

汇报目录

- ① 研究动机与因子轮动的理论背景
- ② 轮动基础：数据与市场状态实证
- ③ 因子轮动策略框架与权重生成机制
- ④ 模型训练数据集的构建与样本划分
- ⑤ 构建因子轮动预测模型
- ⑥ 动态因子策略回测结果
- ⑦ 结论与研究展望

研究动机：因子配置是否应当保持不变？

- 传统多因子投资框架通常假设：
 - 各类风格因子在长期内能够持续提供稳定的风险溢价；
 - 因而在实务中多采用静态或准静态的因子权重配置方式。
- 然而，近年来的理论与实证研究表明：
 - 因子收益并非时间不变；
 - 其表现呈现出明显的阶段性、周期性与结构性变化。

因子收益非平稳性的经济直觉

- 不同因子对应不同的风险补偿机制，其有效性依赖于市场环境：
 - **动量因子**依赖趋势延续，在方向性明确的市场中更为有效；
 - **价值因子**反映均值回归，在风险溢价上升或情绪悲观阶段具备防御属性；
 - **质量因子**强调盈利能力与稳健性，在震荡或不确定性较高的环境中表现相对占优；
 - **低波动与规模因子**往往与风险偏好与资金结构变化密切相关。
- 因此，单一或静态的因子配置方式：
 - 可能在特定阶段表现良好；
 - 却在其他阶段出现系统性失效。

因子轮动的核心思想

- 因子轮动理论的基本观点：
 - 因子溢价具有显著的时间变化特征；
 - 因子表现与市场所处的宏观与风险状态密切相关。
- 若能够在一定程度上刻画当前市场环境，并预测下一阶段不同因子的相对强弱：
 - 则有可能通过动态调整因子权重；
 - 在风险可控的前提下提升组合的风险调整后收益。

因子轮动在实践中的关键挑战

- 因子轮动策略在实际应用中面临两项核心困难：
- **挑战一：市场状态不可直接观测**
 - 市场环境的划分往往具有较强主观性；
 - 缺乏客观、可重复的数据驱动识别方法。
- **挑战二：因子收益的动态演化复杂**
 - 因子收益具有显著的时序相关性与噪声特征；
 - 简单规则或静态判断难以刻画其变化过程。

研究思路：从因子轮动到数据驱动建模

- 本研究将因子轮动问题拆解为两个相互关联的子问题：
- **第一步：市场状态识别**
 - 基于指数层面的市场特征，识别不同市场状态；
 - 从实证角度检验因子收益的状态依赖性是否存在。
- **第二步：因子收益预测与配置**
 - 利用机器学习模型刻画因子收益的时间序列结构；
 - 预测下一期各因子的相对表现，并生成动态因子权重。

研究对象与研究目标

- 研究对象选择：
 - 以上证50指数成分股作为股票池；
 - 兼顾样本稳定性与市场代表性。
- 研究目标：
 - 重点关注动态因子配置在样本外阶段的表现；
 - 检验其是否能够相较于静态多因子策略，
 - 实现更优的风险调整后收益。
- 最终目标：
 - 构建一套可复现、可扩展的数据驱动因子轮动研究范式。

研究对象与数据结构：两类数据、统一月度对齐

- **股票池**：上证50指数成分股（大市值、流动性强、噪声与交易约束相对更小）。
- **时间频率**：月度（用于状态识别、因子收益序列建模与调仓回测）。
- **两类核心数据**（按月严格对齐）：
 - ① 因子层面：**因子多空收益**时间序列（刻画因子溢价随时间变化）。
 - ② 指数层面：**市场特征**（用于识别市场状态 Regime）。
- **关键一致性处理**：所有日期统一映射为对应月份的**第一天**作为时间索引，确保调仓点信息不穿越。

风格因子选择：经典且可解释

- 研究目标是验证“轮动机制”而非穷举因子，因此选择经典且经济含义清晰的风格因子：
 - 动量 (Momentum): MOM20, MOM120, RSI¹
 - 价值 (Value): PB^{-1} , PE^{-1} , DIV^2
 - 质量 (Quality): ROE, PROFIT_GR³
 - 风险相关 (Risk-related): VOL, BETA⁴
- 这样做的好处：
 - 降低高维因子带来的过拟合风险；
 - 保留清晰的经济直觉，便于解释“为什么轮动”。

¹ Jegadeesh, N. and Titman, S. (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*; Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H. (2012), "Time Series Momentum," *Journal of Financial Economics*; Lo, A.W., Mamaysky, H. and Wang, J. (2000), "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," *Journal of Finance*.

