着，以 multi_backtest.py 回测这些保存的组合，回测结果在 HRP-MST/Scripts/multi_backtest_result 路径下(这一部分工作还没有完成，我们计划后续推进)。temp.ipynb 检验了这些结果，并保存在 all_result.csv 中。

图 1.1 展示了项目的目录层级和文件大小.

文件夹	父级百分比	大小	分配	项目	文件	文件夹
D:\ProjectNeedSpeed\0Projects\	100.0 %	2.4 GB	2.4 GB	383	373	10
Scripts	98.7 %	2.3 GB	2.3 GB	366	361	5
multi_backtest_result	68.5 %	1.6 GB	1.6 GB	232	232	0
original_alpha	29.5 %	704.7 MB	704.7 MB	42	41	1
[13 文件在 D:\ProjectNeedSp	1.9 %	46.2 MB	46.2 MB	17	13	4
selected_stocks	0.0 %	45.0 KB	144.0 KB	72	72	0
__pycache__	0.0 %	45.6 KB	52.0 KB	3	3	0
WashedData	1.0 %	24.9 MB	24.9 MB	3	3	0
Reference	0.2 %	4.0 MB	4.1 MB	2	2	0
说明文档&复现报告.docx	0.1 %	2.0 MB	2.0 MB			
.history	0.0 %	2.6 KB	8.0 KB	6	5	1
~\$文档&复现报告.docx	0.0 %	162 字节	0			

图 1.1 项目文件概览

另外, 作为整个项目的基础, `framework_portfolio_weight_performace.py` 提供了 HRP 算法, 基准算法 IV, 相关系数矩阵标准化和去除全局运动以及投资组合表现评估的实现. 这些代码的一部分(HRP 算法, IV 算法以及部分表现评估的代码)是我在毕业论文中使用的, 其余的是为本研究而完成的. `pypfopt_temp.py` 提供了作为基准算法的均值方差优化的实现, 这些代码来自于第三方库 `PyPortfolioOpt`³, 由于版本兼容性问题, 我对它们进行了删改和移植, 仅保留了本研究会用到的代码段.

在本次复现研究中, 项目地址中有几个程序并没有被用到. `multi_select_stock.py` 是我尝试的在不同窗口和组内证券数 N 下进行排列组合并进行选股的程序, 其选股结果储存在 `selected_stocks`, 并以 `multi_select_stock.py` 进行回测. 这些程序均采用了多进程的方式以加快选股, 回测速度. 我们认为关于窗口和组内证券数的排列组合的研究仍然需要打磨, 因此没有在报告中展开论述, 但我们保留了这些代码供您阅览.

³ `PyPortfolioOpt` 库的官方文档: <https://pyportfolioopt.readthedocs.io/en/latest/>

2. 涉及的模型/工具/算法

2.1 相关系数矩阵及其全局运动

以下内容是对 Cho(2023)原文的汉化移植.

设 $P_i(t)$ 为时间 t 时第 i 资产的收盘价；那么对数回报率 $R_i(t)$ 为 $\ln P_i(t) - \ln P_i(t-1)$ 。设 $\mu(t)$ 和 $\sigma(t)$ 分别为移动窗口长度 T 内对数回报率的算术平均值和标准差；那么标准化回报率 $\tilde{R}_i(t)$ 为 $(R_i(t) - \mu(t))/\sigma(t)$ 。

由 N 个资产的标准化对数回报向量为 $\tilde{\mathbf{R}}_{t,l} = [\tilde{R}_l(t-T+1), \dots, \tilde{R}_l(t)]$ ，其中 $i = 1, 2, \dots, N$ ，则时间 t 的 FC（相关性）可定义为：

$$\rho_{t,ij} = (1/T)(\tilde{\mathbf{R}}_{t,l} \cdot \tilde{\mathbf{R}}_{t,j}) - (1/T^2)(\tilde{\mathbf{R}}_{t,l} \cdot \mathbf{1}_T)(\tilde{\mathbf{R}}_{t,j} \cdot \mathbf{1}_T)$$

其中 $\mathbf{1}_T$ 表示长度为时间窗 T 的全 1 向量。记 $\mathbf{x} \cdot \mathbf{y}$ 为两个向量的内积。

设 Σ 为任一移动窗口的 FC 矩阵，其元素为上述等式（2）中的 ρ_{ij} 。首先，Marchenko–Pastur 分布描述了随机矩阵的特征值分布。根据该分布，随机矩阵的特征值有一个理论上限。因此，实证相关性矩阵的特征值中，超出上限的对应特征向量可以认为是非随机互动的结果。从这个角度出发，实证 FC 矩阵的特征值概率密度函数可以分为三部分：

$$\Sigma = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \sum_{i=2}^{N_g} \lambda_i \mathbf{u}_i \mathbf{u}_i^T + \sum_{i=N_g+1}^N \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

其中 λ_i 和 \mathbf{u}_i 分别表示 Σ 的按降序排列的特征值列表和相应的特征向量， N_g 表示超过理论特征值分布上界的特征值数量。理论上界和 N_g 随着移动窗口的滑动而变化。特别地，第一部分 $\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$ 可被视为全球运动。全球运动指的是整个市场对外部信息的集体反应。

最后，GMSC（全球运动剔除矩阵）矩阵 $\tilde{\Sigma}$ 的定义如下：

$$\tilde{\Sigma} = \Sigma - \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$$

理论上，GMSC 矩阵专注于通过排除因共同外部信息引起的市场风险，来研究资产之间特有风险的内部互动。

图 2.1 展示了对相关系数矩阵做的一系列处理.

