



如何利用机器学习赋能红利高股息投资

摘要

- 本文以机器学习算法为基础，在不同红利高股息投资场景下，应用不同类别的机器学习模型赋能红利高股息投资。应用一：利用基于决策树的机器学习模型选取不同风格红利指数；应用二：针对现有红利投资的传统因子，利用强化学习算法深入挖掘不同因子的时序表现，对不同因子进行动态复权。

● 机器学习应用一：红利主题指数配置

本文选取包括中证红利指数在内共计 12 个红利主题指数作为红利指数配置标的，采用 DART 增强方式的 LightGBM 算法作为红利主题资产配置模型，构建红利主题指数因子。经回测，红利主题指数因子自 2021 年 1 月至 2024 年 6 月 RankIC 为 17.11%，RankIC 胜率为 63.41%，多头组合（Top3）年化收益 15.39%，相对等权组合（7.33%）年化超额 8.06%。剔除无跟踪 ETF 产品指数后，红利主题指数因子 RankIC 为 18.92%，RankIC 胜率 63.41%，多头组合（Top3）年化收益 14.09%，相对等权组合年化超额 6.76%。

● 机器学习应用二：选股因子动态复权

泛高股息指数：定义“基本+分红”筛选条件，按照月频（每月最后一个交易日收盘）调仓，以股息率加权构建泛高股息指数。截至 2024 年 6 月 30 日，泛高股息股票池中共有 1619 只股票。自 2013 年 4 月 30 日至 2024 年 6 月 30 日，泛高股息指数年化收益率 11.05%，整体来看相对跑输中证红利全收益指数，跑赢万得全 A。行业配置方面，泛高股息指数配置比例较高的分别为医药、机械、基础化工、电新行业，占比分别为 9.97%、9.14%、8.94%、6.91%。

泛高股息选股因子探究：本节分别测试了量价、质量、盈利、估值、成长五大类因子共计 30 个小类因子，综合考虑 IC 及多头组合表现，每大类保留 2 个因子。除此之外，本文综合考虑单只股票股息率时序变化以及股息率变化在股票价格上的兑现情况，构建股息兑现因子。上述六大类因子股息兑现、量价、质量、盈利、估值、成长自 2013 年 4 月至 2024 年 6 月 IC 分别为 4.94%、7.63%、2.33%、5.19%、1.58% 以及 2.42%，等权合成后因子 IC 为 7.60%，多头组合年化收益率 17.68%。

强化学习 DQN 动态因子复权：本节采取 DQN 算法构建相应的因子轮动策略。以 t-1 期因子表现为当前状态，以 t 期动态复权因子为奖励。在最大化累计奖励的基础上构建 DQN 因子（动态复权因子）。经测试，自 2021 年 1 月 31 日至 2024 年 6 月 30 日，DQN 因子 IC 为 9.55%，多头年化收益率 17.58%，相对中证红利全收益指数年化超额收益率 5.93%，多头组合信息比率 0.90，表现均优于等权因子。

- **风险提示：**报告对应的相关结论完全基于公开的历史数据进行算法构建、统计以及计算，文中部分数据有一定滞后性，同时也存在第三方数据提供不准确或者缺失等风险；策略效果结论仅针对于回测区间得出，并不预示其未来表现，也不能保证未来的可持续性，亦不构成投资收益的保证或投资建议。

西南证券研究发展中心

分析师：郑琳琳

执业证号：S1250522110001

邮箱：zhengll@swsc.com.cn

相关研究

1. 相同工作日动量因子构建与策略探究 (2024-08-02)
2. 量价时序特征挖掘模型在深度学习因子中的应用 (2024-07-11)
3. 金融工程 2024 年中期投资策略：顺势而谋，顺势而动 (2024-07-11)
4. 审时度势：趋势共振量化大小盘风格轮动策略 (2024-05-23)
5. 换手率因子在股债中的差异性与增强作用——量化方法在债券研究中的应用二 (2024-04-03)
6. 抽丝剥茧：多视角构建长短期美债利率择时信号 (2024-03-07)
7. 多维探究黄金定价逻辑，定量择时配置有迹可循 (2024-02-22)
8. 基于宏观的大类资产多维择时框架构建 (2024-01-26)
9. 聚沙成塔：微盘股行情解析与微盘股投资策略 (2024-01-15)
10. 立足全球资产，注重组合策略对抗不确定性——2024 年金融工程投资策略 (2024-01-08)

目 录

1 前言	1
2 红利高股息投资概览—以中证红利指数(000922)为例	1
3 机器学习应用一：红利主题指数配置	3
3.1 决策树算法简介	4
3.2 基于 LightGBM 的红利指数配置策略	8
4 机器学习应用二：选股因子动态复权	11
4.1 泛高股息股票池构建	11
4.2 泛高股息选股因子探究	13
4.3 强化学习算法简介	21
4.4 基于 DQN 的泛高股息因子动态复权	23
5 总结	26

图 目 录

图 1: 中证红利指数中信一级行业分布.....	1
图 2: 中证红利指数成分股市值权重分布.....	1
图 3: 中证红利指数历史走势.....	2
图 4: XGBoost 算法原理示意.....	5
图 5: lightGBM_直方图算法示意.....	7
图 6: lightGBM_Leaf-wise 示意.....	7
图 7: 红利主题因子 RankIC.....	8
图 8: 红利主题因子 RankIC-调整后.....	8
图 9: 红利主题因子多头组合净值.....	9
图 10: 红利主题因子多头分年度表现.....	9
图 11: 红利主题因子多头分年度表现-调整后.....	9
图 12: 红利主题因子空头组合净值.....	10
图 13: 红利高股息股票池各期数量.....	11
图 14: 泛高股息指数净值 (全区间).....	12
图 15: 泛高股息指数净值 (分年度).....	12
图 16: 泛高股息指数行业配置情况 (%).....	12
图 17: 泛高股息指数板块配置情况 (%).....	12
图 18: 成长-EPS 同比分组测试.....	16
图 19: 成长-ROE 环比分组测试.....	16
图 20: 质量-ROE 标准差分组测试.....	16
图 21: 质量-ROE 均值/标准差.....	16
图 22: 盈利-ROIC 分组测试.....	16
图 23: 盈利-营业利润率分组测试.....	16
图 24: 估值-对数市值分组测试.....	17
图 25: 估值-股息率分组测试.....	17
图 26: 量价-60 日上行波动分组测试.....	17
图 27: 量价-60 日价格动量分组测试.....	17
图 28: 股息兑现 120_60.....	18
图 29: 股息兑现 90_90.....	18
图 30: 股息兑现 120_90.....	19
图 31: 股息兑现 90_120.....	19
图 32: 等权泛高股息因子 IC.....	20
图 33: 等权泛高股息因子分组.....	20
图 34: 强化学习示意.....	21
图 35: 贝尔曼方程示意.....	22
图 36: DQN 模型示意.....	23
图 37: DQN 因子 IC.....	25
图 38: DQN 因子分组测试.....	25
图 39: DQN 模型历史各大类因子权重.....	25

表 目 录

表 1: 中证红利指数长期业绩表现.....	2
表 2: 中证红利指数短期业绩表现.....	2
表 3: 红利主题相关指数.....	3
表 4: 红利主题相关指数跟踪产品.....	3
表 5: 红利资产相关性.....	4
表 6: LightGBM 模型数据集划分.....	8
表 7: 红利主题因子回测情况.....	8
表 8: 红利主题因子多头组合分年度表现.....	10
表 9: 红利主题因子空头表现.....	10
表 10: 泛高股息指数业绩对比 (全区间).....	12
表 11: 泛高股息指数业绩 (分年度).....	13
表 12: 基本因子分类及描述.....	13
表 13: 五大类基本因子测试.....	14
表 14: 各大类因子入选小类因子.....	15
表 15: 股息兑现因子回溯窗口测试.....	18
表 16: 股息兑现因子多空.....	19
表 17: 因子相关性分析.....	19
表 18: 小类因子等权合成大类因子测试结果.....	20
表 19: 等权泛高股息因子测试结果.....	20
表 20: 等权泛高股息因子测试结果.....	24
表 21: DQN 因子多空测试.....	25

1 前言

红利投资是一种能够为投资者提供稳定收入和长期增长潜力的策略。在量化研究领域，当前主流红利投资策略为基于股息率及公司其它基本面信息构建相应的选股因子进行选股。除此之外，随着红利投资的逐渐发展，不同风格的红利投资也逐渐成为红利投资的研究方向之一。

传统的多因子策略依赖于人工挖掘的选股因子，人工挖掘选股因子具备较强的逻辑性，但很难选出表现优秀且稳定的因子，且传统的多因子策略大多通过固定权重的方式合成各个因子，没有考虑到不同因子的时序表现相对强弱，该方法具备较大的 alpha 挖掘提升空间。

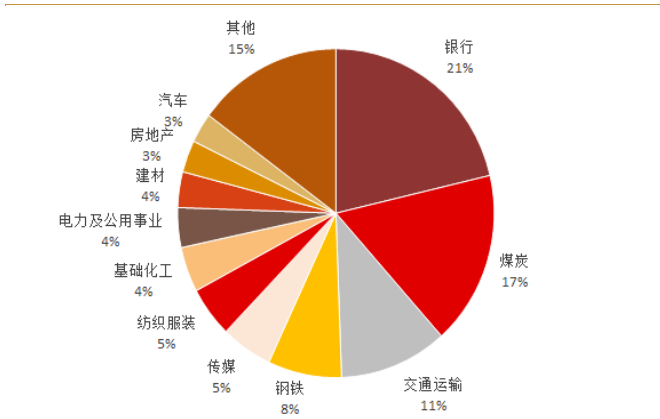
本文以机器学习算法为基础，考虑不同种类的机器学习模型在红利投资的两方面应用。**应用一：**利用基于决策树的机器学习模型选取不同风格红利指数；**应用二：**针对现有红利投资的传统因子，利用强化学习算法深入挖掘不同因子的时序表现，对不同因子进行动态复权。

2 红利高股息投资概览—以中证红利指数(000922)为例

中证红利指数(000922.CSI)是目前具备代表性的红利指数，该指数选取 100 只现金股息率高、分红较为稳定，并具有一定规模及流动性的上市公司证券作为指数样本，以反映高股息率上市公司证券的整体表现。指数以 2004 年 12 月 31 日为基日，以 1000 点为基点。

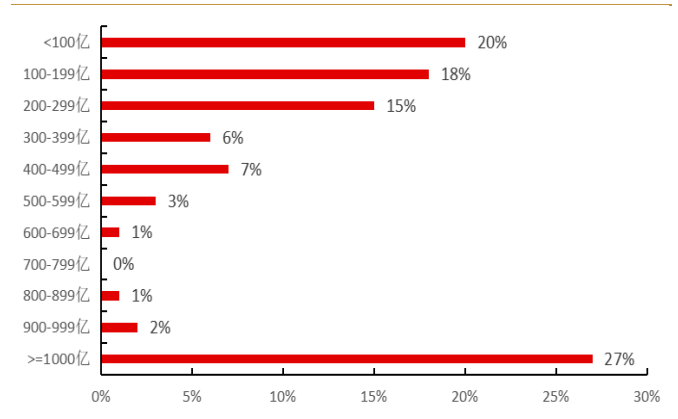
截至 2024 年 6 月 30 日，中证红利指数占比最高的前五名行业为银行、煤炭、交通运输、钢铁和传媒，占比分别为 21%、17%、11%、8%和 5%。中证红利指数的市值分布呈现两极分化的情况，市值规模超过 1000 亿的成分股占比为 27%，市值规模小于 100 亿的成分股占比为 20%。

图 1：中证红利指数中信一级行业分布



数据来源：wind、西南证券整理，数据截至 2024/06/30

图 2：中证红利指数成分股市值权重分布



数据来源：wind、西南证券整理，数据截至 2024/06/30

中证红利指数长期收益优于其他主流宽基指数，收益风险比优秀。自 2013 年 1 月 1 日至 2024 年 6 月 30 日，中证红利指数区间收益达 111.33%，年化收益达 6.99%，高于沪深 300 指数的 2.89%、万得全 A 的 5.91%和中证 800 的 3.08%。收益风险比为 0.33，同样高于沪深 300、万得全 A 和中证 800。

表 1：中证红利指数长期业绩表现

收益风险指标	沪深 300	万得全 A	中证 800	中证红利指数
区间收益(%)	37.13	88.75	39.96	111.33
年化收益率(%)	2.89	5.91	3.08	6.99
年化波动率(%)	21.75	23.93	21.83	21.47
收益风险比	0.13	0.25	0.14	0.33
最大回撤(%)	64.49	55.99	58.33	62.23

数据来源：wind、西南证券整理，数据截至 2024/06/30

图 3：中证红利指数历史走势


数据来源：wind、西南证券整理，数据截至 2024/06/30

中证红利指数今年以来的收益率为 1.96%，优于沪深 300 的-1.18%、万得全 A 指数的-10.23%和中证 800 指数的-4.44%。

表 2：中证红利指数短期业绩表现

收益风险指标	沪深 300	万得全 A	中证 800	中证红利指数
近 3 月涨跌幅(%)	-6.43	-9.06	-8.42	-7.79
近 6 月涨跌幅(%)	2.63	-2.19	0.80	-1.92
今年以来涨跌幅(%)	-1.18	-10.23	-4.44	1.96

数据来源：wind、西南证券整理，数据截至 2024/06/30

3 机器学习应用一：红利主题指数配置

本节主要讨论不同风格的红利指数资产配置问题。除中证红利指数(000922.CSI)以外，目前仍有很多其他不同风格的红利主题指数，本文选取包括中证红利指数在内共计 12 个红利主题指数，构建相应的红利指数配置策略。