² Fama, E.F. and French, K.R. (1992), "The Cross-Section of Expected Stock Returns," *Journal of Finance*; Lakonishok, J., Shleifer, A. and Vishny, R.W. (1994), "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance*.

³ Novy-Marx, R. (2013), "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics*; Fama, E.F. and French, K.R. (2015), "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*.

⁴ Frazzini, A. and Pedersen, L.H. (2014), "Betting Against Beta," *Journal of Financial Economics*; Ang, A., Hodrick, R.J., Xing, Y. and Zhang, X. (2006), "The Cross-Section of Volatility and Expected Returns," *Journal of Finance*.

从横截面信号到时间序列：因子多空收益（Long-Short）

- 目的：把“因子暴露”转换为可建模的**因子收益序列**。
- 月度调仓日：对成分股按某因子值排序，构造五分位组合（G1–G5）。
- 定义因子多空收益：

$$r_{t,f}^{LS} = r_{t,f}^{G5} - r_{t,f}^{G1}$$

- 解释：通过多空构造在较大程度上剥离市场整体方向，更直接度量因子溢价的时间变化。

市场状态信息：用指数层面特征刻画“环境”

- Regime 识别不使用个股微观信息，而使用上证50指数特征，刻画整体环境：
 - 月度指数收益率（ret_month）
 - 20日（及扩展：60日）年化波动率（vol_20_annual_month_end等）
 - 月度成交量总和（vol_month_sum）及成交量变化率（可选扩展）
- 三个维度对应直觉：
 - 收益：方向与趋势
 - 波动：风险强度/不确定性
 - 成交：交易活跃度/情绪与资金参与
- 工程处理：所有特征进入模型前**标准化**，避免量纲影响聚类。

[illegible]

- 加拿大政府也于 1997 年 12 月 1 日, 正式宣布将 1997 年 11 月 1 日以前

三类市场状态的经济含义：收益-风险-成交的稳定区分

- **Regime 0：趋势上行/风险偏好较高**
 - 指数月度收益均值为正；
 - 20日年化波动率较低；
 - 成交量温和（环境更“顺风”）。
- **Regime 1：高波动下行/风险集中释放**
 - 收益显著偏低（存在明显负收益月份）；
 - 波动率显著高；
 - 成交量波动大（情绪剧烈、冲击更强）。
- **Regime 2：低方向性震荡/交易活跃**
 - 收益接近零或小幅波动；
 - 波动率中等；
 - 成交量维持高位（震荡但参与度高）。

实证发现：因子收益对 Regime 高度敏感（方向一致且可解释）

- **Regime 0（正收益、低波动）：动量类因子整体更强**
 - MOM20 / MOM120 / RSI 的多空收益均值更高；
 - 解释：趋势延续更显著，行为惯性与方向性更强。
- **Regime 1（低收益、高波动）：动量显著弱化，价值与防御因子相对占优**
 - 动量出现回撤或失效；
 - PB^{-1} 、 PE^{-1} 与低波动/低 Beta 更稳健；
 - 解释：风险偏好下降，均值回归与防御暴露更有利。
- **Regime 2（震荡、高成交）：质量类因子更稳健**
 - ROE、PROFIT_GR 风险调整后表现更友好；
 - 解释：不确定性与震荡环境下，基本面稳健性溢价更易体现。

本节结论：为何可以进入“预测 → 动态权重”主线？

- 结论1：市场状态可由指数特征客观识别，且三类状态在收益/风险/成交维度具有稳定区分。
- 结论2：因子收益存在显著状态依赖性，不同风格因子在不同 Regime 下强弱系统性变化。
- 因此：
 - 静态因子权重难以跨环境保持最优；
 - 将问题转化为可学习任务：用历史因子收益（及 Regime 辅助信息）预测下一期因子相对表现，
 - 再据此生成可执行的动态因子权重。