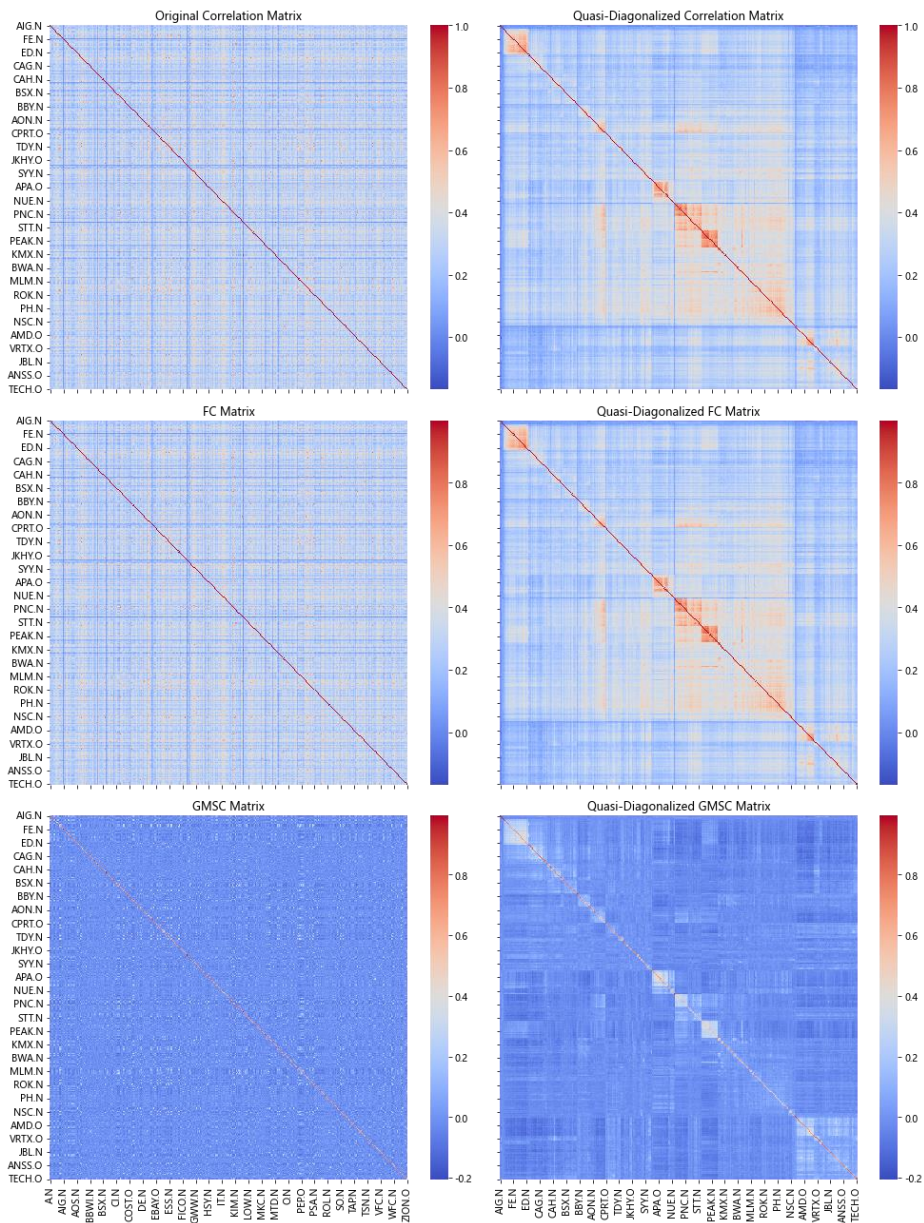


图 2.1 相关系数矩阵相关处理⁴

2.2 基于最小生成树的选股策略

关于采用最小生成树进行选股,或者说采用混合边缘性度量方法减少组合中资产的数量的做法在原文的附录 C 中. 此处仅做定性陈述. 这个方法提供了一种

⁴ 过程见 demo_single_portfolio.ipynb

混合度量值, 该值越高则节点约边缘. 边缘的资产拥有更优秀的分散风险能力, 因此可以尝试选择最边缘的 N 个资产构建组合. 式 2-1 展示了这一评分的机制.

$$\begin{aligned}
 X &= \frac{(R_D^w + R_D^{uw} + R_{BC}^w + R_{BC}^{uw} - 4)}{4} \times (N - 1) \\
 Y &= \frac{(R_E^w + R_E^{uw} + R_C^w + R_C^{uw} + R_{EC}^w + R_{EC}^{uw} - 6)}{6} \times (N - 1) \\
 P &= X + Y
 \end{aligned} \tag{2-1}$$

其中, R_D^w 表示加权的度中心性; R_D^{uw} 表示非加权的度中心性; R_{BC}^w 和 R_{BC}^{uw} 分别代表加权和加权的介数中心性; R_C^w 和 R_C^{uw} 代表加权和加权的接近中心性; R_E^w 和 R_E^{uw} 代表加权和加权的偏心率; R_{EC}^w 和 R_{EC}^{uw} 代表加权和加权的特征向量中心性. 更详细的介绍可以参照 Pozzi(2013).

需要说明的是, 我们按照论文中的方法复现的选股方式得到的结果与论文中展示的图片并不一致. 我们得到的节点分布更为集中. 我们编写的 `demo_single_portfolio.ipynb` 文件下 `draw_mst_mixed` 函数对全部测试时期的股票进行的选择如图 2.1 所示, 相较 Cho(2023)原文中所选的股票几乎均匀地分布在 MST 网络多个枝杈的情况不同, 我们的选股集中在几个枝杈之内.

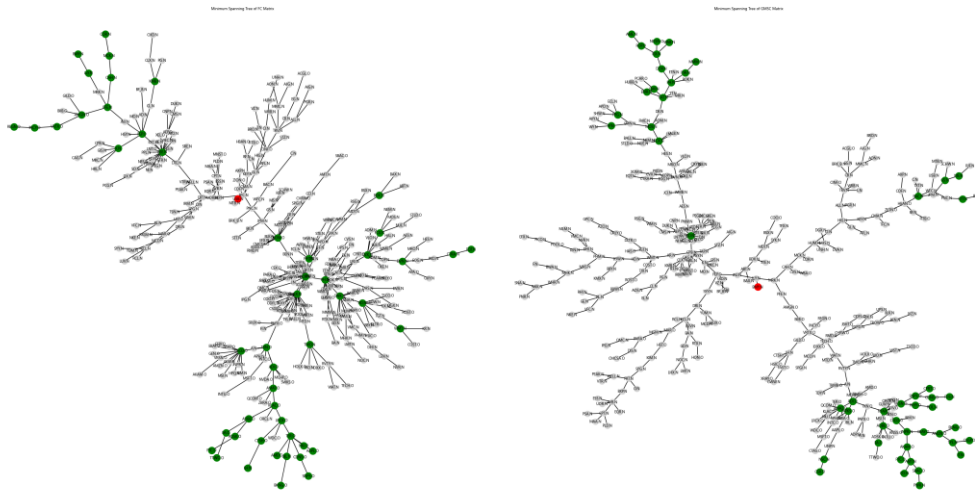


图 2.1 混合评分下基于 FC(左)和 GMSC(右)的选股, 绿点为被选择的股票

我们多次尝试调整参数, 最终认为如图 2.2 展示的算法是与原文结果相近的. 这种分配方案直接按照 MST 的 degree 进行排序, 也即优先选择连接数量小的节点. 我们认为与不成功的混合评分排序算法相比, 这种方法更能够简单直观地

体现 HRP-FC/GMSC 依托 MST 进行分散风险与多样化的选股思路. 因此我们在后续的计算中采用的都是这一方法.