表 3：红利主题相关指数

代码	简称	全收益代码	全收益简称	基期
000922.CSI	中证红利	H00922.CSI	中证红利全收益	2004-12-31
H30269.CSI	红利低波	H20269.CSI	红利低波(全)	2005-12-30
000824.CSI	中证国企红利	H00824.CSI	中证国企红利全收益	2009-06-30
000825.CSI	中证央企红利	H00825.CSI	中证央企红利全收益	2008-12-31
931468.CSI	红利质量	921468.CSI	红利质量全收益	2004-12-31
H30089.CSI	红利潜力	H20089.CSI	红利潜力全收益	2005-12-30
H30366.CSI	高息策略	H20366.CSI	高息策略全收益	2004-12-31
HSYLV.HI	恒生港股通高股息低波动	HSYLV.DV.HI	恒生港股通高股息低波动指数 R	2010-09-03
000015.SH	红利指数	H00015.SH	红利收益	2004-12-31
399324.SZ	深证红利	CN2324.CNI	深证红利 R	2002-12-31
H30270.CSI	红利价值	H20270.CSI	红利价值(全)	2005-12-30
931130.CSI	红利成长低波	H21130.CSI	红利成长低波全收益	2005-12-30

数据来源：wind、西南证券整理，数据截至 2024/06/30

表 4：红利主题相关指数跟踪产品

代码	简称	跟踪产品代码	跟踪产品简称	成立期	基金经理	规模(亿元)
000922.CSI	中证红利	515180.OF	易方达中证红利 ETF	2019/11/26	林伟斌,宋剑贤	65.94
H30269.CSI	红利低波	512890.OF	华泰柏瑞红利低波动 ETF	2018/12/19	柳军	59.23
000824.CSI	中证国企红利	561060.OF	华安中证国有企业红利 ETF	2023/9/6	苏卿云	0.39
000825.CSI	中证央企红利	561580.OF	华泰柏瑞中证中央企业红利 ETF	2023/5/18	李茜	4.45
931468.CSI	红利质量	159758.OF	华夏中证红利质量 ETF	2021/12/20	张弘弢	5.09
H30089.CSI	红利潜力	007671.OF	建信中证红利潜力指数 A	2019/9/11	赵云煜	0.54
H30366.CSI	高息策略	563180.OF	银华中证高股息策略 ETF	2024/5/6	张亦驰,谭跃峰	2.44
HSYLV.HI	恒生港股通高股息低波动	513950.OF	富国恒生港股通高股息低波动 ETF	2023/4/19	田希蒙	2.44
000015.SH	红利指数	510880.OF	华泰柏瑞红利 ETF	2006/11/17	柳军,李茜	177.36
399324.SZ	深证红利	159905.OF	工银深证红利 ETF	2010/11/5	赵栩	25.85
H30270.CSI	红利价值	--	--	--	--	--
931130.CSI	红利成长低波	--	--	--	--	--

数据来源：wind、西南证券整理，数据截至 2024/06/30

对于上述 12 类红利主题资产，其中有 10 类资产包含可投资的 ETF 产品，红利价值指数及红利成长低波指数暂无相关跟踪的 ETF 产品。

针对上述待选的 12 个红利主题相关指数，本文计算了其日度收益率的相关性。整体而言，各个红利资产相关性保持在 70% 以上，恒生港股通高股息低波动与其他资产相关性相对较低，相关性维持在 70% 以下。

表 5：红利资产相关性

	红利质量	深证红利	红利指数	中证国企红利	中证央企红利	中证红利	红利潜力	红利低波	红利价值	高息策略	红利成长低波	恒生港股通高股息低波动
红利质量	1.00	0.90	0.82	0.75	0.66	0.88	0.86	0.86	0.85	0.64	0.87	0.53
深证红利	0.90	1.00	0.86	0.81	0.76	0.90	0.94	0.87	0.87	0.76	0.86	0.63
红利指数	0.82	0.86	1.00	0.96	0.93	0.97	0.80	0.94	0.95	0.90	0.92	0.63
中证国企红利	0.75	0.81	0.96	1.00	0.96	0.96	0.81	0.97	0.97	0.90	0.92	0.65
中证央企红利	0.66	0.76	0.93	0.96	1.00	0.90	0.75	0.93	0.94	0.92	0.87	0.66
中证红利	0.88	0.90	0.97	0.96	0.90	1.00	0.84	0.96	0.96	0.83	0.94	0.65
红利潜力	0.86	0.94	0.80	0.81	0.75	0.84	1.00	0.82	0.80	0.77	0.78	0.65
红利低波	0.86	0.87	0.94	0.97	0.93	0.96	0.82	1.00	0.99	0.90	0.95	0.66
红利价值	0.85	0.87	0.95	0.97	0.94	0.96	0.80	0.99	1.00	0.91	0.94	0.66
高息策略	0.64	0.76	0.90	0.90	0.92	0.83	0.77	0.90	0.91	1.00	0.82	0.69
红利成长低波	0.87	0.86	0.92	0.92	0.87	0.94	0.78	0.95	0.94	0.82	1.00	0.63
恒生港股通高股息低波动	0.53	0.63	0.63	0.65	0.66	0.65	0.65	0.66	0.66	0.69	0.63	1.00

数据来源：wind、西南证券整理，数据截至 2024/06/30

传统的资产配置策略通常基于主观逻辑及相对宏观的指标进行资产择时，此类策略主要痛点在于宏观指标数据量较少，且主观逻辑容易受主观经验制约，难以最大化利用数据信息，且针对较高相似度的各类资产，传统的资产配置策略难以得到行之有效的策略。为进一步剔除主观因素的影响，捕捉各类红利资产的细微差异，本文将通过机器学习算法构建相应红利指数配置策略。

3.1 决策树算法简介

决策树是一类目前应用广泛且较为经典的机器学习模型，其原理在于决策树通过递归地选择最优特征进行分割来构建。最优特征的选择通常基于某种度量标准，如信息增益、基尼指数或方差减少。每次分割都会将数据集分成更小的子集，直到满足停止条件（如达到最大深度或子集纯度）为止。

针对决策树模型，常见的算法包括 ID3、C4.5 以及 CART 算法。其中 ID3 是决策树算法的早期版本，主要基于信息增益来选择划分属性。信息增益是通过比较划分前后的熵差来衡量的，其中熵是衡量数据集不确定性的指标。ID3 算法通过计算每个特征的信息增益，选择信息增益最大的特征进行划分，从而构建决策树。而 C4.5 算法是 ID3 算法的改进版，引入了增益率来选择划分属性，并且能够处理连续属性，通过剪枝技术来防止过拟合，解决了 ID3 算法不能处理连续属性、以及在构建决策树时容易产生过拟合的问题。CART 算法采用基尼不纯度作为划分属性的标准，通过递归地将数据集分割成更小的子集，直到满足停止条件为止，可用于分类和回归问题。

单一的决策树有时难以达到令人满意的预测效果，而 Boosting 作为一种集成学习方法可解决单一决策树预测效果不佳的方法。Boosting 通过结合多个弱学习器（weak learner）

来形成一个强学习器（strong learner），具体方式为顺序地训练弱学习器，并且每个新的学习器都尝试纠正前一个学习器的错误，来逐渐提高整体的预测精度。

常见的 Boosting 算法包括 GBDT (Gradient Boosting Decision Tree) 及 DART (Dropouts meet Multiple Additive Regression Trees)，其中 GBDT 主要原理为将前一棵树的预测残差（或损失函数的梯度）作为后一棵树的目标值，如此迭代，来优化回归模型的性能。而 DART 是 GBDT 的一个变种，它将深度学习的 dropout 技术引入梯度提升树模型中，通过随机丢弃部分树，DART 旨在减少模型过拟合的风险，并提高模型的泛化能力。

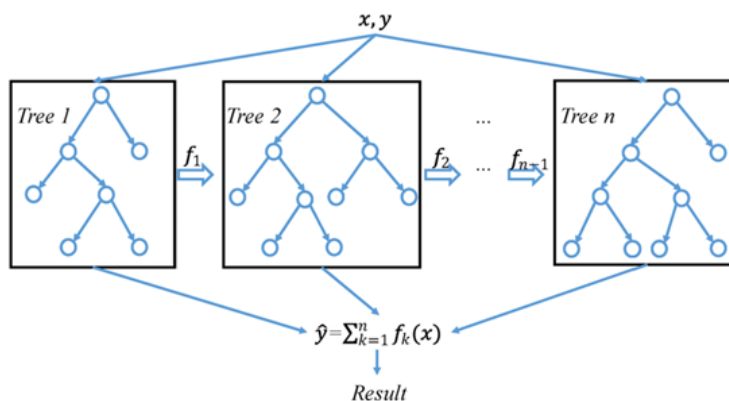
以上集成学习思路具备多种实现模型，本节着重介绍其中两类应用较为经典的 XGBoost 以及 LightGBM。

● XGBoost

XGBoost (eXtreme Gradient Boosting) 是一种高效、灵活且可扩展的机器学习算法，由陈天奇 (Tianqi Chen) 在 2016 年提出。它基于 GBDT 的框架实现，并在此基础上进行了多种优化，显著提升了计算速度和模型性能。其在提高模型准确率、加快训练速度和防止过拟合等方面拥有独特优势，在各类数据挖掘和机器学习任务中广受青睐。该算法具备如下特点：

- 1) **正则化**：XGBoost 引入了 L1 (Lasso) 和 L2 (Ridge) 正则化项，可以有效防止模型过拟合，提升模型的泛化能力。
- 2) **加权的量化特征 (Weighted Quantile Sketch)**：该方法可以高效处理稀疏数据，并支持分布式计算，使得 XGBoost 在处理大规模数据时表现尤为出色。
- 3) **并行处理**：通过分裂点并行化 (Split Finding Algorithm)，XGBoost 可以充分利用多核处理器，显著加速模型训练过程。
- 4) **树结构优化**：XGBoost 使用了基于贪心算法的近似算法 (Greedy Algorithm)，进行最佳分裂点的搜索，同时引入了剪枝 (Pruning) 策略，进一步优化树结构，提高模型性能。
- 5) **支持自定义损失函数**：XGBoost 允许用户自定义损失函数，灵活适应不同的业务需求和应用场景。

图 4：XGBoost 算法原理示意



数据来源：wind、西南证券整理

XGBoost 核心算法在于：**首先**，不断地添加树，进行特征分裂来生长一棵树。每次添加一棵树，学习一个新函数 $f(x)$ ，去拟合上次预测的残差；**其次**，当训练完成得到 k 棵树，就根据这个样本的特征，在每棵树中落到对应的一个叶子节点，每个叶子节点就对应一个分数；**最后**，将每棵树对应的分数加起来就是该样本的预测值。

XGBoost 的目标函数由两部分组成：

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

n 是样本数量， y_i 是第 i 个样本的真实值， \hat{y}_i 是模型的预测值， K 是树的数量， f_k 是第 k 棵树的预测结果。针对目标函数的不同组成部分。其中 $l(\hat{y}_i, y_i)$ 表示损失函数，用于衡量预测值 \hat{y}_i 与真实值 y_i 之间的差异，常用的损失函数包括均方误差 (MSE)、对数损失 (Log Loss) 等； $\Omega(f_k)$ 是正则化项，目的是控制模型复杂度，防止过拟合，共包含两个部分（叶子节点数和叶子节点权重的惩罚项）：

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

其中 T 是树的叶子节点数， w_j 是第 j 个叶子节点的权重， γ 和 λ 是正则化参数，用于控制模型复杂度。

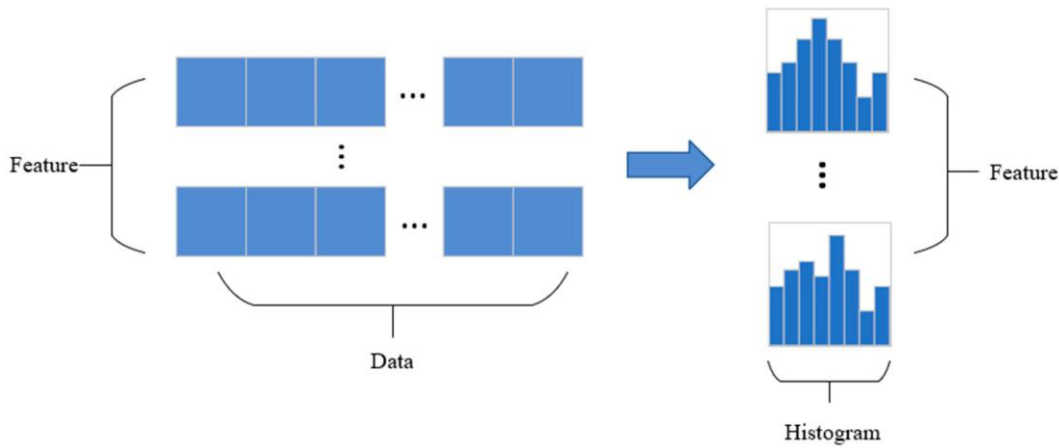
● LightGBM

XGBoost 推出后在多种不同任务类型体现出了极强的预测精度，但训练时长和内存消耗相比之前模型不占优势。而 LightGBM (Light Gradient Boosting Machine) 就是一种由微软提出的解决 GBDT 在遇到海量数据训练速度慢的算法。该算法具备如下特点：

- 1) **计算速度快**：LightGBM 使用直方图算法来加速决策树的构建，并采用基于叶子增长 (Leaf-wise growth) 策略，使其在处理大规模数据时具有显著的速度优势。
- 2) **高效的内存使用**：通过直方图算法和高效的数据存储方式，LightGBM 可以大幅减少内存占用，提高计算效率。
- 3) **支持并行和分布式计算**：LightGBM 支持多线程并行计算和分布式计算，能够在多台机器上进行模型训练，加速模型的构建过程。在工程实现上，采用了包括特征并行、数据并行、投票并行和缓存优化等方式。
- 4) **灵活性强**：LightGBM 支持多种损失函数和评估指标，能够适应不同类型的任务需求，如回归、分类和排序等。