核心问题：预测因子收益之后，如何形成可执行策略？

- 前两部分已表明：
 - 因子收益具有显著的时间变化与状态依赖性；
 - 因子轮动在经济直觉与实证层面均具有合理性。
- 进一步的问题是：
 - 若能够预测下一期各因子的相对表现，
 - 如何将预测结果转化为**连续、可控、可回测**的因子配置权重？
- 本研究关注的是**因子层面的配置问题**，
 - 而非个股择时或选股问题。

因子轮动策略的整体流程

- ① 基于历史数据，对下一期各因子多空收益进行预测；
 - ② 将预测得到的因子收益向量映射为因子配置权重；
 - ③ 在因子层面构建组合，并映射到股票组合进行回测。
- 该流程的核心在于：
 - 因子权重随时间连续变化；
 - 不依赖人工规则或离散切换。

预测目标的定义：预测“因子相对强弱”，而非绝对收益

- 本研究中，模型的预测对象为：
 - 下一期十个因子的多空收益向量
- 关注重点并非预测值本身是否精确，
 - 而是因子之间的相对排序与强弱结构。
- 因此：
 - 即使存在一定预测误差，
 - 只要相对排序具有信息含量，
 - 就有可能改善因子配置效果。

最终建模数据集：统一的时间序列输入

- 在完成因子多空收益与市场状态（Regime）标签合并后，
- 构建用于机器学习建模的**最终数据集**：
 - `final_factor_longshort.csv`
 - 月度频率，是后续模型训练、验证与预测的**唯一中心数据源**
- 数据集中每一行对应一个月份，包含：
 - 十个风格因子的多空收益（G5–G1）
 - 当期市场状态 Regime 标签

数据来源与时间范围

- 最终数据集由以下两类数据按月严格对齐后构建：
- 因子多空收益数据（factor_longshort.csv）
 - 基于上证50成分股
 - 月末横截面排序、五分位分组
 - 计算高分组减低分组（G5—G1）得到因子收益
- 市场状态数据（cluster_info）
 - 基于指数特征的 KMeans 聚类结果
 - 为每个月分配唯一 Regime 标签
- 合并后时间范围：
 - **2015-01-01 至 2024-12-01**
 - 连续、无缺失的月度观测

监督学习样本的构造：滑动窗口设定

- 为构造适用于时间序列模型的监督学习样本，
- 采用长度为 12 个月的滑动窗口（rolling window）。
- 在任意预测时点 t ：
 - **输入** X_t ： $t - 12$ 至 $t - 1$ 的历史信息
 - **输出** y_t ： 第 t 个月的因子多空收益
- 该设定确保模型在任意时点的预测，
 - 仅基于当时可获得的历史数据，
 - 符合真实投资场景下的信息约束。

模型输入特征的组成

- 模型输入特征由两部分在时间维度上拼接而成：
- 因子收益特征
 - 十个因子的历史多空收益序列
 - 刻画因子溢价自身的时间依赖结构
- 市场状态特征
 - Regime 标签经 one-hot 编码后引入
 - 显式刻画不同市场环境下的因子表现差异
- 该联合输入结构使模型能够同时利用：
 - 因子内部动态信息
 - 以及外部环境信息

样本划分原则：严格的时间顺序切分

- 金融时间序列建模中，随机划分样本会导致严重的数据泄漏风险；
- 本研究采用**严格按时间顺序的前推切分**。
- 全部监督学习样本按比例划分为：
 - 训练集（70%）**：用于模型参数学习
 - 验证集（20%）**：用于结构选择与训练过程监控
 - 测试集（10%）**：仅用于样本外预测与回测评估

测试集时间区间与样本外评估

- 测试集对应样本期末的时间区间：
 - **2024-02-01 至 2024-12-01**
- 在该区间内：
 - 模型参数已完全由训练集与验证集确定；
 - 测试集数据不参与任何形式的训练或调参；
 - 仅用于生成因子收益预测与后续动态权重回测。
- 因此，该区间构成本文结论的**核心样本外检验窗口**。

- 本研究在数据构建与模型训练过程中不存在数据穿越或前视偏差：
- 特征层面
 - 因子收益与 Regime 标签均基于当期及之前信息计算；
 - 不使用任何未来数据。
- 样本划分层面
 - 训练、验证、测试集严格按时间顺序切分；
 - 测试集不参与模型训练与参数选择。