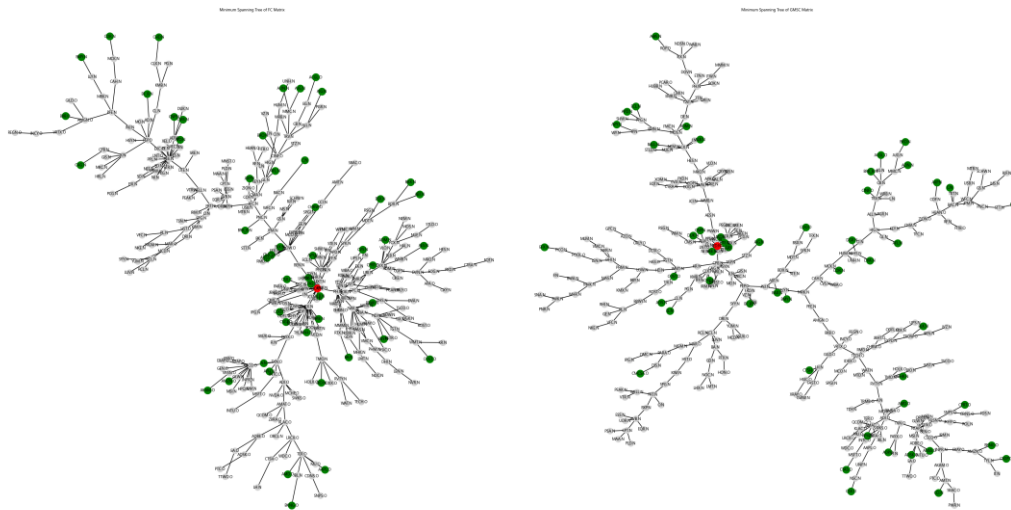


图 2.2 degree 评分下基于 FC(左)和 GMSC(右)的选股, 绿点为被选择的股票⁵

2.3 分层风险平价算法

分层风险平价(Hierarchical Risk Parity, HRP)由 López 在 2016 年提出, 其核心思想是, 金融市场存在一种简单的协方差矩阵无法体现的“层次结构”, 而顺应这种层次结构进行权重分配的组合能够更好地平衡风险和收益。层次结构的一个典型例子是行业的分级, 如伯克希尔哈撒韦公司主营业务为保险, 它的行业分类就是金融业—非银金融—保险业, 而同一行业下的公司的股票价格往往会受到行业层面的影响呈现协整的波动, 同时又因为相互的竞争而出现负相关。这个理念指导下的投资组合非常符合主观投资者的直觉。

在 2016 年的文献中, López 用层次聚类来识别协方差矩阵的层次结构, 用不同簇间的距离对协方差矩阵重新排序以体现层次结构, 并按照该顺序递归地将组合不断二分, 以逆方差的方法为簇内资产赋权。可以认为, 2016 年提出的这种 HRP 本质上就是一种借助机器学习和图论的工具以在资产相关性上体现层次

⁵ 过程见 demo_single_portfolio.ipynb

结构的逆方差算法⁶。

关于具体的流程和公式，本文档不做赘述。详细实现可以参阅本项目 Reference 文件夹下 [López\(2016\)的文章](#)，或者这篇帖子: [The Hierarchical Risk Parity Algorithm: An Introduction](#)。

2.4 基准算法

原文献只涉及了 HRP 一种算法，并发现其相较 S&P500 有一定优势。我们希望检验这种策略下 HRP 能否优于其他基准算法。这些基准算法在原框架的复刻下不会使用，而会在多种窗口长度的检验中被使用。

风险平价算法是基于风险的，与其联系最紧密的是逆方差/逆波动算法 (Inverse Volatility, IV), 如式 (2-2)，以及有效前沿/最小方差算法 (Efficient Frontier, EF), 如式(2-3)。

$$w_n = \frac{\frac{1}{\sigma_n^2}}{\sum_{n=1}^N \frac{1}{\sigma_n^2}} \quad (2-2)$$

$$\begin{aligned} \text{minimize}_w &= \frac{1}{2} w^T \Sigma w \\ \text{s.t. } & w^T \alpha = \alpha_0 \\ & w^T e = 1 \end{aligned} \quad (2-3)$$

2.5 整个策略的流程

HRP-FC/GMSC 策略流程如图 2.2 所示。首先基于历史收益率数据估计相关系数矩阵，基于该矩阵计算出 FC 和 GMSC 矩阵，构建最小生成树，计算混合边缘性度量进行选股。与此同时，采用 HRP 算法为所有的证券赋予权重，再剔除掉没有被选中的(不够外围)的资产，对剩下的权重进行放缩以保证其和为 1。

⁶ 这段文字叙述引自我尚未提交终稿的毕业论文。烦请不要外传。

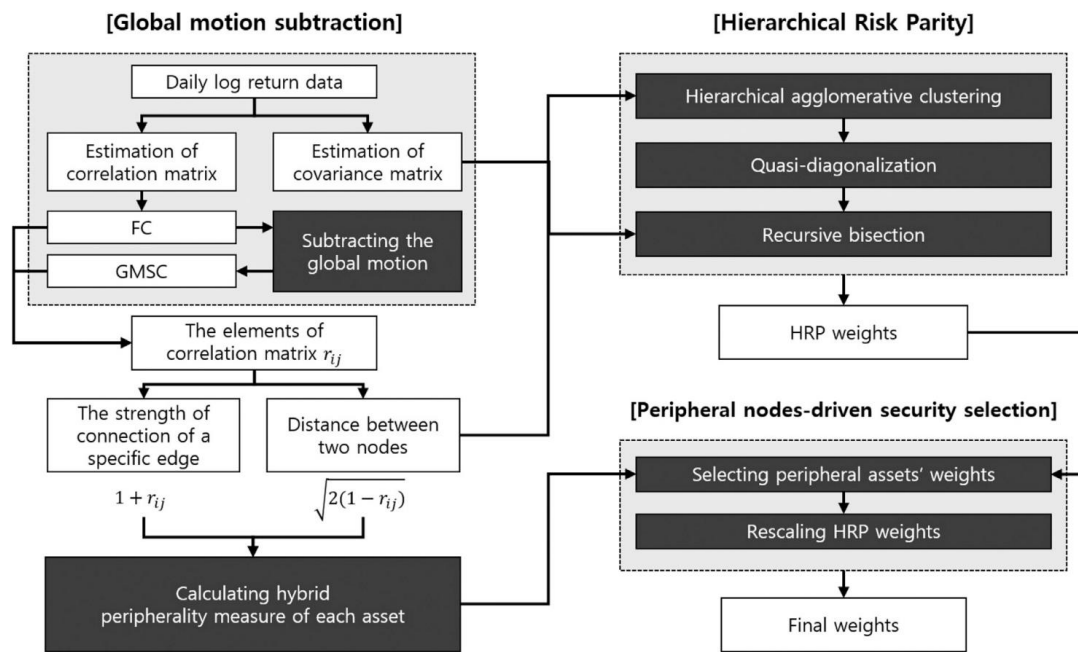


图 2.2 策略整体流程

基于相关系数矩阵进行聚类时，在原有文献中采用的是经历过 FC 或 GMSC 处理的相关系数矩阵，而我们同时回测了基于未经这些处理的相关系数矩阵的方案以供对比。

3. 数据和回测

3.1 数据

本研究采用 S&P500 指数成分股数据, 时间为从 2000/1/1-2024/2/27. 原始数据来自 Wind 金融终端相应表单, 为价格数据. 收益率数据采用百分比收益. 这可能是我们后续复现结果与原文有区别的原因之一, 因为原文采用了对数收益率⁷. 我们选择的 S&P500 指数成分股为 2024 年 2 月 27 日从 Wind 上下载的版本, 共有 500 个, 经过清洗后得到在我们整个实验期间有稳定交易的股票 352 支, 作为回测的数据.

3.2 回测方法

采用滚动窗口回测. 窗口时间长度为 120 日. 为确保我们的回测是样本外的, 我们采用第一个窗口(0-119 日)决定第 120 日的选股和权重, 采用第二个窗口(1-120 日)决定第 121 日的选股和权重, 以此类推.

我们选择了年化收益率, 年化标准差, 夏普比率和最大回撤作为衡量回测表现的指标, 其中计算收益率时扣除了以 0.0001 费率计算的交易费用, 计算夏普比率时采用的无风险收益率为 0.02. 这两个参数的设置在原文献中没有提及, 我们认为, 为模拟真实的交易环境, 这两个参数是必要的.