直方图算法 (Histogram) 基本思想：直方图算法的基本思想是先把连续的浮点特征值离散化成 k 个整数，同时构造一个宽度为 k 的直方图。在遍历数据的时候，根据离散化后的值作为索引在直方图中累积统计量，当遍历一次数据后，直方图累积了需要的统计量，然后根据直方图的离散值，遍历寻找最优的分割点。该算法可以有效降低内存消耗，减少计算代价。

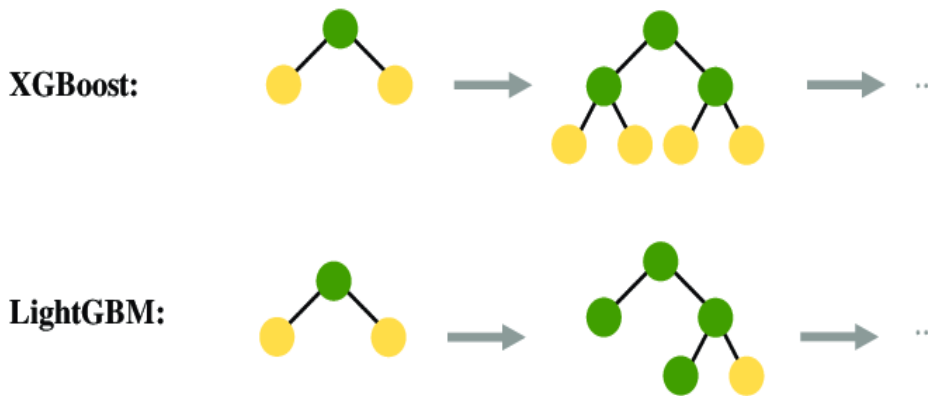
图 5: LightGBM_直方图算法示意



数据来源: 西南证券整理

Leaf-wise 叶子生长策略: LightGBM 抛弃了大多数 GBDT 工具使用的按层生长 (level-wise) 的决策树生长策略, 而使用了带有深度限制的按叶子生长 (leaf-wise) 算法。Leaf-wise 每次从当前所有叶子中, 找到分裂增益最大的一个叶子, 然后分裂, 如此循环。因此同 Level-wise 相比, 在分裂次数相同的情况下, Leaf-wise 可以降低更多的误差, 得到更好的精度。Leaf-wise 的缺点是可能会长出比较深的决策树, 产生过拟合。因此 LightGBM 在 Leaf-wise 之上增加了一个最大深度的限制, 在保证高效率的同时防止过拟合。

图 6: LightGBM_ Leaf-wise 示意



数据来源: 西南证券整理

LightGBM 的目标函数与其他梯度提升模型相似, 同样包含两部分: 损失函数和正则化项, 即可以表示为:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

其中, n 是样本数量, y_i 是第 i 个样本的真实值, \hat{y}_i 是模型的预测值, l 是损失函数, K 是树的数量, f_k 是第 k 棵树的预测结果, Ω 是正则化项。

在后文叙述中, 本文将使用 LightGBM 算法作为红利主题资产配置模型。

3.2 基于 LightGBM 的红利指数配置策略

本文选取前述包括中证红利指数在内共计 12 个红利主题指数作为红利指数配置标的，采用 **DART 增强方式的 LightGBM 算法** 作为红利主题资产配置模型，具体模型设定如下：

训练特征：包含滚动股息率、日频量价及相关技术指标等在内共计 76 个特征；

训练标签：指数未来一个月累计收益（22 交易日，绝对收益）。

数据集划分：2012-01-01 至 2018-12-31 为训练集；2019-01-01 至 2020-12-31 为验证集；2021-01-01 至 2024-06-30 为测试集。暂不考虑滚动训练方式；

数据处理：单个特征截面 Z-Score 标准化，标签截面 MinMax 归一化

表 6：LightGBM 模型数据集划分

数据集类型	数据区间
训练集	2012-01-01 至 2018-12-31
验证集	2019-01-01 至 2020-12-31
测试集	2021-01-01 至 2024-06-30

数据来源：wind、西南证券整理

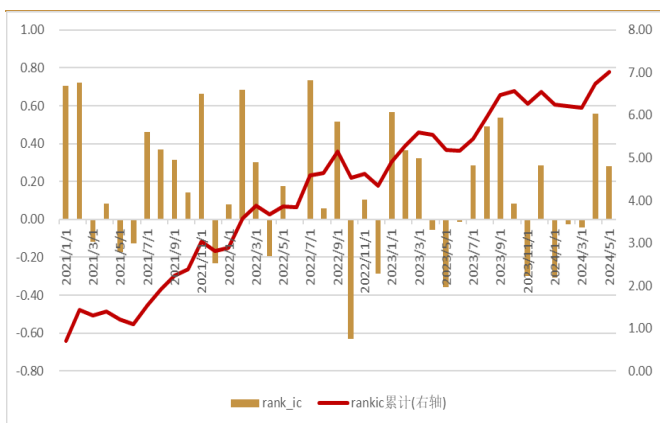
本文根据模型在验证集的表现，利用贝叶斯优化的方式进行超参数调优，并在训练集训练模型得到红利主题指数因子，并每期选择因子值排名前 3 的红利主题指数等权配置，作为多头配置组合。值得注意的是，在模型训练过程中，本文采取的是日频训练，及训练样本的采样频次为日频，而实际调仓的频次为月频。

表 7：红利主题因子回测情况

	IC	ICIR	RankIC 胜率	年化收益	年化波动	信息比率	最大回撤
红利主题因子	17.11%	1.72	63.41%	15.39%	16.44%	0.94	17.34%
红利主题因子-调整后	18.92%	1.68	63.41%	14.09%	16.03%	0.88	14.32%

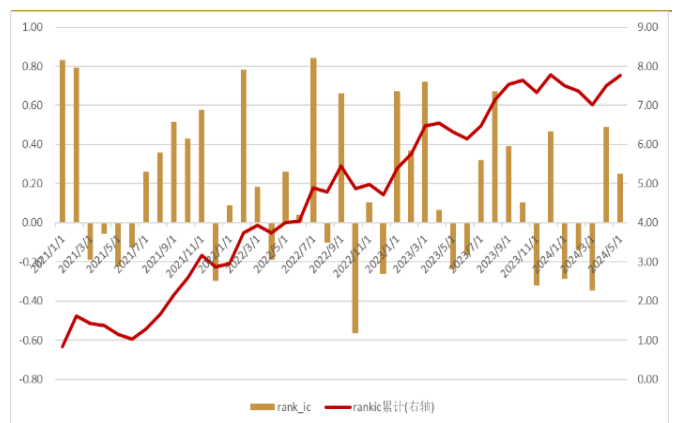
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 7：红利主题因子 RankIC



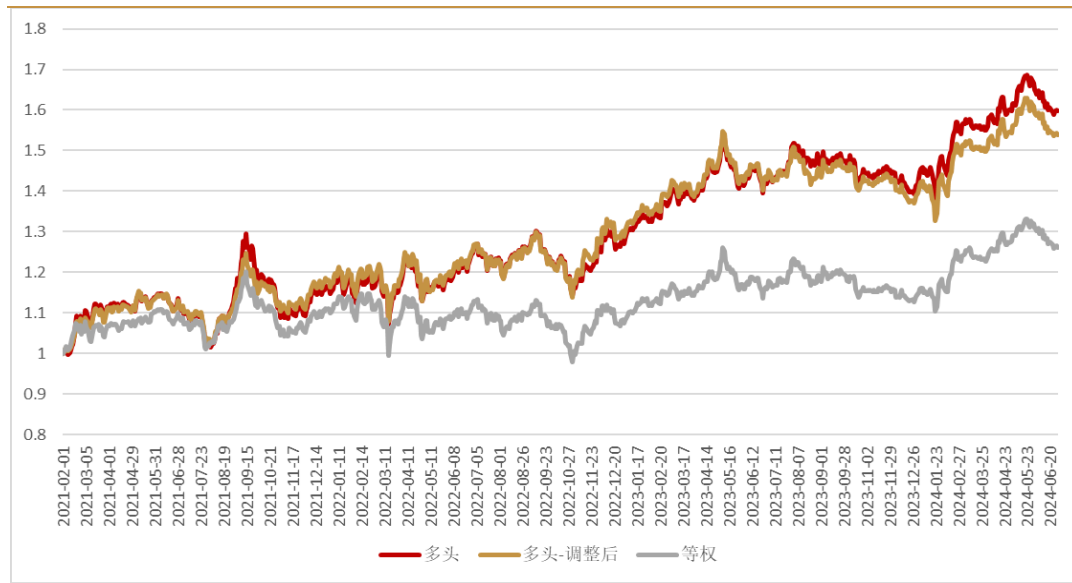
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 8：红利主题因子 RankIC-调整后



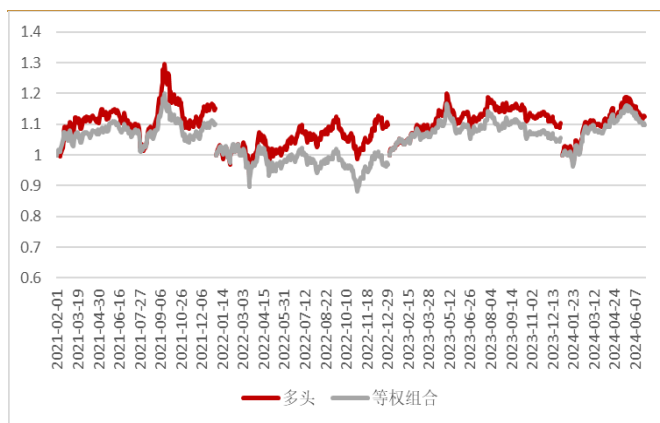
数据来源：wind、西南证券整理，数据截至 2024/06/30

经回测，红利主题指数因子自 2021 年 1 月至 2024 年 6 月 RankIC 为 17.11%，RankIC 胜率为 63.41%，多头组合（Top3）年化收益 15.39%，相对等权组合（7.33%）年化超额收益率 8.06%，最大回撤率 17.34%。剔除无跟踪 ETF 产品指数后，红利主题指数因子 RankIC 为 18.92%，RankIC 胜率 63.41%，多头组合（Top3）年化收益 14.09%，相对等权组合年化超额收益率 6.76%，最大回撤率 14.32%。

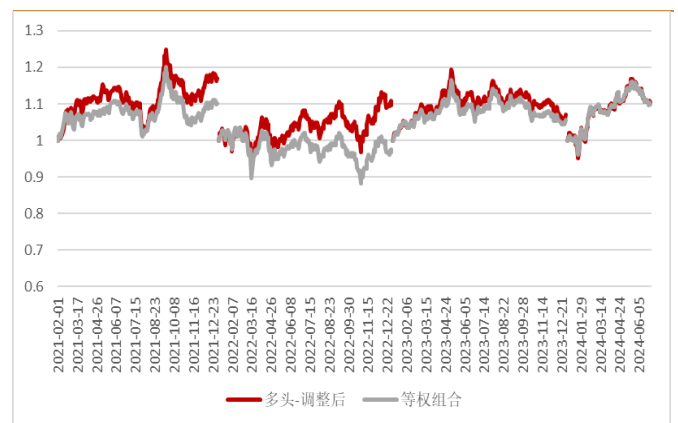
图 9：红利主题因子多头组合净值


数据来源：wind、西南证券整理，数据截至 2024/06/30

分年度来看，自 2021 年至 2024 年 6 月 30 日，红利主题因子多头组合相对等权组合平均年化超额收益率 7.75%，今年以来年化超额收益率 6.67%；剔除无跟踪 ETF 产品的指数后，红利主题因子多头组合相对等权组合平均年化超额收益率 6.43%，今年以来年化超额收益率 1.18%。

图 10：红利主题因子多头分年度表现


数据来源：wind、西南证券整理，数据截至 2024/06/30

图 11：红利主题因子多头分年度表现-调整后


数据来源：wind、西南证券整理，数据截至 2024/06/30

表 8：红利主题因子多头组合分年度表现

	年份	年化超额收益率	年化波动	信息比率	最大回撤率
红利主题因子	2021	5.80%	18.49%	0.94	16.26%
	2022	13.58%	19.24%	0.53	12.34%
	2023	4.94%	11.92%	0.90	9.30%
	2024	6.67%	14.29%	2.03	5.82%
红利主题因子-调整后	2021	7.76%	15.95%	1.21	11.87%
	2022	15.27%	19.57%	0.61	12.34%
	2023	1.50%	12.11%	0.61	11.48%
	2024	1.17%	14.84%	1.59	6.90%