预测—配置—回测的时间一致性

- 在测试集的每一个月份 t ：
 - 模型仅使用 $t - 1$ 及之前 12 个月的信息进行预测；
 - 基于预测的因子收益生成当期因子权重；
 - 再用于 t 期的策略配置与回测。
- 整个流程严格遵循：
 - 先预测、再配置、后实现收益的时间顺序；
- 因此，最终结果具有真实投资场景下的可解释性与可信度。

建模目标：预测下一期因子相对表现

- 在完成数据集构建与样本划分后，进入机器学习建模阶段。
- 本研究的预测目标并非：
 - 个股收益，或市场方向；
- 而是：
 - 下一期十个风格因子的多空收益向量。
- 该设定的核心关注点在于：
 - 因子之间的**相对强弱结构**，
 - 而非单一因子收益预测的绝对精度。

为什么采用联合预测（Multi-output）而非独立模型？

- 十个因子收益并非相互独立：
 - 宏观风险偏好变化可能同时影响多类因子；
 - 市场状态切换往往带来系统性的因子结构变化。
- 若为每个因子单独训练模型：
 - 忽略因子间的相关结构；
 - 在样本量有限的月频数据下更易过拟合。
- 因此，本研究采用：
 - **单一模型同时预测十个因子收益**的多任务回归设定，
 - 以提升统计效率与泛化能力。

模型选择：LSTM 时间序列预测框架

- 因子收益序列具有以下特征：
 - 显著的时间依赖性；
 - 非平稳性与阶段性变化；
 - 对近期历史信息更为敏感。
- 因此选择 LSTM（Long Short-Term Memory）模型：
 - 显式建模时间序列依赖结构；
 - 在有限样本下相对稳健；
 - 能够自然接收多维时序输入。

模型输入输出的张量结构

- 输入样本 X_t :
 - 过去 12 个月的历史序列；
 - 每个时间步包含：
 - 10 个因子多空收益；
 - 市场 Regime 的 one-hot 编码特征。
- 张量维度表示为：

$$X \in \mathbb{R}^{(\text{样本数}, 12, \text{特征数})}$$

- 输出 y_t :
 - 第 t 个月的十维因子多空收益向量；
 - 属于典型的多输出回归问题。

LSTM 网络结构概览

- 网络采用 Keras Sequential 框架构建，结构如下：
- **LSTM 层（64 units）**
 - 输入形状：(12, 特征数)
 - 将历史序列压缩为一个时序表示向量
- **Dropout（0.2）**
 - 用于降低过拟合风险
- **Dense 层（32 units, ReLU）**
 - 提升非线性表达能力
- **输出层（10 units, Linear）**
 - 同时输出十个因子的预测收益

损失函数、优化器与训练策略

- 损失函数：
 - 均方误差 (MSE)，适用于连续型回归问题
- 优化器：
 - Adam (学习率 0.001)，在噪声梯度环境下较为稳健
- 训练策略：
 - 训练轮数：80 epochs
 - 批大小：16
 - 同步监控验证集损失以评估泛化能力

样本外预测与时间索引对齐

- 模型训练完成后，在测试集上执行预测：
 - 输入 X_{test}
 - 输出因子收益预测矩阵 y_{pred}
- 由于采用 12 个月滑动窗口：
 - 预测结果需进行严格的时间索引对齐；
 - 确保每一行预测对应真实的预测月份。
- 最终导出：
 - `factor_prediction.csv`
 - 结构为：month + 各因子预测列 (`_pred`)

本节小结：模型在整体框架中的角色

- 本节构建了一个：
 - 基于历史因子收益与市场状态信息的
 - 多任务时间序列预测模型。
- 模型输出并不直接用于交易，
 - 而是作为后续因子权重生成（softmax 映射）的输入。
- 下一部分将基于预测结果：
 - 构建动态因子权重，
 - 并通过回测系统性评估因子轮动策略的有效性。

回测设定与统一口径

- 回测目标：检验“预测因子收益 \rightarrow softmax 映射 \rightarrow 动态因子权重”是否能在风险可控前提下带来实质性的收益-风险改进。
- 统一口径：
 - 测试区间（样本外）：2024-02-01 – 2024-12-01
 - 基准：上证50指数
 - 调仓频率：月度
 - 组合口径：高因子组合