衡量回测表现时, 我们参考了原文献的时间划分, 设置了五个回测时间段: 子时段 1 覆盖 2001 年 1 月 1 日至 2006 年 12 月 31 日, 这一时期是横盘市场, 市场整体波动较小, 涨跌交替, 没有明显的长期上涨或下跌趋势; 子时段 2 从 2007 年 1 月 1 日至 2009 年 12 月 31 日, 这段时间包括次贷危机, 全球金融市场极度动荡; 子时段 3 从 2010 年 1 月 1 日至 2019 年 12 月 31 日, 这是牛市期间, 市场整体表现强劲, 经济从全球金融危机中恢复; 子时段 4 从 2020 年 1 月 1 日至 2020 年 4 月 30 日, 特别关注 COVID-19 大流行初期带来的市场危机, 股市快速下跌, 经济活动受到严重影响; 子时段 5 从 2020 年 5 月 1 日至 2021 年 8 月 31 日, 市场从疫情初期的震荡中逐渐恢复, 再次进入牛市. 通过这样的

⁷ 我们希望更换收益率计算方式后这个流程仍然稳健.

时段选择, 我们能够在各种市场条件下测试和比较不同投资策略的效果, 为投资者提供更有针对性和实用的指导.

图 3.1 展示了在组内资产数为 10 时, FC 与 GMSC 策略的组合的权重, 日收益率和累计收益率变化.

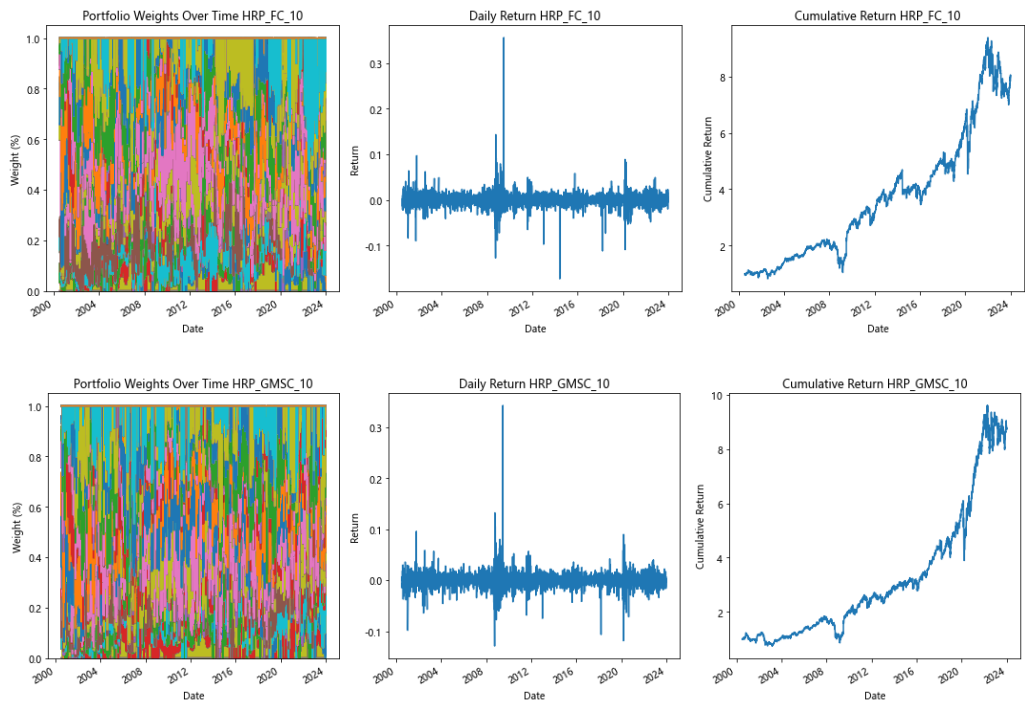


图 3.1 策略整体流程

特别地, 我们设置了 0.0001 的交易费用率, 这在美股市场是一个常见的手续费率. 在 Cho(2023)的文章中并没有提及手续费的计算, 我们认为这是不严谨的, 结合图 3.1 也能够看出, 这是一个换手率极高的策略⁸(事实上, 在我们的回测中, 在不同组内证券数量的情况下, 大部分组合都有接近 0.3 的日换手率).

图 3.2 说明了设置手续费的必要性. 在组合内证券数为 5 和 350 的两种情况下, 这两个组合的无手续费累计收益率(蓝色)明显优于有手续费的累计收益率情况(橙色). 在我们的回测框架中, 无手续费情况是由日收益率累乘而来, 有手续费的情况则是由组合日价格迭代而来, 这两个数据的相互印证也检验了我们自拟框架的准确性.

⁸ 当然, 这也是比较容易优化的部分, Cho(2023)可能只是想提出一个基本的框架.

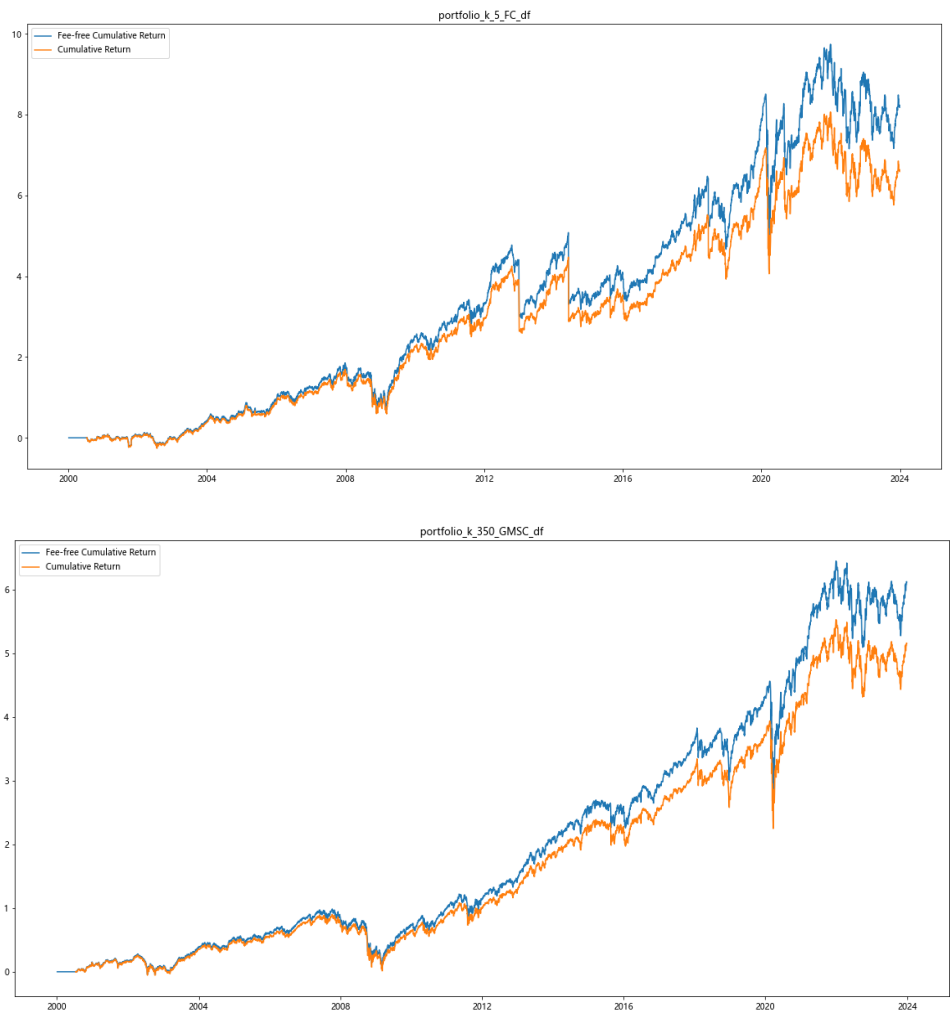


图 3.2 手续费的必要性

3.3 回测表现

3.3.1 采用原始相关系数矩阵进行聚类

表 3.1 展示了 HRP-FC/GMSC 策略的回测表现，其中，FC 和 GMSC 的策略在进行聚类时都采用原始的相关系数矩阵。表格中，标绿的部分为在同样的组内资产数 N 下，FC 或者 GMSC 两个分支中表现更优的一方。我们建议参阅原文策略复现评估.xlsx 以获得更好的阅读体验。