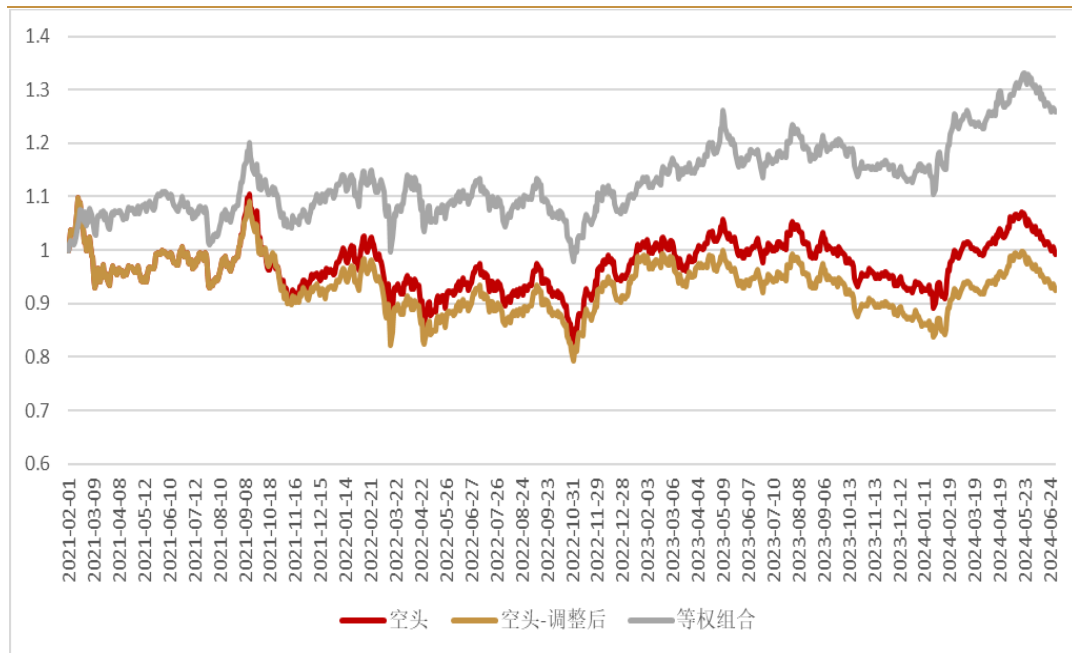
数据来源：wind、西南证券整理，数据截至 2024/06/30

此外，红利指数因子空头(Bottom3)自 2021-01-01 至 2024-06-30 年化收益率为-0.22%，剔除无 ETF 跟踪指数后红利指数因子空头 (Bottom3) 年化收益率为-2.36%。

表 9：红利主题因子空头表现

	年化收益	年化波动	信息比率	最大回撤
空头(bottom3)	-0.22%	16.86%	-0.01	25.20%
空头-调整后(bottom3)	-2.36%	16.83%	-0.14	27.84%

数据来源：wind、西南证券整理，数据截至 2024/06/30

图 12：红利主题因子空头组合净值


数据来源：wind、西南证券整理，数据截至 2024/06/30

4 机器学习应用二：选股因子动态复权

4.1 泛高股息股票池构建

囿于中证红利指数成分股数量仅有 100 只，较少的成分股数量较少，因此本文制定相关红利股票池筛选条件，旨在于全市场范围内筛选出流动性较好，分红能力较好，分红稳定性较强的红利股票池，命名**泛高股息股票池**。具体筛选条件如下（以 252 个交易日表征一年）：

- **基本筛选条件：**

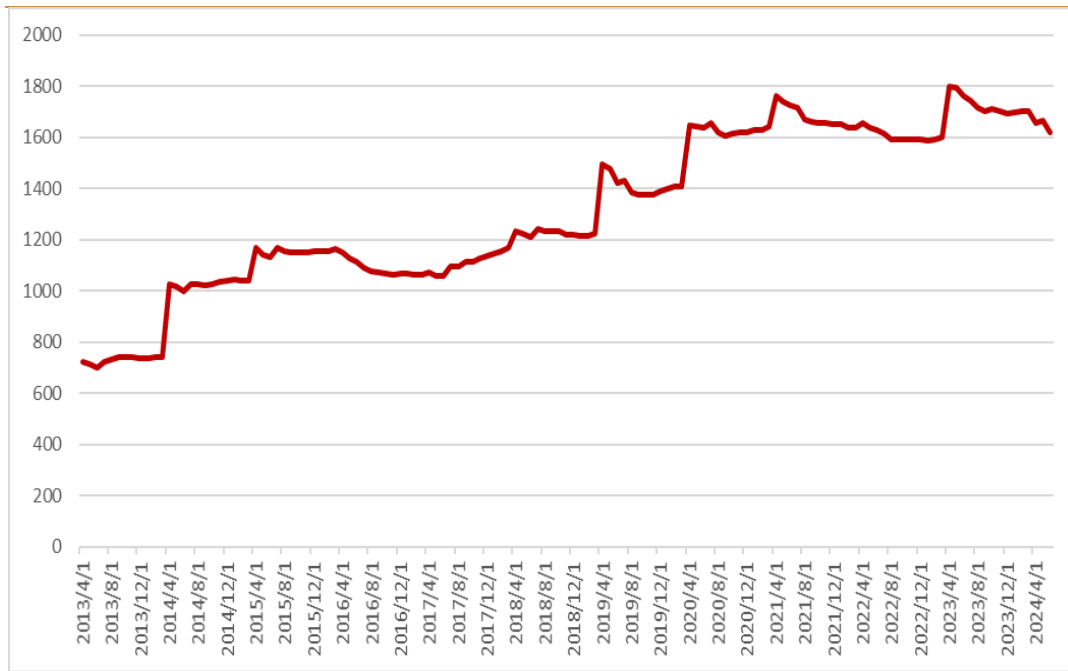
- 1) 剔除 ST 个股；
- 2) 过去一年日均市值排名前 80%；
- 3) 过去一年日均成交额前 80%。

- **分红筛选条件：**

- 1) 近三年连续现金分红；
- 2) 近一年滚动股息率大于 0；
- 3) 近一年股利支付率大于 0 小于 1。

以上述“**基本+分红**”为筛选条件，按照月频（每月最后一个交易日收盘）调仓。自 2013 年 4 月 30 日，红利高股息股票池中 共有 724 只股票，截至 2024 年 6 月 30 日，泛高股息股票池中 共有 1619 只股票，该股票池中股票数量在逐期提升。

图 13：红利高股息股票池各期数量



数据来源：wind、恒生聚源、西南证券整理，数据截至 2024/06/30

参照中证红利指数加权方式，本文以股息率加权构建泛高股息指数。自2013年4月30日至2024年6月30日，泛高股息指数年化收益率11.05%，中证红利全收益指数年化收益率12.14%，万得全A年化收益率6.18%。整体来看，泛高股息指数相对跑输中证红利全收益指数，但整体跑赢万得全A。

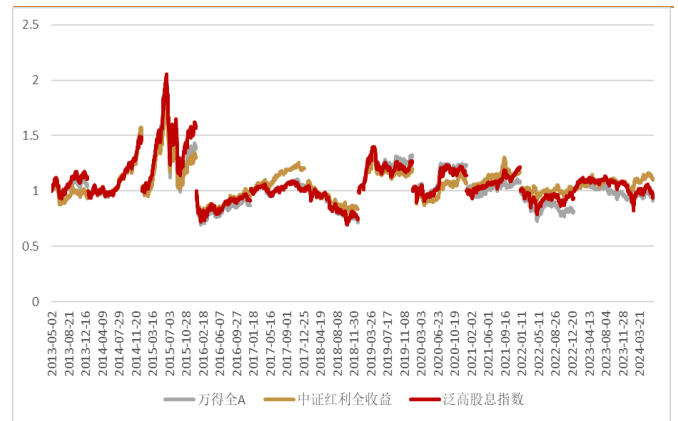
表 10：泛高股息指数业绩对比（全区间）

	年化收益率	年化波动	信息比率	最大回撤率
泛高股息指数	11.05%	23.29%	0.47	50.60%
中证红利全收益	12.14%	21.28%	0.57	45.66%
万得全A	6.18%	24.00%	0.26	55.99%

数据来源：wind、西南证券整理，数据截至2024/06/30

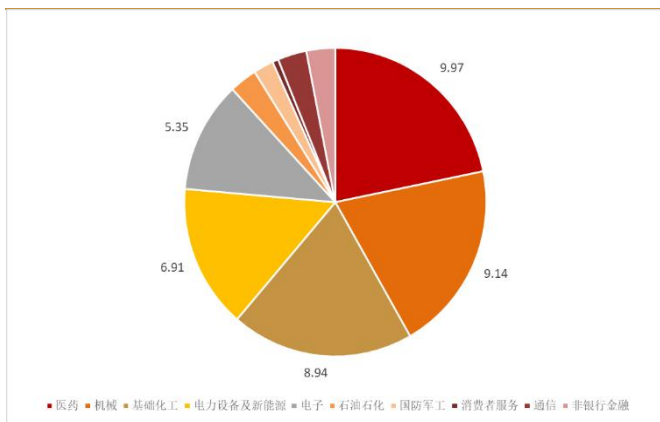
图 14：泛高股息指数净值（全区间）


数据来源：wind、西南证券整理，数据截至2024/06/30

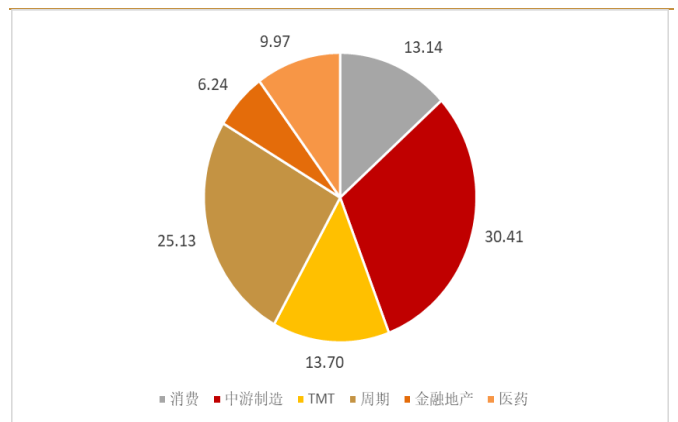
图 15：泛高股息指数净值（分年度）


数据来源：wind、西南证券整理，数据截至2024/06/30

截至最新一期2024年6月30日，按照中信一级行业分类来看，泛高股息指数配置比较高的行业分别为医药、机械、基础化工、电新行业，占比分别为9.97%、9.14%、8.94%、6.91%；消费、中游制造、TMT、周期、金融地产、医药板块配置比例分别为13.14%、30.41%、13.70%、25.13%、6.24%、9.97%。

图 16：泛高股息指数行业配置情况（%）


数据来源：wind、西南证券整理，数据截至2024/06/30

图 17：泛高股息指数板块配置情况（%）


数据来源：wind、西南证券整理，数据截至2024/06/30

表 11：泛高股息指数业绩（分年度）

年份	年化收益率	年化波动	信息比率	最大回撤率	超额(相对中证红利)	超额(相对万得全 A)
2013	12.46%	21.56%	0.58	17.02%	17.91%	6.36%
2014	45.77%	17.95%	2.55	9.44%	-8.82%	-4.17%
2015	59.73%	44.23%	1.35	44.37%	28.58%	19.16%
2016	-8.60%	27.69%	-0.31	28.00%	-3.84%	4.51%
2017	0.40%	11.82%	0.03	9.94%	-20.12%	-4.00%
2018	-25.47%	22.68%	-1.12	33.68%	-9.03%	3.09%
2019	27.22%	21.43%	1.27	19.54%	7.11%	-5.21%
2020	14.37%	22.77%	0.63	13.76%	6.98%	-9.06%
2021	21.30%	13.87%	1.54	9.03%	4.28%	12.78%
2022	-7.17%	21.17%	-0.34	22.27%	-6.14%	11.81%
2023	4.97%	11.98%	0.41	10.68%	-0.82%	11.08%
2024	-5.57%	24.38%	-0.23	17.82%	-15.97%	2.83%

数据来源：wind、西南证券整理，数据截至 2024/06/30

自 2013 年至 2024 年共计 12 个年份以来，相对中证红利全收益指数，泛高股息指数年化超额收益胜率为 41.67% (5/12)，平均年化超额收益率 0.01%；相对万得全 A 指数，泛高股息指数年化超额收益胜率为 66.67% (8/12)，平均年化超额收益率 4.10%。

4.2 泛高股息选股因子探究

基于前述泛高股息股票池的构建，本节将在泛高股息股票池中筛选表现较为优秀的传统因子，并致力于构建表现优秀的特殊因子，以赋能红利高股息投资策略。本文分别从**质量、估值、盈利、成长、量价**五大类中筛选因子分别进行测试，并筛选表现相对较好的基本因子。其中各大类因子下的细分因子信息如下：

表 12：基本因子分类及描述

大类	因子名称	因子助记符	因子计算方式
成长	EPS 同比	EPSYOY	EPS 同比增长率
	净利润同比	NetProfitGrow Rate	净利润同比增长率
	ROE 环比	ROETMMOM	ROE 季度环比增长率
	归母净利润同比	NetProfitGrow Rate	归母净利润同比增长率
质量	归母净利润比负债	SEwithoutMIToTL	财务报表归母净利润比负债
	速动比率	QuickRatio	财务报表速动比率
	流动比率	CurrentRatio	财务报表流动比率
	ROE 标准差	ROE_std	过去三年 ROE 均值
	ROE 均值比标准差	ROE_meanstd	过去三年 ROE 均值/标准差
	净资产负债率	NetAssetLiabilityRatio	财务报表净资产负债率
盈利	营业利润率	OperatingProfitToTOR	财务报表营业利润率
	净利润率	NPTtoTORTTM	财务报表净利润率
	ROETTM	ROETTM	滚动净资产收益率
	ROIC	ROIC	资本回报率

大类	因子名称	因子助记符	因子计算方式
估值	PB 倒数	1/PB	市净率倒数
	PE 倒数	1/PE	市盈率倒数
	PS	PS	市销率
	对数市值	lnMV	当前市值对数值
	股息率	DividendRatioTTM	滚动股息率
	PE 比营收三年增长率	PETMTtoORRate	滚动市盈率/营业收入三年复合增长率
量价	20 日波动	Std20	过去 20 日收益率标准差
	20 日上行波动	Positive20std	过去 20 日正收益率标准差
	20 日下降波动	Negative20std	过去 20 日负收益率标准差
	60 日波动	Std60	过去 60 日收益率标准差
	60 日上行波动	Positive60std	过去 60 日正收益率标准差
	60 日下降波动	Negative60std	过去 60 日负收益率标准差
	20 价格动量	Excess20	过去 20 日超额收益
	60 价格动量	Excess60	过去 60 日超额收益
	90 价格动量	Excess90	过去 90 日超额收益
	120 日价格动量	Excess120	过去 120 日超额收益

