

A 股横截面收益预测中的复杂度与收益关系

市场信息进入 FT-Transformer 的样本外比较与消融证据

AI 金融研学实训营专题项目组

2026 年 7 月 23 日

摘要

本文研究模型复杂度是否改善 A 股未来收益的样本外横截面排序，并进一步检验市场状态信息应以何种方式进入 FT-Transformer。我们使用历史时点沪深北全 A 股票池、50 个公司特征和 7 个市场状态变量，预测信号日之后从第 1 个收盘到第 6 个收盘的完整五日复权收益。训练期截至 2021 年末，验证期为 2022–2023 年，测试期为 2024–2025 年。统一比较 Elastic Net、LightGBM、简化 FT-Transformer、市场变量直接拼接以及 CLS 级条件仿射调制五类模型。测试集共包含 2,238,748 个股票–日期观测、485 个交易日和 4,863 只股票。最重要的结果是复杂度与收益不单调：仅含 50 个线性系数的 Elastic Net 取得最高标准化标签 R^2 和 Top 组成本后均值，LightGBM 的多空成本后均值最高，而三个 FT 变体没有在这些经济指标上提供一致增量。B3 的 Rank IC 最高 (0.113746)，B2 的年度末年化 Rank ICIR 最高 (0.907199，对应 $\sqrt{252}$ 换算值约 14.40)，但这些小幅差异来自单一随机种子，不能解释为稳健结构优势。条件调制 B4 未优于 B2 或 B3，并在高波动期弱于两者。现有证据因此不支持“复杂模型必然更优”或“当前自适应机制有效”；单一随机种子、重叠五日收益、简化交易成本与缺少显著性检验限制了结论强度。

关键词： A 股；横截面收益预测；FT-Transformer；市场状态；样本外检验；Rank IC

1 引言

公司特征与未来收益之间的关系具有非线性、交互性和随市场环境变化的可能性。传统线性模型容易解释，但可能遗漏复杂关系；树模型和深度模型能够学习更灵活的函数，却也更容易在高维、低信噪比的金融数据中产生样本外失效。资产定价机器学习文献因此强调统一的时间顺序检验、强基准与经济指标，而不是只比较训练误差或单次最优结果 [3, 5]。

FT-Transformer 将每个表格特征映射为独立 token，再沿特征轴使用自注意力学习交互。原论文在多个表格任务中取得较强表现，同时明确指出 GBDT 与深度模型之间不存在普遍优胜者 [4]。这一结论对金融研究尤其重要：模型应通过共同股票池、共同标签、共同切分和共同评估进行比较，复杂度本身不能成为有效性的证据。

本文围绕三个可证伪假设展开：

1. H1: LightGBM 与 FT-Transformer 能全面优于 Elastic Net；
2. H2: 市场状态自适应 FT-Transformer 优于原始 FT-Transformer；
3. H3: 自适应机制的增量主要出现在状态切换或高波动阶段。

为区分“增加模型复杂度”“加入市场信息”和“改变条件化结构”三类效应，我们设置从 B0 到 B4 的递进对照。本文首先发现 Elastic Net 在 R^2 和 Top 组成后均值上没有被更复杂模型超越，说明 50 因子场景中的复杂度-收益关系并不单调；其次，B3 的横截面排序均值虽最高，但 B2 的排序稳定性更高，且差异不足以越过单 seed 的证据边界；最后，B4 的结构创新未形成稳健增量。这些结果共同表明，模型容量和结构新颖性本身不能替代样本外证据。

2 相关研究

2.1 机器学习与横截面收益预测

已有研究系统比较线性、树模型和神经网络，展示公司特征中的非线性与交互可能改善样本外收益预测 [5]。非参数研究进一步讨论特征的独立作用、非线性与冗余 [1]；相关综述则区分机器学习在因子建模、收益预测和资产定价中的不同任务 [3]。本文沿用“强基准、时间切分、样本外排序与组合指标并重”的研究框架，但不把任何海外结论直接外推为 A 股有效性证据。

在中国市场语境下，涨跌幅限制会影响价格发现、波动与可成交性，不能把无约束的分组收益直接解释为可实现策略收益 [2]。工程上，Qlib 提供了面向量化研究的数据、模型与回测工作流范式 [9]；本文仅借鉴其 Alpha158 特征命名与研究治理思想，数据抽取、时点对齐和正式评估均由本项目独立实现。股票交易、停牌与价格限制的具体处理以交易所规则为制度背景 [10]。

Elastic Net 通过 ℓ_1 与 ℓ_2 联合惩罚兼顾稀疏性和共线性处理 [11]；LightGBM 使用梯度提升树学习非线性和交互 [6]。二者分别构成可解释线性基准和强非线性表格基准。