表 3.1 主策略回测表现-采用原始的相关系数矩阵

N strategy_type	5		10		20		30		50	
	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC
Sideways market										
p-1 an~return	0.169	0.087	0.129	0.098	0.153	0.099	0.169	0.092	0.118	0.093
p-1 an~std	0.227	0.226	0.182	0.183	0.158	0.163	0.143	0.148	0.134	0.142

p-1 sharpe	0.613	0.252	0.541	0.372	0.783	0.422	0.972	0.418	0.662	0.442
p-1 max~draw	-0.314	-0.488	-0.259	-0.373	-0.247	-0.341	-0.242	-0.280	-0.255	-0.276
Sub-prime mortgage crisis										
p-2 an~return	0.170	-0.013	0.216	0.169	0.071	0.075	0.030	0.077	-0.006	0.047
p-2 an~std	0.296	0.290	0.468	0.447	0.293	0.328	0.253	0.297	0.224	0.272
p-2 sharpe	0.473	-0.149	0.397	0.311	0.140	0.137	0.000	0.160	-0.162	0.063
p-2 max~draw	-0.421	-0.544	-0.529	-0.556	-0.453	-0.493	-0.444	-0.454	-0.441	-0.430
Bull market										
p-3 an~return	0.056	0.098	0.074	0.113	0.119	0.135	0.122	0.106	0.117	0.115
p-3 an~std	0.231	0.175	0.176	0.158	0.154	0.150	0.145	0.148	0.131	0.138
p-3 sharpe	0.112	0.392	0.249	0.524	0.574	0.698	0.636	0.510	0.663	0.620
p-3 max~draw	-0.431	-0.241	-0.281	-0.199	-0.213	-0.168	-0.216	-0.185	-0.182	-0.172
COVID-19 pandemic crisis										
p-4 an~return	-0.369	-0.439	-0.357	-0.434	-0.392	-0.393	-0.300	-0.372	-0.157	-0.403
p-4 an~std	0.610	0.571	0.560	0.586	0.556	0.576	0.519	0.567	0.490	0.546
p-4 sharpe	-0.653	-0.821	-0.692	-0.791	-0.760	-0.734	-0.636	-0.708	-0.381	-0.793
p-4 max~draw	-0.367	-0.359	-0.347	-0.364	-0.372	-0.361	-0.343	-0.372	-0.296	-0.369
Bull market										
p-5 an~return	0.155	0.407	0.399	0.441	0.407	0.392	0.341	0.357	0.343	0.339
p-5 an~std	0.229	0.203	0.176	0.190	0.159	0.167	0.146	0.155	0.142	0.152
p-5 sharpe	0.545	1.856	2.092	2.167	2.380	2.162	2.135	2.103	2.202	2.038
p-5 max~draw	-0.228	-0.174	-0.137	-0.126	-0.083	-0.088	-0.082	-0.082	-0.074	-0.083
N	100		200		300		350		\	
strategy_type	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC	SP500	mean-HRP
Sideways market										
p-1 an~return	0.093	0.087	0.101	0.098	0.092	0.087	0.087	0.088	0.013	0.092
p-1 an~std	0.130	0.137	0.123	0.133	0.125	0.129	0.126	0.126	0.172	0.128
p-1 sharpe	0.485	0.415	0.576	0.513	0.494	0.445	0.455	0.458	-0.101	0.480
p-1 max~draw	-0.256	-0.282	-0.244	-0.258	-0.257	-0.265	-0.257	-0.257	-0.426	-0.260
Sub-prime mortgage crisis										
p-2 an~return	-0.003	0.023	-0.036	-0.024	-0.035	-0.035	-0.035	-0.033	-0.079	-0.022
p-2 an~std	0.230	0.267	0.243	0.267	0.248	0.256	0.252	0.253	0.304	0.252
p-2 sharpe	-0.143	-0.025	-0.271	-0.204	-0.261	-0.254	-0.259	-0.251	-0.360	-0.208
p-2 max~draw	-0.447	-0.466	-0.476	-0.484	-0.474	-0.488	-0.475	-0.474	-0.568	-0.473
Bull market										
p-3 an~return	0.111	0.105	0.115	0.108	0.116	0.115	0.114	0.114	0.117	0.112
p-3 an~std	0.130	0.138	0.127	0.135	0.128	0.132	0.129	0.129	0.149	0.131
p-3 sharpe	0.620	0.546	0.670	0.578	0.671	0.643	0.653	0.650	0.581	0.629
p-3 max~draw	-0.196	-0.179	-0.186	-0.193	-0.177	-0.185	-0.176	-0.176	-0.198	-0.184
COVID-19 pandemic crisis										
p-4 an~return	-0.176	-0.367	-0.233	-0.379	-0.260	-0.303	-0.268	-0.272	-0.282	-0.282
p-4 an~std	0.502	0.550	0.509	0.555	0.519	0.542	0.527	0.528	0.548	0.529
p-4 sharpe	-0.412	-0.722	-0.517	-0.736	-0.559	-0.615	-0.566	-0.572	-0.569	-0.587
p-4 max~draw	-0.319	-0.364	-0.328	-0.370	-0.338	-0.352	-0.342	-0.343	-0.339	-0.344
Bull market										
p-5 an~return	0.360	0.362	0.357	0.368	0.342	0.353	0.335	0.335	0.405	0.351
p-5 an~std	0.143	0.158	0.142	0.157	0.144	0.150	0.145	0.146	0.163	0.148

p-5 sharpe	2.310	2.103	2.298	2.153	2.164	2.146	2.101	2.098	2.305	2.172
p-5 max~draw	-0.077	-0.085	-0.079	-0.085	-0.078	-0.082	-0.078	-0.079	-0.096	-0.081

注: p-x 表示特定的时段,其中 x 是时段编号; an~return 是年化收益率, 衡量资产或投资组合的平均年回报; an~std 是年化标准差, 反映投资的波动性或风险; sharpe 是夏普比率, 用于评估每单位风险所获得的超额回报; max~draw 是最大回撤, 显示期间的最大潜在亏损

3.3.2 采用 FC/GMSC 处理的相关系数矩阵进行聚

接着, 表 3.2 报告了采用经过了 FC/GMSC 处理的相关系数矩阵进行聚类的回测结果.