对照策略定义（四策略）

- 等权多因子（Equal-weight）：10 个因子固定等权（各 10%），不使用 Regime/预测。
- 规则配置（Rule-based）：规则驱动的静态配置（不依赖机器学习预测）。
- 动态权重（有 Regime）：使用预测 + Regime 特征，softmax 生成当期权重，月度更新。
- 动态权重（无 Regime）：仅使用预测（不含 Regime 特征），softmax 生成当期权重，月度更新。

策略核心指标总览

策略	CAGR	Vol	Sharpe	Sortino	Calmar	MaxDD	DD(月)	WinRate
等权配置	21.24%	20.56%	1.0294	4.6733	5.4106	-3.93%	3	0.50
规则配置	23.13%	32.75%	0.7754	2.3725	1.6603	-13.93%	6	0.50
动态权重(有状态)	36.71%	29.77%	1.1856	3.4556	3.8853	-9.45%	3	0.60
动态权重(无状态)	40.09%	26.87%	1.3806	6.3496	13.1028	-3.06%	2	0.70

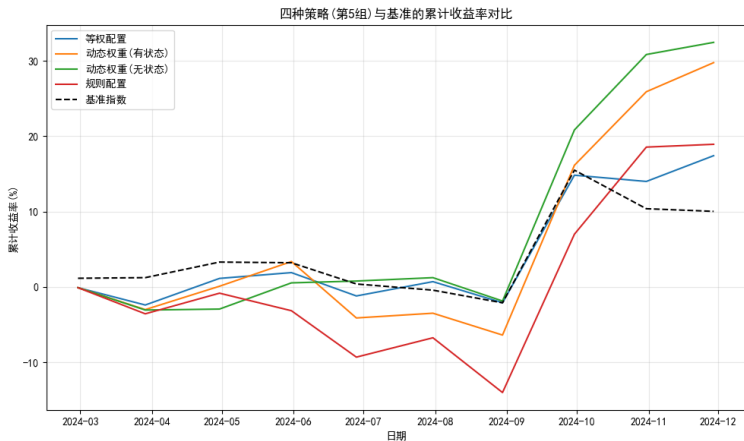
- 动态权重（无状态）在 CAGR、Sharpe、Calmar 与最大回撤控制上同时占优。

累计收益：期末值对比

策略	期末累计收益	基准期末累计收益
等权配置	17.42%	10.03%
动态权重(有状态)	29.77%	10.03%
动态权重(无状态)	32.46%	10.03%
规则配置	18.93%	10.03%

累计收益曲线

- 建议插入：四策略与基准的累计收益率曲线图。



相对基准表现 (CAPM / IR)

策略	年化超额收益	Tracking Error	IR	CAPM Alpha(年化)
等权配置	0.0909	0.0733	1.0552	0.0910
动态权重(有状态)	0.2455	0.1743	1.2547	0.2014
动态权重(无状态)	0.2794	0.1600	1.4792	0.2363
规则配置	0.1098	0.2154	0.5554	0.0986

- 动态权重（无状态）在年化超额收益与信息比率（单位跟踪误差效率）上同时领先。

信息比率对比（建议放图）

- 建议插入：四策略 Information Ratio 柱状图。

回撤曲线对比（建议放图）

- 建议插入：四策略（第5组）回撤曲线图，用于展示尾部风险差异与修复速度。

Sharpe 对比（建议放图）

- 建议插入：四策略 Sharpe Ratio 柱状图（与上页 IR 柱状图构成“风险调整后收益 + 相对基准”双证据）。

换手率与单位换手收益（交易成本代理）

策略	平均换手率(%)	平均月度收益(%)	单位换手收益
等权配置	51.1111	1.7640	0.0345
动态权重(有状态)	24.4444	2.9410	0.1203
动态权重(无状态)	42.2222	3.0910	0.0732
规则配置	32.2222	2.1160	0.0657

- 含 Regime 版本在更低换手下单位效率更高；无 Regime 版本收益更高但换手相对更大。

因子层面诊断：多空收益（Top-Bottom）与 IC

策略	多空CAGR	多空Vol	多空Sharpe	多空MaxDD	多空WinRate
等权配置	0.0074	0.0964	0.1195	-0.0445	0.40
动态权重(有状态)	-0.2450	0.1678	-1.5749	-0.2403	0.30
动态权重(无状态)	-0.2655	0.2161	-1.3010	-0.2506	0.50
规则配置	-0.0715	0.2393	-0.1966	-0.2295	0.40