2.2 FT-Transformer 与条件调制

对于第 j 个数值特征，FT-Transformer 的 tokenizer 为

$$T_{i,j} = b_j + x_{i,j}W_j, \quad (1)$$

其中 $W_j, b_j \in \mathbb{R}^d$ 。拼接可学习的 [CLS] token 后，Transformer 在特征 token 之间建模关系，并使用最终 CLS 表示完成预测 [4, 8]。

条件仿射调制使用外部状态生成缩放和平移参数，其思想与通用条件化层 FiLM 相近 [7]。本文没有声称复现 FiLM 的视觉任务设定，而是将这一机制作为金融市场状态进入模型的结构化对照。重要的是，本文的 B4 不修改 Transformer Encoder 各层内部的 LayerNorm；它只在 Encoder 输出的 CLS 表示经过 head LayerNorm 后执行一次条件仿射调制。

3 数据、标签与时点约束

3.1 股票池与样本切分

股票池为历史时点沪深北全 A，保留退市历史，不按当前上市状态回溯筛选。时间切分固定为：训练期从数据起点至 2021 年 12 月 31 日，验证期为 2022–2023 年，测试期为 2024–2025 年。正

式测试集覆盖 485 个交易日、4,863 只股票和 2,238,748 行观测。测试集只在 B2、B3、B4 三个 checkpoint 全部冻结后加载。

训练和验证分别从 7,231,111 与 2,083,083 行中采用固定系统抽样得到 500,000 与 250,000 行。抽样并非随机抽样，但覆盖训练期全部 2,917 个交易日和验证期全部 484 个交易日。测试使用完整样本，不做抽样。

3.2 公司特征与市场状态

公司输入共 50 个特征，包括 30 个 Alpha158-Core 行情技术特征，以及规模估值、盈利质量、成长、财务风险和流动性特征。模型输入按日进行行业中位数与全市场中位数两级填充、5 MAD 去极值和横截面 z-score；财务数据仅在公告信息可得后的下一交易日生效。

市场状态共 7 维：全 A 等权 1、5、20 日收益，20 日波动率，总成交额，20 日成交额相对值和上涨家数占比。标准化均值与样本标准差只用 2010–2021 年训练期拟合，同一交易日的所有股票共享同一个市场状态向量 m_t 。市场状态不计入 50 个公司特征。

3.3 标签与交易时点

信号在交易日 t 形成，主标签为从 $t+1$ 收盘进入、到 $t+6$ 收盘结束的完整五日复权收益：

$$y_{i,t} = \frac{\text{AdjClose}_{i,t+6}}{\text{AdjClose}_{i,t+1}} - 1. \quad (2)$$

训练目标 label_z 是每日横截面处理后的标签；评估 IC、Rank IC 和分组收益使用与信号日 t 对齐的原始收益 label_raw。这一显式对齐避免把 Wangwen 默认 close 快捷入口误当成相同的五日持有口径。

4 模型与训练

4.1 五模型对照

B0 Elastic Net

检验稀疏线性关系，并作为最低复杂度基准。

B1 LightGBM

检验非线性与特征交互。

B2 简化 FT-Transformer

复现数值特征 tokenizer、CLS 和 Transformer Encoder 核心模块，不声称复现原论文全部结构、数据集或调参协议。

B3 FT + 市场变量直接拼接

在 head LayerNorm 后的 CLS 表示上拼接 7 维市场状态，检验市场信息本身是否有增量。

B4 市场自适应 FT

使用市场状态生成 CLS 级条件缩放与平移，检验条件调制是否优于直接拼接。

4.2 B2、B3 与 B4 的精确实现

B2、B3、B4 共用 50 个独立的数值 tokenizer，每个特征映射为 8 维 token。加入可学习 CLS 后，输入序列形状为 51×8 。Encoder 包含 2 层 pre-LN Transformer、2 个注意力头（每头 4 维）和 $8 \rightarrow 16 \rightarrow 8$ 的 GELU 前馈网络，随后执行 Encoder final LayerNorm 并取 CLS。预测 head 再执行一次 LayerNorm。

B2 的输出为

$$\hat{y}_{i,t} = w_2^\top h_{i,t} + b_2, \quad (3)$$

其中 $h_{i,t} \in \mathbb{R}^8$ 。B3 直接拼接市场状态：

$$\hat{y}_{i,t} = w_3^\top [h_{i,t}; m_t] + b_3. \quad (4)$$

B4 的 conditioner 为 $7 \rightarrow 16 \rightarrow 16$ ：

$$q_t = \text{Dropout}_{0.1}(\text{GELU}(W_1 m_t + b_1)), \quad (5)$$

$$[\Delta\gamma_t; \beta_t] = W_2 q_t + b_2, \quad (6)$$

$$\tilde{h}_{i,t} = (1 + \Delta\gamma_t) \odot h_{i,t} + \beta_t, \quad (7)$$

$$\hat{y}_{i,t} = w_o^\top \tilde{h}_{i,t} + b_o. \quad (8)$$

其中 $\Delta\gamma_t, \beta_t \in \mathbb{R}^8$ 。Conditioner 末层 W_2, b_2 零初始化，使模型初始时退化为 B2 的 head 表示；零初始化的不是预测输出层。B2、B3、B4 的参数量分别为 2,049、2,056 和 2,449。