表 3.2 主策略回测表现-采用 FC/GMSC 处理后的相关系数矩阵

k	5	5	10	10	20	20	30	30	50	50
strategy_type	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC
Sideways market										
p-1 an~return	0.130	0.021	0.112	0.045	0.114	0.062	0.115	0.068	0.093	0.073
p-1 an~std	0.215	0.221	0.176	0.182	0.158	0.163	0.146	0.148	0.133	0.142
p-1 sharpe	0.467	-0.040	0.466	0.080	0.530	0.196	0.583	0.254	0.475	0.300
p-1 max~draw	-0.329	-0.511	-0.301	-0.408	-0.275	-0.372	-0.267	-0.297	-0.262	-0.284
Sub-prime mortgage crisis										
p-2 an~return	0.156	-0.016	0.124	0.121	0.044	0.039	0.026	0.047	0.003	0.023
p-2 an~std	0.297	0.289	0.368	0.364	0.277	0.299	0.247	0.280	0.223	0.266
p-2 sharpe	0.423	-0.159	0.256	0.251	0.052	0.030	-0.016	0.059	-0.122	-0.025
p-2 max~draw	-0.409	-0.516	-0.532	-0.534	-0.431	-0.486	-0.417	-0.451	-0.415	-0.433
Bull market										
p-3 an~return	0.093	0.107	0.088	0.114	0.123	0.127	0.129	0.101	0.118	0.115
p-3 an~std	0.203	0.181	0.167	0.161	0.150	0.151	0.140	0.150	0.130	0.138
p-3 sharpe	0.308	0.428	0.347	0.520	0.623	0.641	0.705	0.476	0.677	0.613
p-3 max~draw	-0.316	-0.244	-0.265	-0.186	-0.184	-0.159	-0.183	-0.175	-0.164	-0.165
COVID-19 pandemic crisis										
p-4 an~return	-0.399	-0.367	-0.345	-0.381	-0.368	-0.338	-0.266	-0.303	-0.201	-0.360
p-4 an~std	0.605	0.565	0.551	0.593	0.544	0.577	0.507	0.564	0.487	0.544
p-4 sharpe	-0.710	-0.703	-0.682	-0.693	-0.732	-0.638	-0.585	-0.590	-0.474	-0.716
p-4 max~draw	-0.381	-0.357	-0.336	-0.363	-0.363	-0.352	-0.336	-0.359	-0.302	-0.362
Bull market										
p-5 an~return	0.251	0.497	0.398	0.474	0.385	0.351	0.316	0.339	0.324	0.337
p-5 an~std	0.217	0.206	0.173	0.191	0.157	0.169	0.144	0.155	0.143	0.151
p-5 sharpe	1.020	2.269	2.122	2.327	2.267	1.894	1.978	1.992	2.054	2.035
p-5 max~draw	-0.191	-0.149	-0.134	-0.121	-0.092	-0.089	-0.088	-0.078	-0.080	-0.079
k	100	100	200	200	300	300	350	350	\	
strategy_type	FC	GMSC	FC	GMSC	FC	GMSC	FC	GMSC	SP500	mean-HRP
Sideways market										
p-1 an~return	0.081	0.086	0.091	0.093	0.083	0.083	0.079	0.080	0.013	0.084
p-1 an~std	0.128	0.135	0.122	0.132	0.123	0.127	0.124	0.124	0.172	0.127
p-1 sharpe	0.400	0.413	0.500	0.479	0.427	0.416	0.397	0.399	-0.101	0.429
p-1 max~draw	-0.268	-0.269	-0.244	-0.255	-0.250	-0.259	-0.250	-0.250	-0.426	-0.256
Sub-prime mortgage crisis										
p-2 an~return	0.004	0.006	-0.027	-0.026	-0.030	-0.032	-0.030	-0.029	-0.079	-0.021

p-2 an~std	0.230	0.264	0.243	0.265	0.248	0.255	0.252	0.252	0.304	0.251
p-2 sharpe	-0.115	-0.092	-0.236	-0.211	-0.242	-0.241	-0.240	-0.232	-0.360	-0.201
p-2 max~draw	-0.436	-0.468	-0.466	-0.482	-0.467	-0.481	-0.468	-0.466	-0.568	-0.467
Bull market										
p-3 an~return	0.111	0.105	0.118	0.111	0.119	0.118	0.117	0.117	0.117	0.115
p-3 an~std	0.129	0.138	0.127	0.135	0.128	0.131	0.129	0.129	0.149	0.131
p-3 sharpe	0.626	0.543	0.697	0.600	0.696	0.674	0.679	0.675	0.581	0.649
p-3 max~draw	-0.189	-0.184	-0.185	-0.186	-0.177	-0.181	-0.175	-0.175	-0.198	-0.182
COVID-19 pandemic crisis										
p-4 an~return	-0.237	-0.334	-0.258	-0.387	-0.272	-0.310	-0.279	-0.281	-0.282	-0.295
p-4 an~std	0.496	0.548	0.507	0.555	0.517	0.541	0.525	0.526	0.548	0.527
p-4 sharpe	-0.538	-0.664	-0.569	-0.752	-0.585	-0.629	-0.589	-0.592	-0.569	-0.615
p-4 max~draw	-0.324	-0.357	-0.330	-0.371	-0.338	-0.352	-0.341	-0.342	-0.339	-0.344
Bull market										
p-5 an~return	0.340	0.350	0.352	0.361	0.340	0.348	0.332	0.334	0.405	0.344
p-5 an~std	0.143	0.156	0.141	0.156	0.143	0.150	0.144	0.145	0.163	0.147
p-5 sharpe	2.173	2.045	2.273	2.121	2.168	2.125	2.091	2.099	2.305	2.137
p-5 max~draw	-0.080	-0.083	-0.077	-0.084	-0.077	-0.081	-0.077	-0.077	-0.096	-0.079

在整体上, 对相关系数矩阵进行 FC/GMSC 处理对算法的回测表现影响较小, 进行处理的比不进行处理的有微弱的提升. 进行处理的组的所有组合的夏普比率的均值为 0.4669, 而不进行处理的相应均值为 0.4935, 这主要是因为在一个时期, 不进行 FC/GMSC 处理的组合的表现大幅优于处理的组合, 而在其余四个时期, 进行处理的组合以微弱的优势领先. 在 FC 于 GMSC 这两个算法的相对优势方面, 不论是否处理相关系数矩阵, 基于 GMSC 选股的策略都在第三, 四和五个时期的含有 20 个以内证券的组合上相较 FC 有优势, 而在其他方面往往是 FC 策略占优.

我们将与原文献的结果一起检验我们的回测结果. 图 3.1 展示了原文献的回测结果. 在第一个历史时期 Sideways market 中, 我们发现 HRP-FC/GMSC 不论在收益率还是波动性上的表现都优于 S&P500, 而在两种处理收益率相关系数矩阵的方法中 FC 的表现优于 GMSC, 这与 Cho(2023)的结论基本一致.