数据来源: wind、西南证券整理

根据上述五大类因子共计 30 个小类因子, 本文分别进行因子测试, 其中因子测试的相关基本设定如下:

回测区间: 2013 年 4 月 30 日至今。

调仓频次: 月频, 每月最后一个交易日收盘后调仓。

因子处理: 3 倍 sigma 去极值, 除市值因子外均进行行业市值标准化。

表 13: 五大类基本因子测试

大类	因子名称	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
成长	EPS 同比	1.88%	0.76	9.46%	26.73%	0.35	56.20%	-1.58%
	净利润同比	1.47%	0.55	8.08%	26.83%	0.30	60.20%	-2.96%
	ROE 环比	2.36%	1.20	11.68%	26.96%	0.43	59.25%	0.64%
	归母净利润同比	1.51%	0.55	7.90%	26.81%	0.29	59.62%	-3.14%
质量	归母净利润比负债	0.66%	0.15	9.89%	28.05%	0.35	68.63%	-1.15%
	速动比率	0.74%	0.18	10.58%	28.60%	0.37	69.63%	-0.46%
	流动比率	0.51%	0.13	11.03%	28.32%	0.39	67.90%	-0.01%
	ROE 标准差	-1.99%	-1.12	9.82%	25.36%	0.39	64.78%	-1.22%
	ROE 均值比标准差	2.40%	1.12	12.94%	22.87%	0.57	47.81%	1.90%
	净资产负债率	-0.66%	-0.15	9.89%	28.05%	0.35	68.55%	-1.15%
盈利	营业利润率	1.34%	0.50	7.47%	22.62%	0.33	51.76%	-3.57%
	净利润率	1.12%	0.44	6.77%	22.57%	0.30	54.67%	-4.27%
	ROETTM	1.48%	0.41	6.79%	23.82%	0.29	48.70%	-4.25%
	ROIC	1.69%	0.51	8.22%	24.49%	0.34	45.20%	-2.82%

大类	因子名称	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
估值	BP	4.12%	0.70	11.49%	23.36%	0.49	40.36%	0.45%
	EP	4.43%	0.91	12.25%	22.51%	0.54	35.58%	1.21%
	PS	-2.74%	-0.57	11.92%	24.64%	0.48	49.08%	0.88%
	对数市值	-2.10%	-0.38	20.48%	29.73%	0.69	55.98%	9.44%
	股息率	4.76%	0.97	14.23%	21.51%	0.66	35.99%	3.19%
	PE/营收三年增长率	-0.58%	-0.22	4.58%	26.90%	0.17	73.17%	-6.46%
量价	20日波动	-6.94%	-1.26	8.67%	14.69%	0.59	61.45%	-2.37%
	20日上行波动	-7.29%	-1.84	9.90%	18.80%	0.53	55.77%	-1.14%
	20日下降波动	-4.10%	-0.83	11.89%	17.43%	0.68	53.87%	0.85%
	60日波动	-7.44%	-1.28	10.80%	15.83%	0.68	57.35%	-0.24%
	60日上行波动	-7.96%	-1.82	12.21%	19.50%	0.63	52.43%	1.17%
	60日下降波动	-5.69%	-1.03	13.32%	16.89%	0.79	44.81%	2.28%
	20价格动量	-4.73%	-1.09	8.34%	29.19%	0.29	61.44%	-2.70%
	60价格动量	-5.29%	-1.12	11.14%	29.53%	0.38	55.92%	0.10%
	90价格动量	-3.78%	-0.77	9.87%	29.28%	0.34	60.20%	-1.17%
	120日价格动量	-3.28%	-0.65	9.16%	28.93%	0.32	64.74%	-1.88%

数据来源: wind、西南证券整理, 数据截至 2024/06/30

经过因子在 IC 和多头组合表现综合来看, **成长因子**中表现较好的因子为 ROE 环比及 EPS 同比, IC 均值分别为 2.36%及 1.88%, 多头组合年化收益率为 11.68%及 9.46%; **质量因子**中表现较好的因子为 ROE 均值比标准差以及 ROE 标准差, IC 均值分别为 2.40%及 -1.99%, 多头年化收益为 12.94%及 9.82%; **盈利因子**表现较好的因子为 ROIC 及营业利润率, IC 均值分别为 1.69%及 1.34%, 多头组合年化收益率分别为 8.22%及 7.47%; **估值因子**表现较好的因子为对数市值及股息率, IC 均值分别为 -2.10%及 4.76%, 多头组合年化收益率分别为 20.48%及 14.23%。量价类表现较好的因子有 60 日上行波动和 60 日价格动量, IC 均值分别为 -7.96%及 -5.29%, 多头组合年化收益分别为 12.21%以及 11.14%。

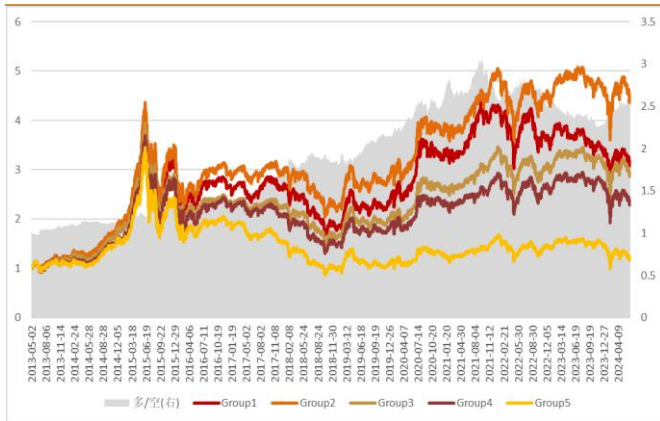
本文在上述五大类因子中分别筛选出表现相对较好的 1-2 个因子如下:

表 14: 各大类因子入选小类因子

大类	因子名称	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
成长	EPS 同比	1.88%	0.76	9.46%	26.73%	0.35	56.20%	-1.58%
	ROE 环比	2.36%	1.20	11.68%	26.96%	0.43	59.25%	0.64%
质量	ROE 标准差	-1.99%	-1.12	9.82%	25.36%	0.39	64.78%	-1.22%
	ROE 均值比标准差	2.40%	1.12	12.94%	22.87%	0.57	47.81%	1.90%
盈利	营业利润率	1.34%	0.50	7.47%	22.62%	0.33	51.76%	-3.57%
	ROIC	1.69%	0.51	8.22%	24.49%	0.34	45.20%	-2.82%
估值	对数市值	-2.10%	-0.38	20.48%	29.73%	0.69	55.98%	9.44%
	股息率	4.76%	0.97	14.23%	21.51%	0.66	35.99%	3.19%
量价	60 日上行波动	-7.96%	-1.82	12.21%	19.50%	0.63	52.43%	1.17%
	60 价格动量	-5.29%	-1.12	11.14%	29.53%	0.38	55.92%	0.10%

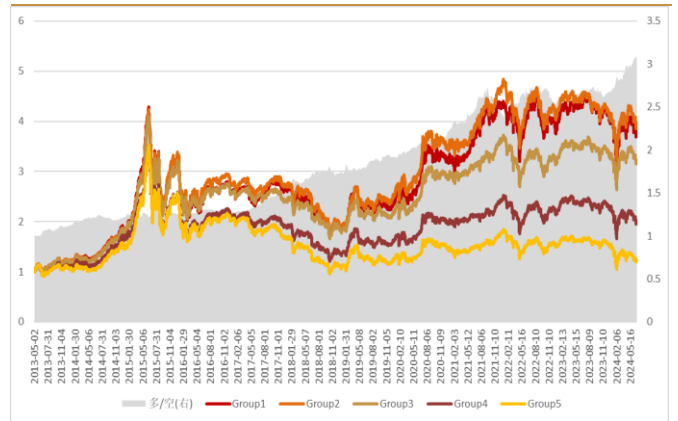
数据来源: wind、西南证券整理, 数据截至 2024/06/30

图 18：成长-EPS 同比分组测试



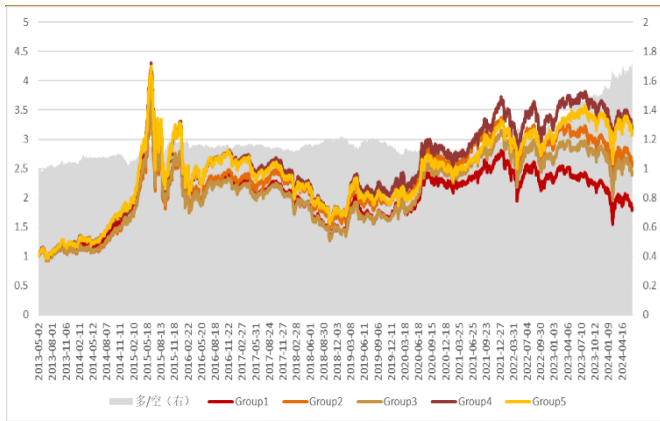
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 19：成长-ROE 环比分组测试



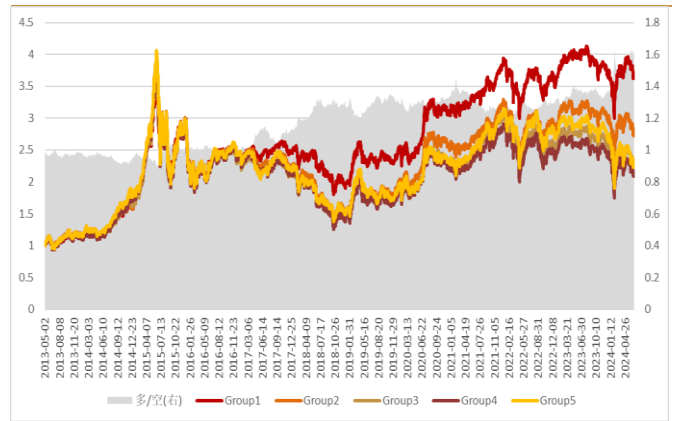
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 20：质量-ROE 标准差分组测试



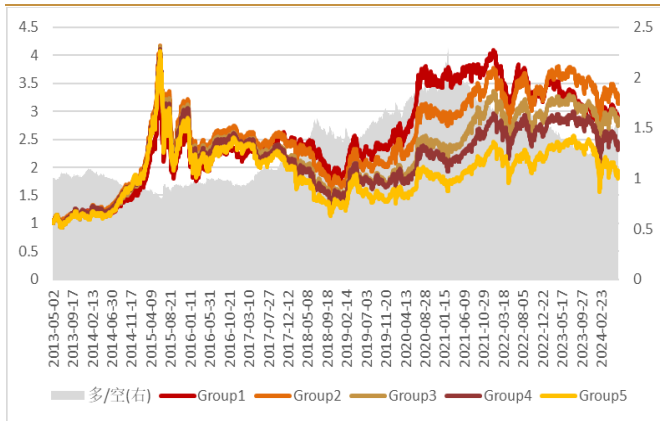
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 21：质量-ROE 均值/标准差



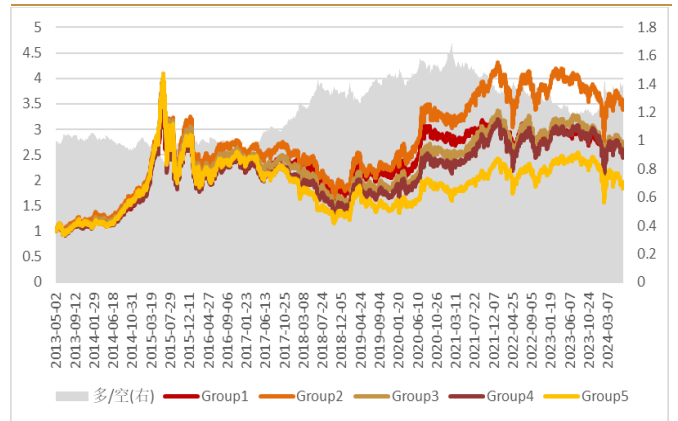
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 22：盈利-ROIC 分组测试

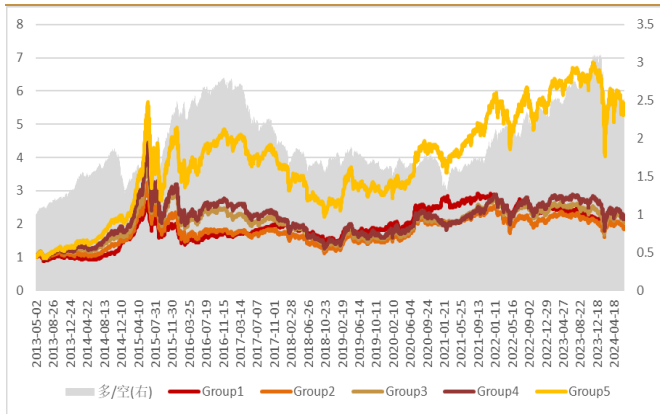


数据来源：wind、西南证券整理，数据截至 2024/06/30

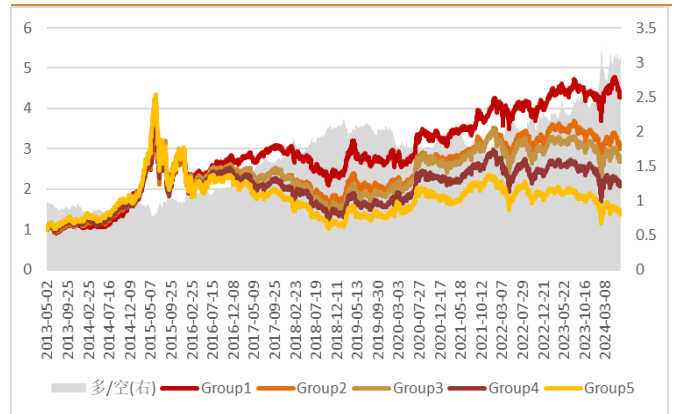
图 23：盈利-营业利润率分组测试



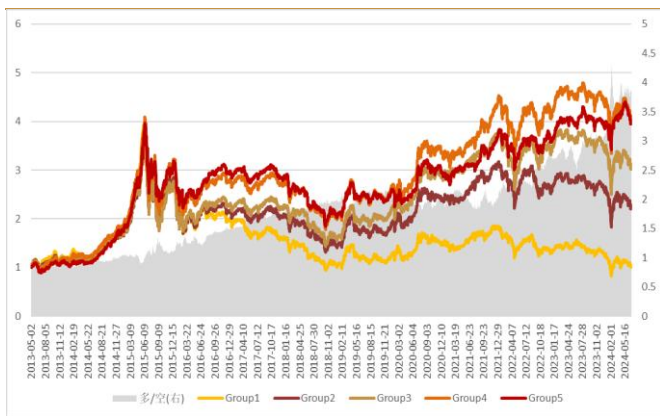
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 24：估值-对数市值分组测试