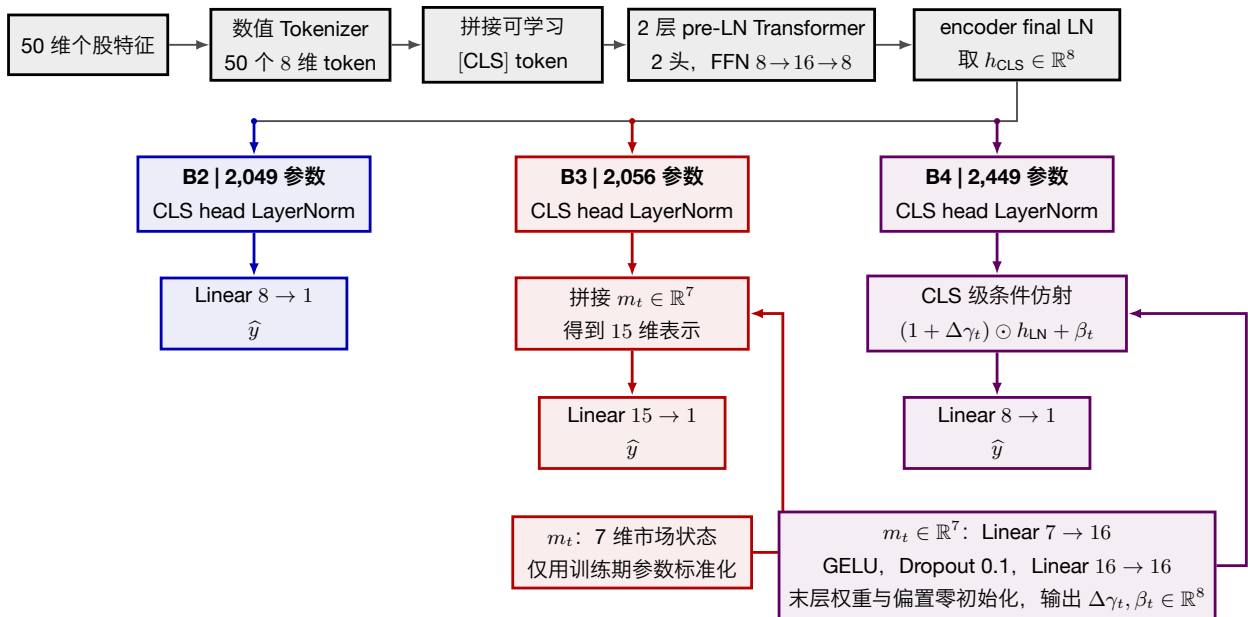


图 1: B2、B3、B4 的实际代码结构。B4 在最终 CLS 的 head LayerNorm 后调制，不改写 Encoder 内部 LayerNorm。

4.3 训练预算与选择边界

三种深度模型统一使用 AdamW (学习率 10^{-3} 、权重衰减 10^{-5})、Huber Loss ($\delta = 1$)、batch size 4,096、最大 10 个 epoch 和 patience 3。Attention、FFN 与残差 dropout 均为 0.1，B4 conditioner

dropout 为 0.1。固定 seed 为 20260722，以验证集 Huber Loss 选择 checkpoint；B2、B3 选择第 9 个 epoch，B4 在第 7 个 epoch 早停并选择第 4 个 epoch。训练未使用学习率调度、梯度裁剪或混合精度。

5 实验治理与评价指标

5.1 预冻结与版本证据

正式 P1 run 为 20260723T014153。preregistration.json 在测试数据加载前固定三模型、训练预算、验证选择规则和单一 seed；training_complete.json 记录三个 checkpoint 冻结时间；verification.json 记录预测键、行数、日期数、股票数、文件哈希和 Wangwen 对账。B0/B1 来自 P0.2 run 20260722T093202，B2 使用 P1 公平重跑，避免用原 P0.2 的五轮训练结果与新模型进行不等预算比较。

正式 P1 manifest 记录模型代码 commit 为 0a08beb，且当时 worktree 为 dirty。因而本文把 run manifest 与 code_manifest.json 的文件哈希作为权威版本证据，不把当前 HEAD 或后续 commit 描述成正式训练的干净快照。这一限制不改变已保存预测和指标，但降低了仅凭 Git commit 恢复完整当时代码的能力。