策略	IC Mean	IC Std	IC_IR (年化)
等权配置	-0.0300	0.1416	-0.7339
动态权重(有状态)	-0.0878	0.1892	-1.6072
动态权重(无状态)	-0.0833	0.2575	-1.1209
规则配置	0.0167	0.2819	0.2048

因子区分度检验 (T-Stat / P-Value)

策略	T_Stat	P_Value
等权配置	0.11	54.18
动态权重(有状态)	-1.45	9.29
动态权重(无状态)	-1.19	13.38
规则配置	-0.18	43.13

- 该组“因子层面”统计并未显示传统意义上的显著增强；
- 动态权重的优势更可能来自组合层面权重调整对收益路径/风险暴露的工程性改良。

回测结论

- 动态权重策略整体优于静态策略：CAGR 与 Sharpe/IR 同时提升。
- 最优策略为动态权重（无状态）：
 - CAGR 40.09%，Sharpe 1.3806，IR 1.4792，MaxDD -3.06%。
- Regime 的边际贡献在本样本外区间内有限或不稳定：
 - 含 Regime 版本 CAGR 36.71%，Sharpe 1.1856，IR 1.2547，且回撤更深。

研究问题回顾

- 本研究围绕一个核心问题展开：
 - 在中国股票市场中，
 - 因子收益具有显著时间变化特征的背景下，
 - 是否可以通过数据驱动的方法，
 - 构建有效的因子轮动策略？
- 为回答该问题，本文构建了一套完整流程：
 - 因子构建与多空收益提取；
 - 市场状态（Regime）的无监督识别；
 - 基于时间序列的因子收益预测；
 - 动态因子权重映射与样本外回测评估。

核心结论一：动态因子权重在组合层面显著有效

- 样本外回测结果表明：
 - 基于因子收益预测并通过 softmax 映射得到的动态因子权重，
 - 在收益水平与风险调整后表现上，
 - 均显著优于等权配置与规则型静态策略。
- 更重要的是：
 - 收益提升并非以放大波动或回撤为代价；
 - 在 Sharpe、IR、Calmar 等多维指标下均表现更优。

核心结论二：优势来自权重机制，而非因子显著性增强

- 在因子层面统计检验中：
 - 多空组合收益整体偏弱；
 - IC 均值与显著性检验未显示明显改善。
- 这表明：
 - 动态因子策略的优势，
 - 并非来源于单个因子预测能力的统计显著增强，
 - 而更可能来自组合层面权重调整机制，
 - 对收益路径与风险暴露的工程性改良。

核心结论三：Regime 方法的边际贡献具有样本依赖性

- 在本研究样本区间内：
 - 引入 Regime 信息并未带来稳定的边际收益提升；
 - 含 Regime 的动态权重策略在部分指标上反而略逊。
- 需要强调的是：
 - 该结论并不否定 Regime 方法的理论合理性；
 - 而是反映其在样本期较短、状态转移有限条件下，
 - 难以稳定发挥增量信息的现实约束。

研究局限性

- 样本期长度有限
 - 样本外测试区间较短；
 - 年化指标对个别月份较为敏感。
- Regime 识别的简化假设
 - 基于有限的指数特征进行聚类；
 - 未显式刻画宏观或政策变量。
- 交易成本与市场冲击的近似处理
 - 回测主要通过换手率进行间接评估；
 - 未显式建模冲击成本与流动性约束。

未来研究方向

- 样本扩展与跨市场检验
 - 延长时间跨度；
 - 在其他指数或市场中检验稳健性。
- 更精细的市场状态刻画
 - 引入宏观、流动性或情绪变量；
 - 探索连续状态或状态转移模型。
- 权重映射与风险约束的改进
 - 在 softmax 映射中引入风险或换手惩罚；
 - 探索收益-风险联合优化的权重生成机制。

总结

- 本研究实证表明：
 - 即便在因子统计显著性并不突出的环境中，
 - 数据驱动的动态权重调整，
 - 仍可在组合层面显著改善收益-风险结构。
- 该发现为：
 - 因子轮动策略的工程实现，
 - 以及机器学习在多因子投资中的应用，
 - 提供了一种可复现、可扩展的研究范式。