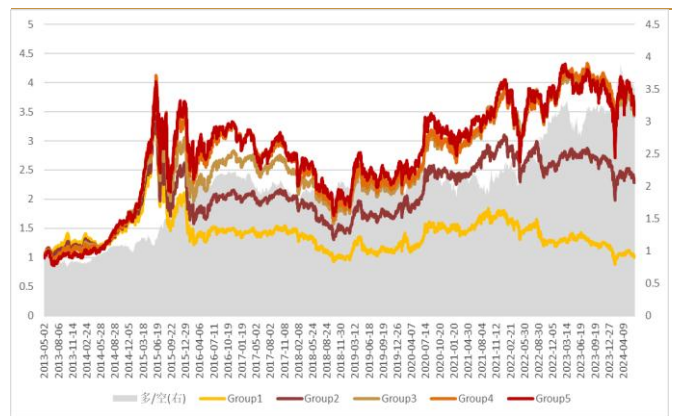
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 25：估值-股息率分组测试


数据来源：wind、西南证券整理，数据截至 2024/06/30

图 26：量价-60 日上行波动分组测试


数据来源：wind、西南证券整理，数据截至 2024/06/30

图 27：量价-60 日价格动量分组测试


数据来源：wind、西南证券整理，数据截至 2024/06/30

● 股息兑现因子

根据本文前一节表述，股息率因子在泛高股息股票池内具备较好的选股效果，其中 RankIC 为 4.76%，多头组合年化收益率 14.23%，相对泛高股息指数年化超额收益率 3.19%。但传统的股息率因子仅考虑截面上股息率对比进行选股，并未考虑到单只股票时序上的股息率变化情况，且未考虑股息率变化是否及时在股票价格的变化有所兑现。

为了综合考虑单只股票股息率时序变化以及股息率变化在股票价格上的兑现情况，本文构建股息兑现因子如下：

$$\text{bias_divprice}_{t,i} = \text{divrank}_{t,i}(n) / \text{exprice}_{t,i}(n, n1)$$

其中 n 表示滚动分位数计算窗口， $n1$ 表示超额收益计算窗口； $\text{bias_divrank}_{t,i}(n)$ 表示 t 时刻股票 i 股息率在过去 n 日的分位数， $\text{exprice}_{t,i}(n, n1)$ 则表示 t 时刻股票 i 滚动 $n1$ 日超额收益在过去 n 日的分位数，因此若当前股息率时序分位数较高且其股价超额分位数较低，判断该股票未来上涨概率更高。

为了探究超额收益计算窗口以及滚动排名分位数的回溯时间窗口对股息兑现因子本身效果的影响，本文分别设定超额收益计算窗口 $n1$ 为滚动 20、60、90 及 120 天、设定排名计算回溯窗口 n 为滚动 20、60、90 及 120 天，并测试不同窗口 $n1$ 及 n 下的因子表现如下：

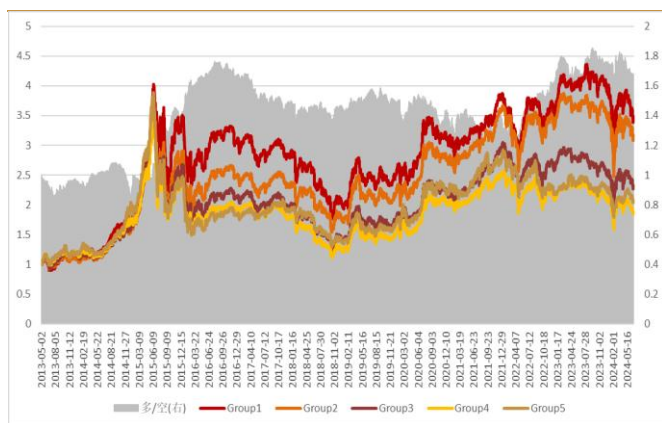
表 15：股息兑现因子回溯窗口测试

大类	因子名称	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
股息兑现	股息兑现 20_20	2.33%	0.68	3.52%	26.88%	0.13	68.54%	-7.53%
	股息兑现 60_20	3.01%	0.83	4.30%	27.80%	0.15	70.80%	-6.74%
	股息兑现 90_20	3.24%	0.86	5.21%	28.19%	0.18	67.33%	-5.84%
	股息兑现 120_20	3.39%	0.91	5.04%	27.46%	0.18	62.09%	-6.01%
	股息兑现 20_60	3.64%	0.95	6.79%	26.93%	0.25	61.46%	-4.25%
	股息兑现 60_60	4.28%	1.12	8.71%	27.18%	0.32	65.34%	-2.33%
	股息兑现 90_60	5.04%	1.19	10.33%	27.43%	0.38	56.24%	-0.72%
	股息兑现 120_60	5.17%	1.27	10.16%	27.18%	0.37	57.87%	-0.88%
	股息兑现 20_90	4.03%	1.05	8.02%	26.81%	0.30	58.66%	-3.02%
	股息兑现 60_90	4.75%	1.20	9.59%	27.14%	0.35	60.44%	-1.45%
	股息兑现 90_90	5.15%	1.20	11.53%	27.27%	0.42	55.63%	0.49%
	股息兑现 120_90	5.13%	1.25	11.56%	26.80%	0.43	54.21%	0.52%
	股息兑现 20_120	4.32%	1.12	8.41%	26.62%	0.32	53.61%	-2.63%
	股息兑现 60_120	5.08%	1.26	10.94%	27.18%	0.40	54.53%	-0.10%
	股息兑现 90_120	4.94%	1.16	13.73%	26.97%	0.51	48.46%	2.69%
	股息兑现 120_120	4.75%	1.16	12.59%	26.50%	0.48	54.10%	1.55%

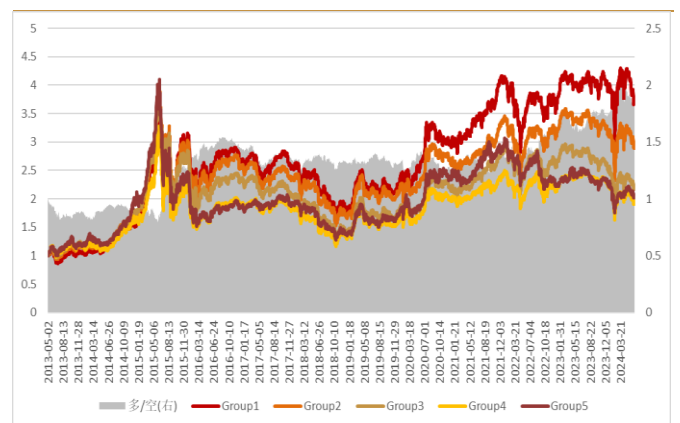
数据来源：wind、西南证券整理，数据截至 2024/06/30

整体来看，随着超额回溯计算窗口 n_1 及滚动分位数计算窗口 n 的扩大，因子表现逐渐提升。其中当 $n_1=120$ ， $n=60$ 时因子 IC 最高为 5.17%，当 $n_1=90$ ， $n=120$ 时因子多头组合表现最好，多头年化收益率为 13.73%，相对泛高股息指数年化超额收益率 2.69%。结合多头组合表现来看，在多头组合跑赢泛高股息指数的前提下， $n_1, n=90, 90$ 及 $n_1, n=120, 90$ 因子 IC 表现同样不俗，分别为 5.15% 及 5.13%。

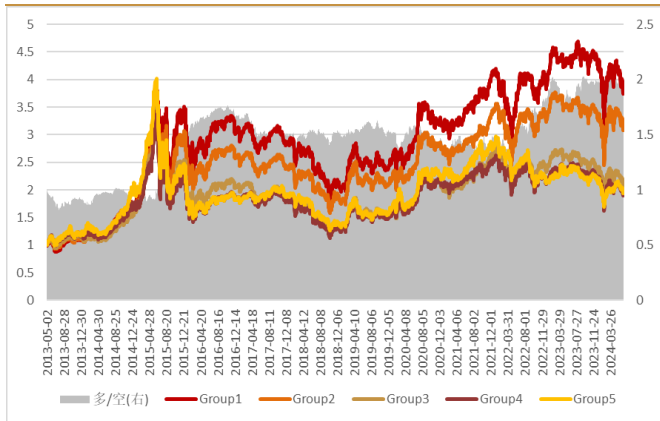
为了进一步选取表现相对稳定的因子参数，本文对上述提到的四组参数下的因子进行分组及多空测试：

图 28：股息兑现 120_60


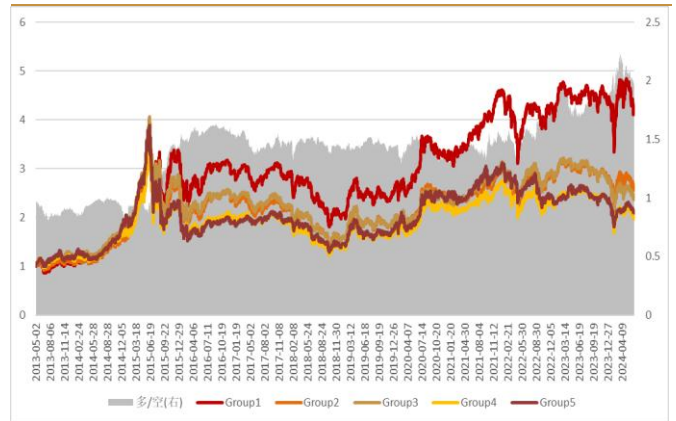
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 29：股息兑现 90_90


数据来源：wind、西南证券整理，数据截至 2024/06/30

图 30：股息兑现 120_90


数据来源：wind、西南证券整理，数据截至 2024/06/30

图 31：股息兑现 90_120


数据来源：wind、西南证券整理，数据截至 2024/06/30

表 16：股息兑现因子多空

	年化收益率	年化波动	信息比率	最大回撤率
股息兑现因子 120_60	3.75%	14.27%	0.26	35.94%
股息兑现因子 90_90	4.54%	14.45%	0.31	28.59%
股息兑现因子 120_90	5.09%	14.11%	0.36	32.56%
股息兑现因子 90_120	5.28%	14.37%	0.37	25.38%

数据来源：wind、西南证券整理，数据截至 2024/06/30

综合因子 IC、多头组合以及多空组合表现，选取超额收益计算窗口 $n1=90$ ，滚动分位数回测窗口 $n=120$ 作为股息兑现价格因子计算参数。

表 17：因子相关性分析

	股价兑现	60 日上行波动	60 价格动量	营业利润率	ROIC	EPS 同比	ROE 环比	对数市值	股息率	ROE 标准差	ROE 均值/标准差
股价兑现	--	-7.5%	-26.3%	0.3%	2.4%	-0.1%	0.0%	-1.4%	1.6%	-0.2%	-0.3%
60 日上行波动	-7.5%	--	35.8%	0.0%	0.0%	0.0%	0.1%	-1.2%	-20.8%	0.0%	3.3%
60 价格动量	-26.3%	35.8%	--	0.5%	0.2%	0.8%	0.1%	6.1%	-6.4%	-0.3%	5.1%
营业利润率	0.3%	0.0%	0.5%	--	-0.1%	0.7%	-0.1%	2.2%	7.7%	-0.2%	0.0%
ROIC	2.4%	0.0%	0.2%	-0.1%	--	0.1%	0.0%	0.0%	10.3%	12.8%	0.0%
EPS 同比	-0.1%	0.0%	0.8%	0.7%	0.1%	--	-0.5%	2.0%	0.1%	-0.2%	0.0%
ROE 环比	0.0%	0.1%	0.1%	-0.1%	0.0%	-0.5%	--	-0.3%	-0.4%	0.1%	0.0%
对数市值	-1.4%	-1.2%	6.1%	2.2%	0.0%	2.0%	-0.3%	--	15.1%	-1.2%	12.0%
股息率	1.6%	-20.8%	-6.4%	7.7%	10.3%	0.1%	-0.4%	15.1%	--	0.0%	15.2%
ROE 标准差	-0.2%	0.0%	-0.3%	-0.2%	12.8%	-0.2%	0.1%	-1.2%	0.0%	--	0.0%
ROE 均值/标准差	-0.3%	3.3%	5.1%	0.0%	0.0%	0.0%	0.0%	12.0%	15.2%	0.0%	--

数据来源：wind、西南证券整理，数据截至 2024/06/30

整体来看，股息兑现因子与量价因子呈现出一定的负相关，与 60 日上行波动、60 日价格动量相关性分别为 -7.5% 及 -26.3%。除此之外，股息兑现因子与其他各因子相关性较低。

本文将上述股息兑现因子作为第六大类因子，并将“量价、盈利、估值、质量、成长”五大类因子中的小类因子分别进行 rank 等权求和，具体求和方式为：

$$\text{factor_level1}_t = \sum_{i=1}^n \text{rank}(\text{factor_level2}_{t,i}) * \text{pool}_{t,i}$$