5.2 预测与经济指标

对每个交易日计算 Pearson IC 与 Spearman Rank IC。本文报告

$$\text{RankICIR} = \frac{\text{mean}_t(\text{RankIC}_t)}{\text{std}_t(\text{RankIC}_t)}, \quad (9)$$

其中标准差为样本标准差。为避免与常见年化口径混淆，本文把该值明确记为“日度未年化 Rank ICIR”，并同时给出 $\text{RankICIR}_{\text{ann}} = \sqrt{252} \text{RankICIR}$ 的机械换算；该换算不消除五日收益重叠或序列相关。样本外 R^2 使用横截面标准化标签 label_z，不等同于原始收益 R^2 。

每天按预测分数从低到高分 5 组和 10 组。Top 组换手率采用相邻日期持仓集合的对称成员变更率；修订后的评估代码把首日建仓记为缺失并从均值中排除。冻结主表仍保留原始 run 的数值，首日处理只影响 485 日中的 1 日且不改变模型结论；下一轮正式评估将使用修订口径。交易成本假设为单边 10bp：

$$r_{\text{top},t}^{\text{net}} = r_{\text{top},t} - c\tau_{\text{top},t}, \quad (10)$$

$$r_{\text{LS},t}^{\text{net}} = r_{\text{top},t} - r_{\text{bottom},t} - c(\tau_{\text{top},t} + \tau_{\text{bottom},t}), \quad (11)$$

其中 $c = 0.001$ 。由于相邻信号日对应的五日持有期高度重叠，本文只把这些数值解释为统一比较指标，不把它们逐日复利为可交易净值，也不据此声称可实现的年化收益。

6 样本外结果

6.1 没有单一模型全面胜出

表 1 与图 2 汇总 2024–2025 年完整测试集结果。B0 的 $label_z$ R^2 与 Top 成本后五日均值最高，B1 的多空成本后五日均值最高且 Top 换手最低；这是本文关于复杂度–收益关系的核心发现。B3 的 Rank IC 最高，为 0.113746；B2 的日度未年化 Rank ICIR 最高，为 0.907199。B4 未在任何主指标上领先。

模型	名称	R^2	Rank IC	ICIR	Top 净收益	多空净收益	Top 换手
B0	Elastic Net	0.0096	0.0999	0.816	1.158%	1.498%	34.1%
B1	LightGBM	0.0065	0.1130	0.853	1.078%	1.512%	30.4%
B2	FT-Transformer	0.0055	0.1106	0.907	0.837%	1.236%	38.4%
B3	FT + 市场拼接	0.0065	0.1137	0.856	0.991%	1.440%	35.6%
B4	市场自适应 FT	0.0049	0.1076	0.894	0.791%	1.130%	40.2%

表 1: 五模型统一样本外指标。Top 与多空净收益为扣除 10bp 简化成本后的平均五日观察；ICIR 未年化。

按 $\sqrt{252}$ 机械换算，B0–B4 的年化 Rank ICIR 依次为 12.958、13.549、14.401、13.590 和 14.199。主表保留冻结产物中的日度未年化列，以维持结果追溯；两种口径必须连同定义一起报告，不能把换算值解释为独立显著性证据。

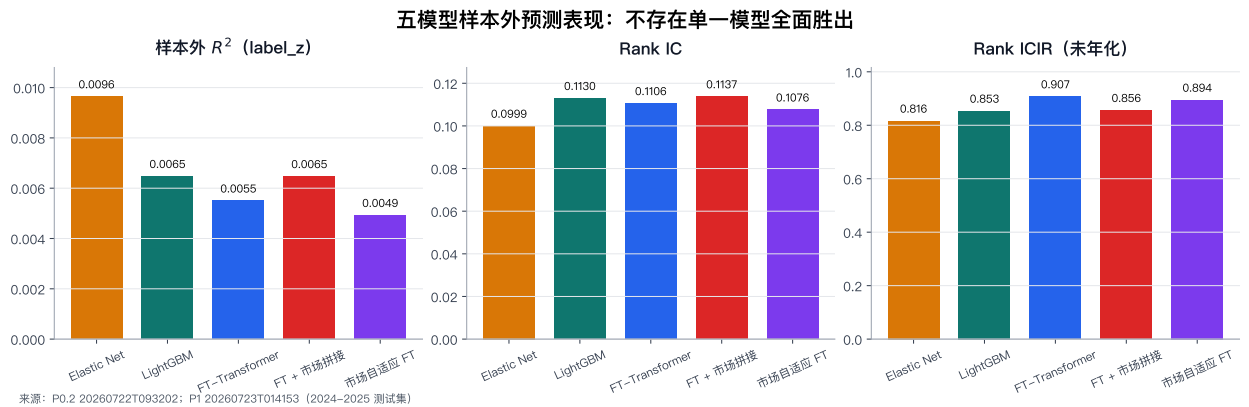


图 2: 五模型预测指标。不同指标由不同模型领先，复杂模型没有形成全面优势。

B3 相对公平重跑 B2 的 Rank IC 提高 0.003190， R^2 提高 0.000971，多空成本后五日均值提高 0.002043，但 Rank ICIR 降低 0.051085。由于只有一个冻结 seed，且没有置信区间或配对显著性检验，这些差异只能作为描述性增量，不能宣称其为稳定的结构性优势。