Sub-period portfolio performance analysis.											
		$\kappa = 5$	$\kappa = 10$	$\kappa = 20$	$\kappa = 30$	$\kappa = 50$	$\kappa = 100$	$\kappa = 200$	$\kappa = 300$	All	S&P500
Sub-period 1: 01/2001-12/2006 (Sideways market)											
C/R	FC-HRP	0.2272 ^a	0.1430 ^a	0.1086 ^a	0.0887 ^a	0.0864	0.0946 ^a	0.0747	0.0727	0.0822 ^a	0.0176
	GMSC-HRP	-0.0222	-0.0167	0.0655	0.0752	0.0870 ^a	0.0849	0.0838 ^a	0.0740 ^a	0.0784	
S/D	FC-HRP	0.2320	0.1905	0.1956	0.1815	0.1635	0.1492 ^a	0.1445 ^a	0.1394 ^a	0.1354 ^a	0.1707
	GMSC-HRP	0.2307 ^a	0.1859 ^a	0.1678 ^a	0.1608 ^a	0.1569 ^a	0.1536	0.1472	0.1448	0.1356	
MDD	FC-HRP	-0.3627 ^a	-0.3249 ^a	-0.4818	-0.4615	-0.3920	-0.3464	-0.3107	-0.3126 ^a	-0.2943 ^a	-0.4605
	GMSC-HRP	-0.5785	-0.5176	-0.4087 ^a	-0.3634 ^a	-0.3490 ^a	-0.3380 ^a	-0.3081 ^a	-0.3163	-0.3000	
SH/R	FC-HRP	0.9794 ^a	0.7504 ^a	0.5550 ^a	0.4888 ^a	0.5284	0.6339 ^a	0.5173	0.5217 ^a	0.6073 ^a	0.1031
	GMSC-HRP	-0.0961	-0.0899	0.3902	0.4679	0.5544 ^a	0.5528	0.5693 ^a	0.5111	0.5781	
CA/R	FC-HRP	0.6263 ^a	0.4400 ^a	0.2254 ^a	0.1922 ^a	0.2204	0.2730 ^a	0.2406	0.2326	0.2793 ^a	0.0382
	GMSC-HRP	-0.0383	-0.0323	0.1602	0.2070	0.2492 ^a	0.2512	0.272 ^a	0.234 ^a	0.2613	
Sub-period 2: 01/2007-12/2009 (Sub-prime mortgage crisis)											
C/R	FC-HRP	-0.2082	-0.1414	-0.1059	-0.1235	-0.0847	-0.0651	-0.0723	-0.0723	-0.0909	-0.0710
	GMSC-HRP	-0.0019 ^a	-0.0686 ^a	-0.0425 ^a	-0.0668 ^a	-0.0048 ^a	-0.0323 ^a	-0.0406 ^a	-0.0609 ^a	-0.0867 ^a	
S/D	FC-HRP	0.4384 ^a	0.3719 ^a	0.3243 ^a	0.3200 ^a	0.3024 ^a	0.2917 ^a	0.2823 ^a	0.2744 ^a	0.2672	0.2995
	GMSC-HRP	0.4503	0.4001	0.3481	0.3515	0.3343	0.3179	0.2994	0.2873	0.2651 ^a	
MDD	FC-HRP	-0.7923	-0.7056	-0.6552	-0.6726 ^a	-0.6172	-0.5976 ^a	-0.6156	-0.6074 ^a	-0.6197	-0.6103
	GMSC-HRP	-0.7216 ^a	-0.6810 ^a	-0.6393 ^a	-0.6865	-0.6018 ^a	-0.6239	-0.6048 ^a	-0.6080	-0.6127 ^a	
SH/R	FC-HRP	-0.4749	-0.3802	-0.3265	-0.3861	-0.2801	-0.2234	-0.2560	-0.2635	-0.3403	-0.2372
	GMSC-HRP	-0.0043	-0.1714	-0.1221	-0.1900	-0.0143	-0.1017	-0.1357	-0.2121	-0.3271	
CA/R	FC-HRP	-0.2628	-0.2004	-0.1616	-0.1837	-0.1372	-0.1090	-0.1174	-0.119	-0.1467	-0.1164
	GMSC-HRP	-0.0027	-0.1007	-0.0665	-0.0973	-0.008	-0.0518	-0.0672	-0.1002	-0.1416	
Sub-period 3: 01/2010-12/2019 (Bull market)											
C/R	FC-HRP	0.2841 ^a	0.3383 ^a	0.2672 ^a	0.2530 ^a	0.2165 ^a	0.1915	0.1614	0.1736	0.1788	0.1855
	GMSC-HRP	0.2592	0.2507	0.1733	0.2414	0.2138	0.2096 ^a	0.1685 ^a	0.1843 ^a	0.1814 ^a	
S/D	FC-HRP	0.2204	0.1794 ^a	0.1620 ^a	0.1593 ^a	0.1511 ^a	0.1438 ^a	0.1397 ^a	0.1330 ^a	0.1257 ^a	0.1479
	GMSC-HRP	0.2018 ^a	0.1834	0.1722	0.1701	0.1613	0.1571	0.1473	0.1410	0.1257 ^a	
MDD	FC-HRP	-0.3307	-0.2557 ^a	-0.2564 ^a	-0.2441 ^a	-0.2630	-0.2394 ^a	-0.2171 ^a	-0.2081 ^a	-0.1849	-0.2070
	GMSC-HRP	-0.3180 ^a	-0.2809	-0.2710	-0.2732	-0.2509 ^a	-0.2495	-0.2463	-0.2271	-0.1831 ^a	
SH/R	FC-HRP	1.2892 ^a	1.8851 ^a	1.6500 ^a	1.5888 ^a	1.4330 ^a	1.3324	1.1555 ^a	1.3055	1.4208	1.2546
	GMSC-HRP	1.2847	1.3666	1.0061	1.4194	1.3257	1.3341 ^a	1.1443	1.3074 ^a	1.4431 ^a	
CA/R	FC-HRP	0.8592 ^a	1.3229	1.0421 ^a	1.0367 ^a	0.8231	0.8001	0.7435 ^a	0.8343 ^a	0.9670	0.8962
	GMSC-HRP	0.8151	0.8924	0.6394	0.8835	0.8521 ^a	0.8402 ^a	0.6842	0.8118	0.9907 ^a	
Sub-period 4: 01/2020-04/2020 (COVID-19 pandemic crisis)											
C/R	FC-HRP	-0.8415	-0.8528	-0.7915	-0.6895	-0.6395	-0.5317	-0.4609	-0.4609	-0.4169	-0.3258
	GMSC-HRP	-0.5311 ^a	-0.5091 ^a	-0.5988 ^a	-0.5379 ^a	-0.5860 ^a	-0.4944 ^a	-0.4126 ^a	-0.3715 ^a	-0.3795 ^a	
S/D	FC-HRP	0.7920	0.7520	0.7232	0.6656	0.6584	0.6249	0.5914 ^a	0.5789	0.5424	0.5391
	GMSC-HRP	0.7295 ^a	0.6822 ^a	0.6830 ^a	0.6482 ^a	0.6545 ^a	0.6217 ^a	0.6021	0.5699 ^a	0.5310 ^a	
MDD	FC-HRP	-0.5675	-0.5432	-0.5109	-0.4842	-0.4754 ^a	-0.4515	-0.4197	-0.4150	-0.3927	-0.3610
	GMSC-HRP	-0.4759 ^a	-0.4716 ^a	-0.4781 ^a	-0.4595 ^a	-0.4770	-0.4382 ^a	-0.4161 ^a	-0.3963 ^a	-0.3801 ^a	
SH/R	FC-HRP	-1.0624	-1.1340	-1.0943	-1.0358	-0.9714	-0.8509	-0.7793	-0.7962	-0.7686	-0.6044
	GMSC-HRP	-0.7280	-0.7462	-0.8767	-0.8299	-0.8952	-0.7952	-0.6852	-0.6518	-0.7147	
CA/R	FC-HRP	-1.4828	-1.5698	-1.5493	-1.4239	-1.3452	-1.1778	-1.0981	-1.1107	-1.0616	-0.9025
	GMSC-HRP	-1.1160	-1.0795	-1.2523	-1.1708	-1.2283	-1.1283	-0.9914	-0.9372	-0.9983	
Sub-period 5: 05/2020-08/2021 (Bull market)											
C/R	FC-HRP	0.3915 ^a	0.6137 ^a	0.5231 ^a	0.4802 ^a	0.4746 ^a	0.4961 ^a	0.4421 ^a	0.3486 ^a	0.3196 ^a	0.4483
	GMSC-HRP	0.1893	0.1702	0.3137	0.2685	0.4307	0.3660	0.3433	0.3377	0.3095	
S/D	FC-HRP	0.2765	0.2442	0.1932	0.1880	0.1830	0.1678	0.1643	0.1548	0.1460	0.1623
	GMSC-HRP	0.2043 ^a	0.1893 ^a	0.1859 ^a	0.1753 ^a	0.1766 ^a	0.1667 ^a	0.1640 ^a	0.1616	0.1456 ^a	
MDD	FC-HRP	-0.1472	-0.1244 ^a	-0.0963 ^a	-0.0902 ^a	-0.0902 ^a	-0.0928 ^a	-0.1013 ^a	-0.1002 ^a	-0.0965 ^a	-0.0979
	GMSC-HRP	-0.1244 ^a	-0.1333	-0.1338	-0.1220	-0.1269	-0.1168	-0.1159	-0.1121	-0.0967	
SH/R	FC-HRP	1.4156 ^a	2.513 ^a	2.708 ^a	2.5548 ^a	2.5938 ^a	2.9575 ^a	2.6901 ^a	2.2519 ^a	2.1895 ^a	2.7626
	GMSC-HRP	0.9269	0.8995	1.6873	1.5316	2.4386	2.1962	2.0924	2.090	2.1266	
CA/R	FC-HRP	2.6598 ^a	4.9332 ^a	5.4315 ^a	5.3226 ^a	5.2609 ^a	5.346 ^a	4.3624 ^a	3.4796 ^a	3.311 ^a	4.5785
	GMSC-HRP	1.5216	1.2767	2.3435	2.2002	3.3935	3.133	2.9616	3.0137	3.2012	