其中 factor_level1_t 表示 t 时刻的大类因子， $\text{factor_level2}_{t,i}$ 为 t 时刻大类因子的小类因子 i ， $\text{pool}_{t,i}$ 则为泛高股息股票池示性函数，当股票 i 为泛高股息股票，则取值 1，否则取值 nan。

对各大类因子分别进行合成后，各大类因子相较于小类因子在 IC 和多头表现基本都有不同程度的提升。

表 18：小类因子等权合成大类因子测试结果

大类	因子名称	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
小类因子 -等权	股息兑现	4.94%	1.16	13.74%	26.96%	0.51	48.47%	2.69%
	量价	7.85%	1.85	11.52%	25.24%	0.46	53.76%	0.48%
	质量	2.33%	1.36	12.63%	23.83%	0.53	54.20%	1.58%
	估值	5.19%	1.43	17.02%	27.39%	0.62	56.39%	5.97%
	盈利	1.58%	0.51	9.46%	24.33%	0.39	45.66%	-1.59%
	成长	2.42%	1.02	12.32%	26.77%	0.46	54.73%	1.27%

数据来源：wind、西南证券整理，数据截至 2024/06/30

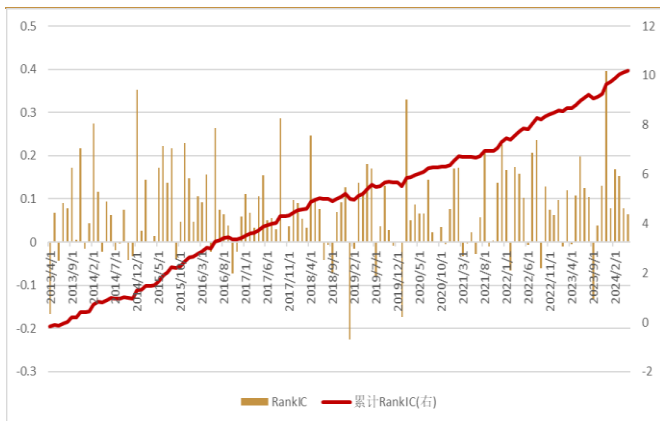
进一步，本文对各大类因子进行等权 rank 求和，得到等权泛高股息因子 IC 为 7.60%，相较于各大类因子有所提升，ICIR（年化）为 2.59，相关性稳定性相较于各大类因子同样有所提升。此外，多头组合年化收益 17.68%，相对泛高股息指数年化超额收益率 6.64%，信息比例 0.72，相较各小类因子有不同程度的提升。

表 19：等权泛高股息因子测试结果

	IC	ICIR	年化收益率	年化波动	信息比率	最大回撤率	年化超额收益率
等权泛高股息因子	7.60%	2.59	17.68%	24.51%	0.72	44.78%	6.64%

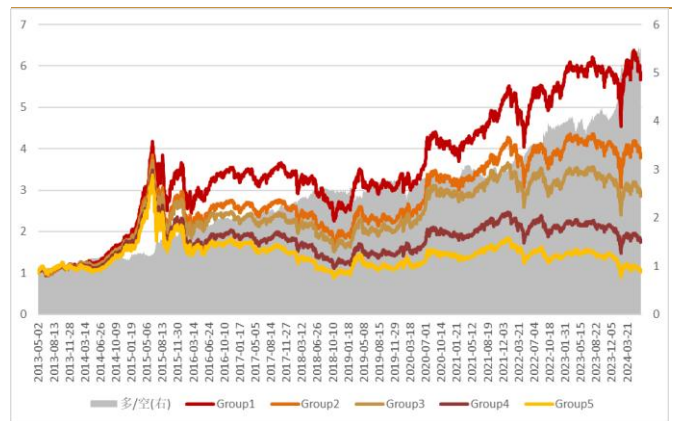
数据来源：wind、西南证券整理，数据截至 2024/06/30

图 32：等权泛高股息因子 IC



数据来源：wind、西南证券整理，数据截至 2024/06/30

图 33：等权泛高股息因子分组



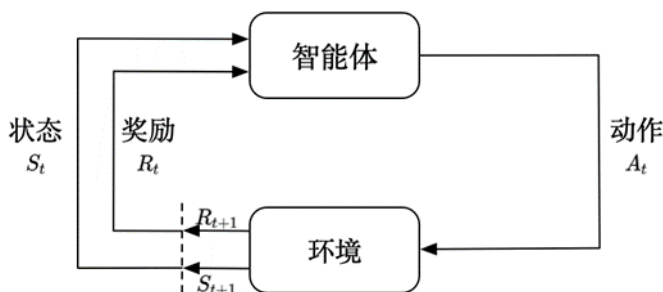
数据来源：wind、西南证券整理，数据截至 2024/06/30

传统多因子策略一般是建立在因子等权重和或因子固定权重和，该策略的理论基础即为将各因子的重要性视作时序不变，而因子的有效性是动态变化的，固定权重和因子往往会忽略因子时序有效性的变化。

4.3 强化学习算法简介

强化学习（Reinforcement Learning, RL），是机器学习的范式和方法论之一，主要研究代理（Agent）如何在所处环境（Environment）中采取行动（Action）使得能够获得最大的奖励（Reward）。强化学习理论受到行为主义心理学启发，侧重在线学习并试图在探索-利用（exploration-exploitation）间保持平衡。不同于监督学习和非监督学习，强化学习不要求预先给定任何数据，而是通过接收环境对动作的奖励（反馈）获得学习信息并更新模型参数。

图 34：强化学习示意



数据来源：西南证券整理

强化学习基本概念：

- 智能体(Agent)：在环境中进行操作的主体，例如机器人、自动驾驶汽车或计算机程序。
- 环境(Environment)：智能体所处的外部世界，智能体在其中进行感知和行动。
- 状态(State,s)：状态是环境的当前快照，它包含了智能体决策所需的所有信息。
- 动作(Action,a)：动作是智能体在特定状态下可以采取的行为。每一个动作都可能导致环境状态的转变，从而引发新的奖励信号。
- 奖励(Reward,r)：奖励是环境对智能体动作的反馈，用于衡量该动作的好坏。奖励可以是正数、负数或零，智能体的目标就是最大化从初始状态开始的累积奖励。
- 策略(Policy, π)：策略是智能体从状态到动作的映射，它指导智能体在给定状态下应该采取何种动作。策略可以是简单的规则，也可以是复杂的函数。
- 值函数(Value Function)：评估某一状态或状态-动作对的价值，通常表示在该状态下采取某一策略所能获得的期望累积奖励。

强化学习主要通过马尔可夫决策（Markov Decision Process, MDP）过程来建模。马尔可夫决策过程为强化学习提供了一个系统的数学框架，帮助定义和解决复杂的决策问题。通过将问题建模为 MDP，强化学习算法可以系统地优化策略，最终达到最大化累积奖励的目标。

马尔可夫决策 MDP 主要由四个要素组成：

- 状态集 (States, S)：所有可能的状态的集合。
- 动作集 (Actions, A)：所有可能的动作的集合。
- 状态转移函数 (State Transition Function, P)：给定当前状态和动作，描述下一个状态的概率分布。即 $P(s' | s, a)$ ，表示在状态 s 执行动作 a 后转移到状态 s' 的概率。
- 奖励函数 (Reward Function, R)：定义在某一状态下或从一个状态到另一个状态时所获得的即时奖励。即 $R(s, a)$ 或 $R(s, a, s')$ 。

强化学习的工作流程：

- 1) 初始化：智能体从某个初始状态开始。
- 2) 选择动作：根据当前状态和策略选择一个动作。
- 3) 执行动作：智能体执行选择的动作，进入下一个状态，并从环境中获得相应的奖励。
- 4) 更新策略：根据获得的奖励和新的状态，更新值函数和策略，以提高未来的决策质量。
- 5) 重复：上述步骤不断重复，直到达到学习目标或满足某个终止条件。

与传统监督学习的异同：

- 1) 数据标签 (label) 依赖性：不需要预先标记的数据。它通过与环境的实时交互来学习，并根据从环境中获得的奖励信号来调整行为。这种交互性使得强化学习能够处理那些难以或不可能获得大量标记数据的任务。
- 2) 实时反馈与探索：强化学习通过尝试不同的行为并观察结果 (奖励或惩罚) 来学习。它鼓励智能体探索未知的行为，以便找到更好的策略。

Q 学习 (Q-Learning) 是一种传统的强化学习算法，通过更新 Q 值 (状态-动作对的值函数) 来学习最优策略。Q 学习是一种无模型的算法，不需要环境模型。在 Q 学习中，我们用 Q 值来量化评价当前行为的价值，其中 Q 值函数的更新基于贝尔曼方程 (Bellman Equation)：

图 35：贝尔曼方程示意

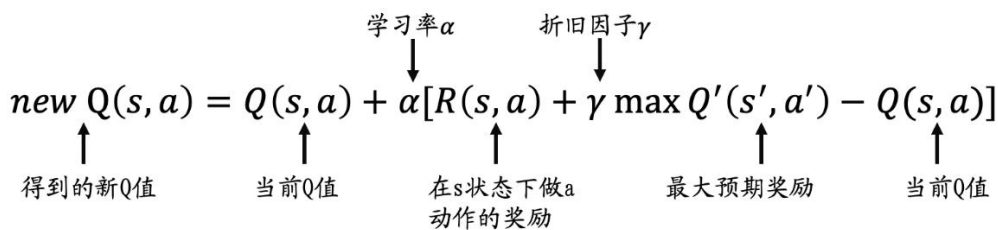
$$\text{new } Q(s, a) = Q(s, a) + \alpha [R(s, a) + \gamma \max_{a'} Q'(s', a') - Q(s, a)]$$


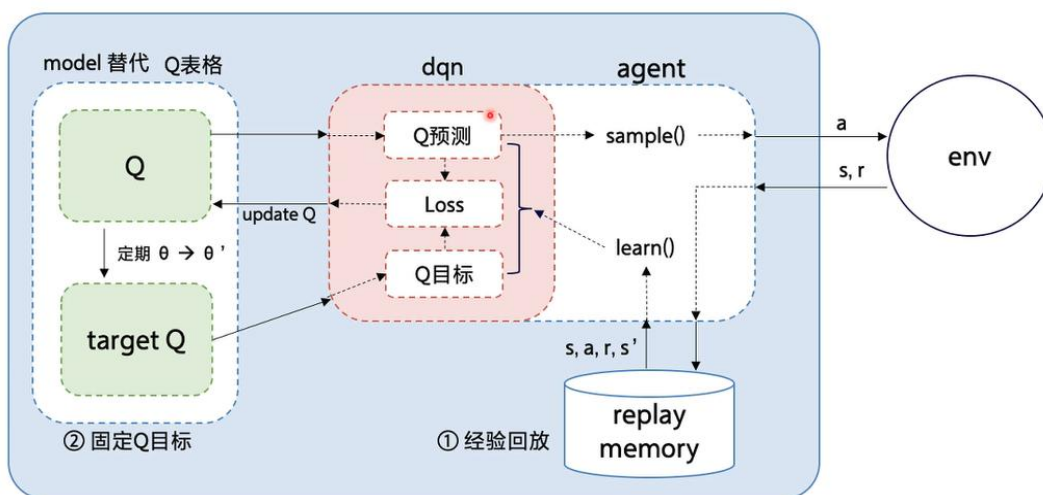
Diagram description: The equation is shown with arrows pointing to each term. Above the equation, '学习率 α ' points to the coefficient α , and '折旧因子 γ ' points to the coefficient γ . Below the equation, '得到的新Q值' points to 'new Q(s, a)', '当前Q值' points to 'Q(s, a)', '在s状态下做a动作的奖励' points to 'R(s, a)', '最大预期奖励' points to ' $\max_{a'} Q'(s', a')$ ', and '当前Q值' points to 'Q(s, a)'.