6.2 分组单调性与交易权衡

图 3 显示五个模型的十分组跨期平均原始五日收益均随预测分数严格递增，说明统一评估方向正确且横截面排序具有经济含义。不过最高组收益并不等同于综合模型优胜：B0 的 Top 组更高，B1 的多空成本后均值更高，B3 在深度模型内部改善收益与换手，B4 则同时表现为更低收益和更

高换手。

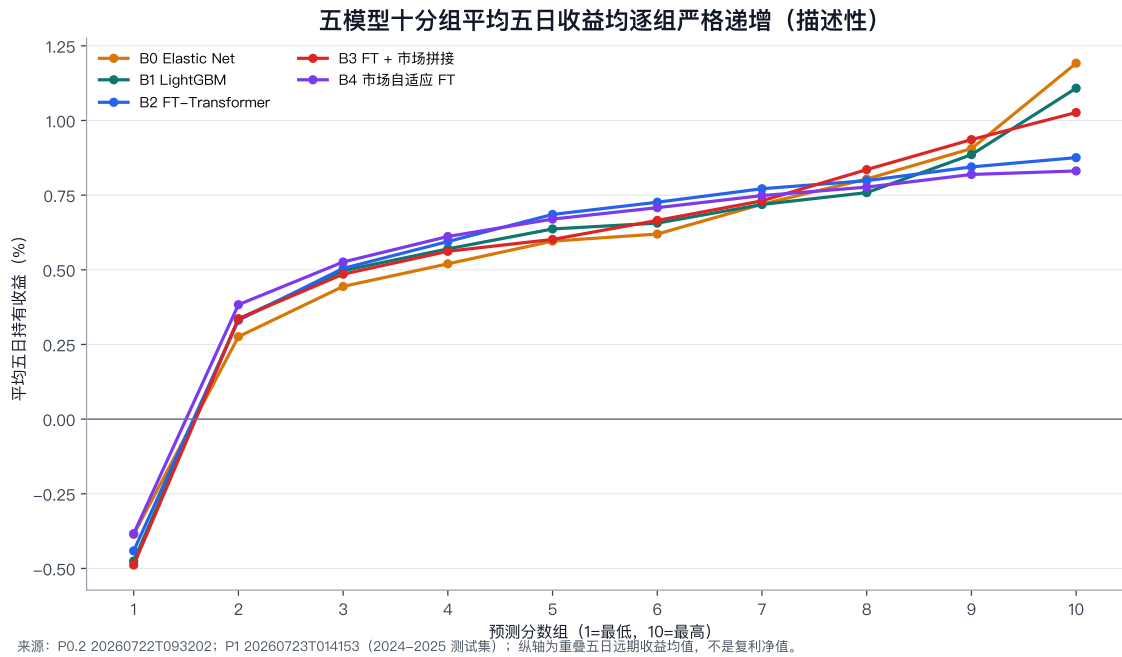


图 3: 十分组平均五日原始收益。每条曲线均严格递增, 但收益为重叠持有期观察。

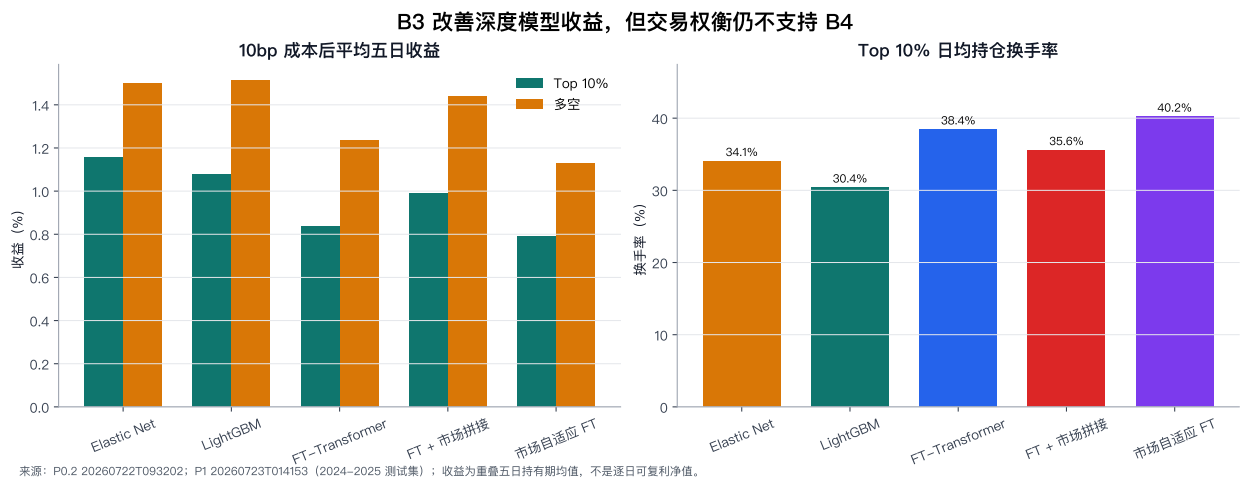


图 4: 10bp 简化成本后的平均五日收益与 Top 组换手率。该图不是净值曲线。

6.3 排序能力随时间变化

图 5 使用 63 个交易日滚动均值降低日频 Rank IC 噪声。五个模型在 2024 年末同步走弱、2025 年上半年同步增强, 说明共同市场环境对模型表现具有显著影响; 但领先关系也会变化, 不能从全样本均值推断任一模型在所有阶段占优。