Note: The abbreviations C/R, S/D, MDD, SH/R, and CA/R refer to the (annualized) compound return, standard deviation, maximum drawdown, Sharpe ratio, and Calmar ratio, respectively.

^aIndicates the higher portfolio performance between the FC- and GMSC-HRP.

图 3.3 原文献回测结果

在图 3.4, 我们汇报了采用 FC/GMSC 处理的组合的累计收益率情况, 并区分了有手续费(红色)与无手续费(蓝色)的情况。

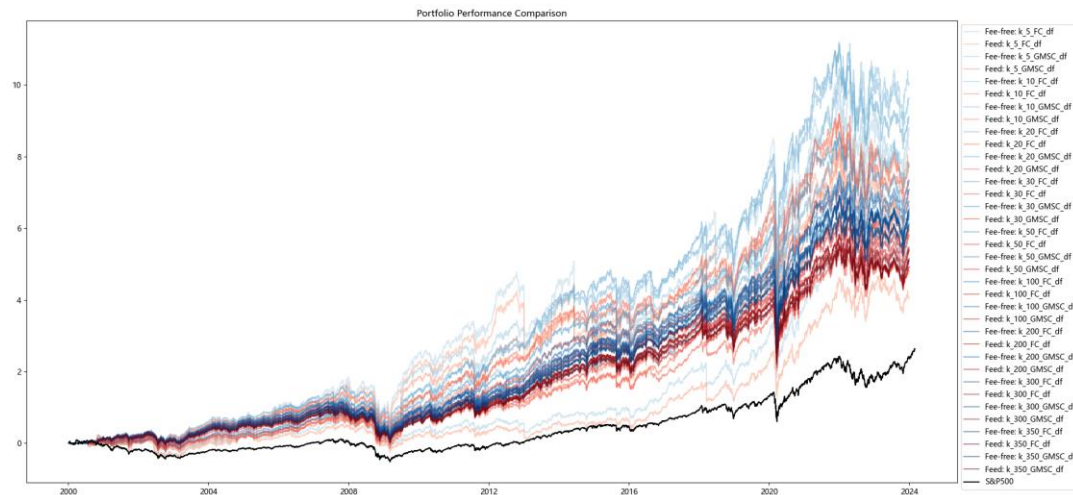


图 3.4 累计收益率情况

需要说明的是, 我们的回测结果在累计收益率图方面看起来比 Cho(2023)的

结果优秀很多, 如其原文中汇报的最优秀的策略在约 20 年的整个测试期间中也仅增值不足 200%, 而我们的策略在 2020 年增值能达到 600%左右, 我们认为造成这种差异的原因主要是原作者错误地采用了对数收益率. 对数收益率的精度缺失在回测期间的约 5000 个交易日中逐渐累积, 造成了巨大差异. 这个差异并不是本复现报告关注的重点, 对于感兴趣的读者, 我们在 [TestLogReturn.xlsx](#) 做了相关说明.

结论

我们尝试对 Cho(2023)提出的一种基于最小生成树思想的选股以及赋予权重的 HRP-FC/GMSC 策略进行复现, 因为未能完整复现其“混合得分”的计算, 采用 MST 的 degree 进行替代, 并取得了相似的效果. 我们认为 HRP-FC/GMSC 在个人和机构投资者的实际投资组合管理工作中存在一定实践意义.

后续研究可考虑降低交易频率, 因为该策略的平均日换手率, 不论含有资产数量, 均在 0.3 左右, 有较大的交易摩擦成本. 另外, 应该考虑适当的止盈, 止损策略, 如在市场连续下跌 k 日后减少仓位, 持有现金.

参考文献

- [1] Burggraf, T. (2021). Beyond risk parity – A machine learning-based hierarchical risk parity approach on cryptocurrencies. *Finance Research Letters*, 38, 101523. <https://doi.org/10.1016/j.frl.2020.101523>
- [2] Cho, Y., & Song, J. W. (2023). Hierarchical risk parity using security selection based on peripheral assets of correlation-based minimum spanning trees. *Finance Research Letters*, 53, 103608. <https://doi.org/10.1016/j.frl.2022.103608>
- [3] López De Prado, M. (2016). Building Diversified Portfolios that Outperform Out of Sample. *The Journal of Portfolio Management*, 42(4), 59–69. <https://doi.org/10.3905/jpm.2016.42.4.059>
- [4] Pozzi, F., Di Matteo, T., & Aste, T. (2013). Spread of risk across financial markets: Better to invest in the peripheries. *Scientific Reports*, 3(1), 1665. <https://doi.org/10.1038/srep01665>