数据来源：西南证券整理

如果状态数以及对应的行为数过多,那么无法将其存储为表格进行更新。因此为了解决这个问题,将深度学习与强化学习结合,即可得到较为经典的深度 Q 网络 (Deep Q-Network, DQN) 模型,该模型通过引入深度神经网络来近似 Q 值函数,并解决了传统 Q 学习的一些问题。其主要使用神经网络模型近似 Q 值函数,具体特点如下:

- 1) 使用神经网络近似 Q 值函数:通过一个深度神经网络 $Q(s, a; \theta)$ 来近似 Q 值函数,其中 θ 是网络参数;选择一个动作 a_t ,把 a_t 输入到环境中,获得新状态 s_{t+1} 和 r ;
- 2) 经验回放 (Experience Replay):经验回放缓冲区 (Replay Buffer) 存储智能体与环境交互过程中获得的经验样本 (s, a, r, s') 。在更新 Q 值函数时,从经验回放缓冲区中随机抽取小批量样本,打破样本间的相关性,提高数据利用效率和学习稳定性。
- 3) 目标网络 (Target Network):DQN 使用两个神经网络:一个是主网络 $Q_{current}$,另一个是目标网络 Q_{target} ,目标网络 Q_{target} 的参数 θ' 每隔一定步数才从主网络 θ 同步一次,且更新 Q 值函数时,使用目标网络 Q_{target} 计算目标 Q 值,减少参数震荡和发散风险。

图 36: DQN 模型示意



数据来源:西南证券整理

在多因子轮动的场景下,我们可以将因子当前或历史的表现作为状态,因子对应的权重作为行动,合成因子下一期的表现作为奖励,交互环境可以定义为股票市场表现,策略即为如何基于当前或历史状态选择最优的合成因子权重。

4.4 基于 DQN 的泛高股息因子动态复权

本文采取强化学习中的 DQN 算法构建相应的因子轮动策略。其中训练数据为 2013 年 4 月 30 日至 2020 年 12 月 31 日的月度因子数据 (6 大类因子),并利用训练好的模型对 2021 年 1 月 31 日至 2024 年 6 月 30 日因子权重进行计算。模型具体细节如下:

Q 值近似网络：全连接神经网络 MLP (6, 24, 24, 6)，激活函数 tanh + relu；

奖励(reward, r)：当前权重下合成的 T 期因子与 T+1 期收益率秩相关系数；

训练方式：定义 episode 为 12 步，步长 1 期 (1 个月)；

超参数：折扣系数 $\gamma = 0.90$ 、学习率 0.001、总回合数 300, $\beta = 0.01$ ；

损失函数：

$$loss = MSE(Q_{current}(state, action), target_f) + \beta \frac{1}{6} \sum_{n=1}^6 (w_i - 1/6)^2$$

state 表示当前状态，即 T-1 期因子与 T 期收益的秩相关系数，而 *next_state* 表示 T 期因子与 T+1 期收益的秩相关系数。*action* 表示根据当前状态做出的动作，即根据当前状态生成新的因子权重， $Q_{current}(state, action)$ 可以视作当前 Q 值，*target_f* 表示目标 Q 网络计算得到的目标 Q 值。为了防止不同时间点下模型的输出权重差异过大，本文在损失函数部分引入当前权重与等权的 L2 罚项 $\beta \frac{1}{6} \sum_{n=1}^6 (w_i - 1/6)^2$ 。

具体训练步骤如下：

Step1：初始化 Q 网络和目标 Q 网络参数；

Step2：交互环节，获取初始状态 *state* 并生成因子权重 *action*，根据该权重计算奖励 *r* 以及下一个状态 *next_state*，并将 (*state, action, r, next_state*) 放入经验回放缓冲区；

Step3：从经验回放缓冲区随机抽取样本，样本数量 *batch_size* = 32，对每个样本计算目标 Q 值 *target_f*，并利用当前网络计算当前 $Q_{current}(state, action)$ ，计算损失函数 *loss*，通过反向传播的方式更新当前 Q 网络参数，计算累计奖励 (*total_reward*)。

Step4：重重上述步骤，在最近的 100 回合训练中仅保留累计奖励最高的模型。

本文将基于 DQN 计算得到的因子权重合成上述 6 类因子得到的新因子命名为 DQN 因子，该因子表现如下：

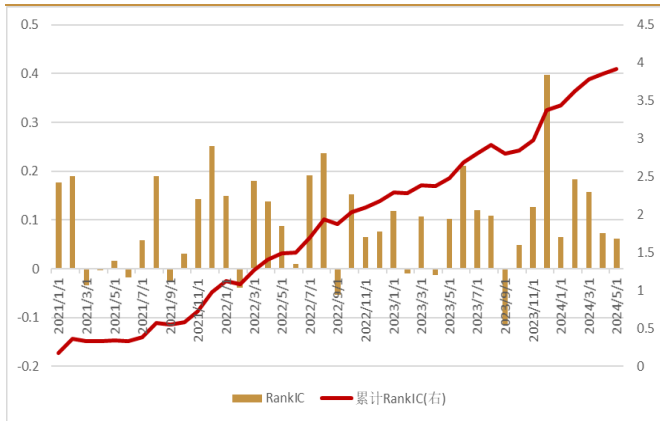
表 20：等权泛高股息因子测试结果

	IC	ICIR	相对等权 IC 胜率	年化收益率	年化波动	信息比率	最大回撤率	相对中证红利超额
DQN 因子	9.55%	3.30	56.10%	17.58%	19.57%	0.90	27.30%	5.93%
等权因子	9.29%	3.17	—	16.42%	19.41%	0.85	27.24%	4.77%

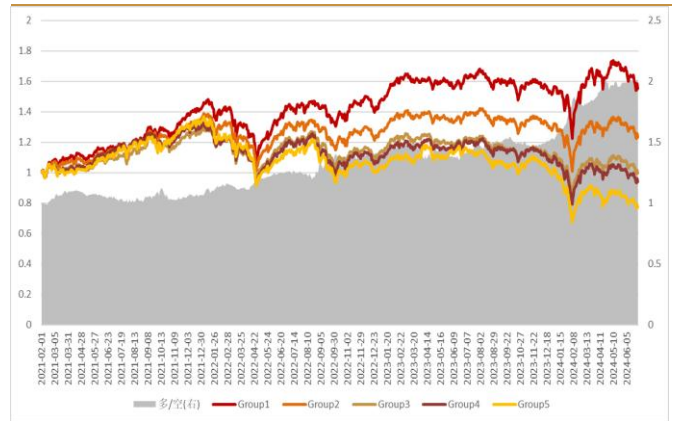
数据来源：wind、西南证券整理，数据截至 2024/06/30

自 2021 年 1 月 31 日至 2024 年 6 月 30 日，DQN 因子 IC 均值为 9.55%，ICIR 为 3.30，IC 表现优于等权合成的泛高股息因子；此外，DQN 因子相对等权因子 IC 胜率为 56.10%。

多头表现方面，DQN 因子多头年化收益率 17.58%，相对中证红利全收益指数年化超额收益率 5.93%，多头组合信息比率 0.90，表现优于等权。

图 37: DQN 因子 IC


数据来源: wind、西南证券整理, 数据截至 2024/06/30

图 38: DQN 因子分组测试


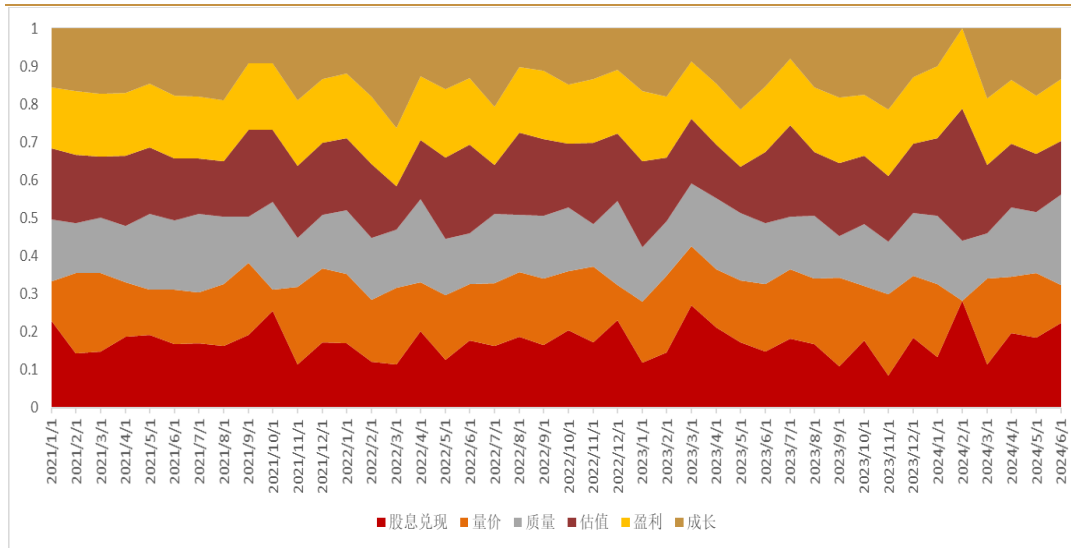
数据来源: wind、西南证券整理, 数据截至 2024/06/30

表 21: DQN 因子多空测试

	年化收益率	年化波动	信息比率	最大回撤率
DQN 因子	20.97%	11.93%	1.76	10.08%
等权因子	19.85%	11.64%	1.71	10.13%

数据来源: wind、西南证券整理, 数据截至 2024/06/30

多空表现方面, DQN 因子多空年化收益率 20.97%, 最大回撤率 10.08%, 表现优于等权因子 (19.85%、10.13%)。

图 39: DQN 模型历史各大类因子权重


数据来源: wind、恒生聚源、西南证券整理, 数据截至 2024/06/30

5 总结

本文以机器学习为工具，分别研究了机器学习在不同风格红利指数选择以及泛高股息股票池内传统多因子动态复权上的应用。

首先，本文选取包括中证红利指数、红利低波指数等在内共计 12 个不同风格的红利指数，利用传统的基于决策树的 LightGBM 模型进行指数选择（红利指数因子）。经过测试，自 2021 年 1 月至 2024 年 5 月红利指数因子 RankIC 为 17.11%，多头组合（Top3）年化收益率 16.96%，相对等权组合（8.79%）年化超额收益率 8.17%，剔除无跟踪 ETF 产品指数后，红利主题指数因子 RankIC 为 18.92%，多头组合（Top3）年化收益 16.21%，相对等权组合年化超额 7.48%，最大回撤率 14.32%，因子表现较为优秀。

其次，针对红利高股息选股问题，本文首先定义**泛高股息股票池**，并分别测试了在泛高股息股票池中表现较好的因子，最终筛选出**量价、质量、估值、成长、盈利**五大类因子。除了上述五大类因子外，本文还构造了**股息兑现因子**，用以刻画高股息股票的股价兑现程度。最后本文利用强化学习 DQN 模型对上述六大类因子进行动态复权，得到 DQN 因子。经过测试，自 2021 年 1 月 31 日至 2024 年 6 月 30 日，DQN 因子 IC 均值为 9.55%，多头年化收益率 17.58%，表现均优于等权合成因子。

区别于传统的多因子模型，强化学习模型综合考虑因子表现状态，挖掘因子时序表现异同，从而给出相对等权更优的动态权重配置，而上文中本文是根据控制与等权偏离的背景下给出的动态权重，事实上也可以根据不同的权重设置偏好给出不同的限制条件。此外，若引入更好的传统选股因子以增加因子维度，也可能使强化学习动态复权因子表现更上一层楼。

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接获取任何形式的补偿。

投资评级说明

报告中投资建议所涉及的评级分为公司评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 个月内的相对市场表现，即：以报告发布日后 6 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准。

公司评级	买入：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅在 20% 以上
	持有：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 10% 与 20% 之间
	中性：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 -10% 与 10% 之间
	回避：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅介于 -20% 与 -10% 之间
	卖出：未来 6 个月内，个股相对同期相关证券市场代表性指数涨幅在 -20% 以下
行业评级	强于大市：未来 6 个月内，行业整体回报高于同期相关证券市场代表性指数 5% 以上
	跟随大市：未来 6 个月内，行业整体回报介于同期相关证券市场代表性指数 -5% 与 5% 之间
	弱于大市：未来 6 个月内，行业整体回报低于同期相关证券市场代表性指数 -5% 以下

重要声明

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与作者在自身所知知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

《证券期货投资者适当性管理办法》于 2017 年 7 月 1 日起正式实施，本报告仅供本公司签约客户使用，若您并非本公司签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告，本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，本公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附录的，本公司将保留向其追究法律责任的权利。

请务必阅读正文后的重要声明部分

西南证券研究发展中心

上海

地址：上海市浦东新区陆家嘴 21 世纪大厦 10 楼

邮编：200120

北京

地址：北京市西城区金融大街 35 号国际企业大厦 A 座 8 楼

邮编：100033

深圳

地址：深圳市福田区益田路 6001 号太平金融大厦 22 楼

邮编：518038

重庆

地址：重庆市江北区金沙门路 32 号西南证券总部大楼 21 楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	座机	手机	邮箱
上海	蒋诗烽	总经理助理、销售总监	021-68415309	18621310081	jsf@swsc.com.cn
	崔露文	销售副总监	15642960315	15642960315	clw@swsc.com.cn
	谭世泽	高级销售经理	13122900886	13122900886	tsz@swsc.com.cn
	李煜	高级销售经理	18801732511	18801732511	yfliyu@swsc.com.cn
	卞黎旸	高级销售经理	13262983309	13262983309	bly@swsc.com.cn
	田婧雯	高级销售经理	18817337408	18817337408	tjw@swsc.com.cn
	张玉梅	销售经理	18957157330	18957157330	zymf@swsc.com.cn
	魏晓阳	销售经理	15026480118	15026480118	wxyang@swsc.com.cn
	欧若诗	销售经理	18223769969	18223769969	ors@swsc.com.cn
	李嘉隆	销售经理	15800507223	15800507223	ljliong@swsc.com.cn
	龚怡芸	销售经理	13524211935	13524211935	gonggy@swsc.com.cn
	孙启迪	销售经理	19946297109	19946297109	sqdi@swsc.com.cn
	蒋宇洁	销售经理	15905851569	15905851569	jjj@swsc.com.c
北京	李杨	销售总监	18601139362	18601139362	yfly@swsc.com.cn
	张岚	销售副总监	18601241803	18601241803	zhanglan@swsc.com.cn
	杨薇	资深销售经理	15652285702	15652285702	yangwei@swsc.com.cn
	姚航	高级销售经理	15652026677	15652026677	yhang@swsc.com.cn
	张鑫	高级销售经理	15981953220	15981953220	zhxin@swsc.com.cn

	王一菲	销售经理	18040060359	18040060359	wyf@swsc.com.cn
	王宇飞	销售经理	18500981866	18500981866	wangyuf@swsc.com
	路漫天	销售经理	18610741553	18610741553	lmtf@swsc.com.cn
	马冰竹	销售经理	13126590325	13126590325	mbz@swsc.com.cn
	郑龔	广深销售负责人	18825189744	18825189744	zhengyan@swsc.com.cn
	杨新意	广深销售联席负责人	17628609919	17628609919	yxy@swsc.com.cn
	张文锋	高级销售经理	13642639789	13642639789	zwf@swsc.com.cn
广深	龚之涵	销售经理	15808001926	15808001926	gongzh@swsc.com.cn
	丁凡	销售经理	15559989681	15559989681	dingfyf@swsc.com.cn
	陈紫琳	销售经理	13266723634	13266723634	chzlyf@swsc.com.cn
	陈韵然	销售经理	18208801355	18208801355	cyryf@swsc.com.cn
	林哲睿	销售经理	15602268757	15602268757	lzt@swsc.com.cn