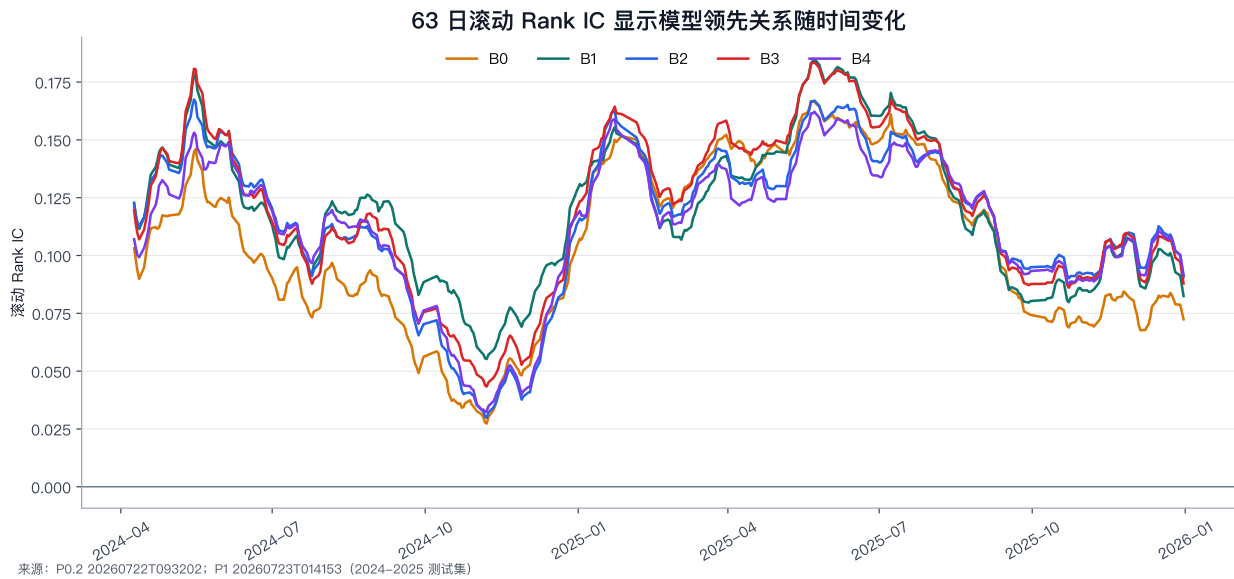


图 5: 63 日滚动 Rank IC。滚动窗口只用于展示，不参与模型选择。

7 市场状态消融

趋势和波动阈值只使用 2010-2021 年市场日期的三分位拟合。测试期包含下跌 149 日、中性 168 日、上涨 168 日；低波动 204 日、中波动 106 日、高波动 175 日；趋势切换日 78 日、稳定日 407 日。表 2 和图 6 仅比较 B2、B3、B4。

市场状态	日期数	B2	B3	B4
下跌	149	0.1001	0.1078	0.0889
中性	168	0.1115	0.1052	0.1112
上涨	168	0.1189	0.1276	0.1205
低波动	204	0.1059	0.1035	0.1032
中波动	106	0.1118	0.1087	0.1146
高波动	175	0.1153	0.1288	0.1084
稳定日	407	0.1114	0.1161	0.1076
切换日	78	0.1062	0.1015	0.1072

表 2: B2-B4 在不同市场状态下的平均 Rank IC。状态分组为描述性分析。

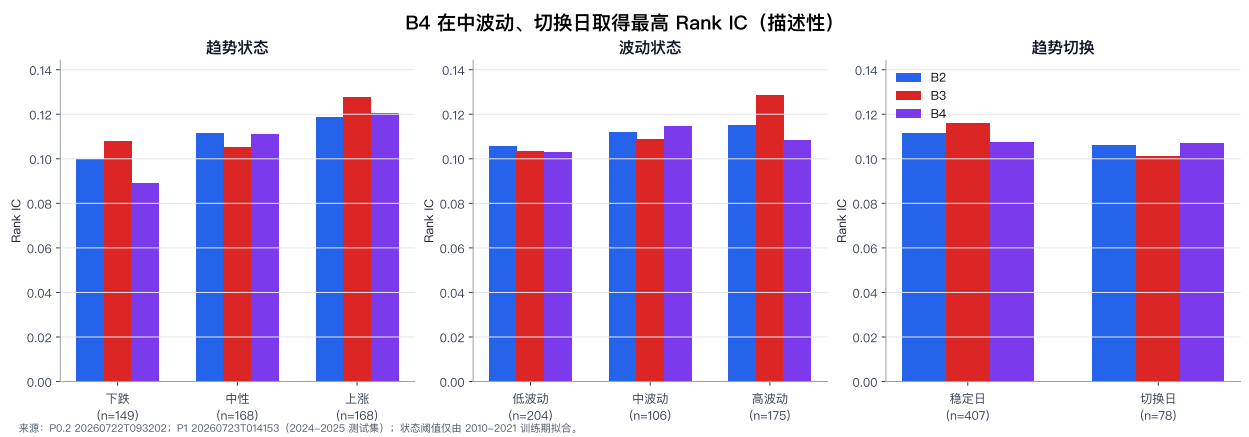


图 6: 市场状态 Rank IC。柱图使用零基线; 阈值只由训练期拟合。

B3 在下跌、上涨、高波动与稳定日领先；B2 在中性和低波动状态领先。相对 B2，B4 在上涨、中波动和切换日略高；但 B4 只有在中波动和切换日同时领先 B2、B3，在上涨状态仍低于 B3。B4 在高波动期明显弱于 B3 和 B2，因此 H3 中“增量主要来自高波动”的部分不成立。切换日 B4 相对 B2 的差异仅约 0.001，且切换样本只有 78 日，未做多重检验或统计推断，不能据此宣称机制已在切换阶段稳定有效。

8 假设判断与讨论

假设	结果	证据与解释
H1	部分支持，不支持“全面优于”	B1/B2/B3 的 Rank IC 高于 B0，但 B0 的 R^2 和 Top 成本后均值更高；模型优势依赖评价目标。
H2	不支持	B4 的总体 Rank IC、 R^2 、Top/多空净收益均低于 B2，换手更高；也未优于 B3。
H3	不支持，仅有局部描述性信号	B4 在上涨、中波动和切换日略高于 B2，但只在后两类状态同时领先 B2、B3；其高波动期更弱，且没有显著性检验。

表 3: 研究假设的样本外判断。

结果首先表明，模型复杂度与经济收益并不单调。Elastic Net 只有 50 个线性系数，却在 R^2 和 Top 组成本后均值上领先；LightGBM 在多空成本后均值上领先；三个 FT 变体只在部分排序指标上占优。对于当前低信噪比、强横截面标准化且训练预算受限的 50 因子场景，正则化线性结构可能具有更有利的偏差-方差权衡。这与文献中“机器学习优势依赖任务、特征和评估目标，而不存在普遍赢家”的判断一致 [4, 5]。

B4 失败可能来自三类机制。第一，条件调制只发生在 8 维最终 CLS 上，不能让市场状态参与 Encoder 内部的特征交互，形成明显信息瓶颈。第二，七维状态变量本身的横截面预测信息可能较弱；尤其 B3 的同日共享状态项在固定模型下只是共同偏移，无法直接改变当日排序。第三，500,000 行确定性系统抽样可能不足以稳定识别额外 conditioner 参数。以上均是由结构与证据边界导出的待检验假说，不是事后调参理由；现有单 seed 结果不能区分三者。下一轮应先预注册多 seed 和新的样本外区间，再比较 Encoder 层内调制、cross-attention 或更稳健的状态定义。

更重要的是，实验没有根据测试结果改变 B4 结构、方向或训练预算。保留无效创新，能够把“市场信息有描述性增量”与“条件调制有效”分开，也与 FT-Transformer 原论文关于表格任务中不存在普遍赢家的结论相一致 [4]。

9 局限与后续工作

1. **单一随机种子。** B2-B4 只使用一个预冻结 seed，细小差异不能视为稳定结构优势。后续应在不使用测试集选择的前提下预注册多个 seed，报告均值、离散度与配对比较。

2. **重叠五日收益。**每个信号日都对应一个五日持有收益，相邻观察高度重叠。现有成本后均值用于同口径比较，不是可逐日复利的策略净值。
3. **交易约束简化。**10bp 成本未覆盖冲击、税费、涨跌停、停牌、实际成交量与组合再平衡调度；冻结结果的换手率把首日建仓记为 1，修订代码已改为排除首日，下一轮正式评估将采用修订口径。
4. **状态分析为描述性。**三分位状态与切换定义预先冻结，但未做置信区间、多重检验或因果识别。
5. **训练预算有限。**深度模型在 CPU 上使用固定系统抽样训练，未扩大样本、参数或调参范围。当前结果只评价这一冻结预算下的实现。
6. **代码快照不完全。**正式 run 为 dirty worktree；核心架构、配置、预测和指标有哈希，但部分当时源码不能只靠 Git commit 恢复。后续正式实验应归档完整 source snapshot 或补丁。

优先级最高的增强不是增加更多模型，而是：预注册多 seed；构建非重叠持有组合或明确五日轮动组合；加入可交易约束与完整成本；为状态差异提供置信区间。只有这些验证通过后，才适合扩大 B4 的条件化位置或模型容量。

10 结论

本文在历史时点全 A、固定标签和严格时间切分下统一比较五类模型。最稳健的模式不是某个复杂模型胜出，而是复杂度-收益关系不单调：Elastic Net 在 R^2 和 Top 组成本后均值上领先，LightGBM 在多空成本后均值上领先，FT 变体没有提供一致经济增量。B3 的总体 Rank IC 最高，但差异在单 seed 条件下只能视为描述性；CLS 级条件仿射调制 B4 未在总体上优于 B2/B3，且在高波动期更弱。

因此，当前最可信的结论是：在本研究的 A 股 50 因子与冻结预算下，增加模型复杂度并不必然改善样本外预测或经济结果；市场信息进入方式与自适应机制仍需多 seed 和新样本外区间验证。这一结论以正式 run、预测哈希、统一评估与保留失败结果为基础，并明确区分预测能力、重叠五日比较指标和可交易净值。

A 证据索引

P2 结果集为 20260723T120956，由 `p2 --config configs/p2.toml` 生成。关键证据如下：

文件	作用
artifacts/p1/20260723T014153/metrics_summary.csv	五模型主指标, B0/B1 映射至 P0.2, B2/B3/B4 映射至 P1。
artifacts/p1/20260723T014153/market_state_rank_ic.csv	趋势、波动与切换状态的 Rank IC 和样本日数。
artifacts/p1/20260723T014153/preregistration.json	测试加载前冻结的模型、预算、选择指标与 seed。
artifacts/p1/20260723T014153/training_complete.json	三个 checkpoint 在测试加载前冻结的记录。
artifacts/p1/20260723T014153/code_manifest.json	正式运行涉及源文件的 SHA-256。
artifacts/p2/20260723T120956/manifest.json	图表、表格、来源文件与输出文件哈希。
artifacts/p2/20260723T120956/verification.json	五模型集合、CSV/JSON 一致性、分组单调性与结论门。

B 可复现命令

```
cd workspace
.venv/bin/python -m p02_baselines.p2_reporting --config configs/p2.toml
.venv/bin/python -m pytest -q
cd paper
TECTONIC_CACHE_DIR=/private/tmp/ai-fin-camp-p2-tectonic-cache \
tectonic -X compile --outdir build --outfmt pdf --print \
--untrusted main.tex
cp build/main.pdf market_adaptive_ft_transformer.pdf
```

参考文献

- [1] Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33(5), 2326–2377. <https://doi.org/10.1093/rfs/hhz123>.
- [2] Chen, G., Kim, K. A., and Rui, O. M. (2005). Price limits and stock market volatility in China. *Journal of Empirical Finance*, 12(4), 585–601. <https://doi.org/10.1016/j.jempfin.2004.04.003>.
- [3] Giglio, S., Kelly, B., and Xiu, D. (2022). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics*, 14, 337–368. <https://doi.org/10.1146/annurev-financial-101521-104735>.

- [4] Gorishniy, Y., Rubachev, I., Khrulkov, V., and Babenko, A. (2021). Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems*, 34, 18932–18943. <https://arxiv.org/abs/2106.11959>.
- [5] Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>.
- [6] Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- [7] Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. (2018). FiLM: Visual reasoning with a general conditioning layer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). <https://doi.org/10.1609/aaai.v32i1.11671>.
- [8] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [9] Yang, X., Liu, W., Zhou, D., Bian, J., and Liu, T.-Y. (2020). Qlib: An AI-oriented quantitative investment platform. *arXiv preprint arXiv:2009.11189*. <https://arxiv.org/abs/2009.11189>.
- [10] 上海证券交易所 (2023)。《上海证券交易所交易规则 (2023 年修订)》。 https://www.sse.com.cn/lawandrules/sselawsrules2023/repeal/rules/c/c_20230217_5716668.shtml.
- [11] Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.