

2026年8月11日

商品指数机器学习择时研究

——宏观量化系列深度报告（三）

分析师：

明道雨
从业资格证号：F03092124
投资咨询证号：Z0018827
电话：021-68758120
邮箱：mingdy@qh168.com.cn

摘要：

➤ 大宗商品价格受宏观流动性、库存周期、风险偏好与突发事件共同驱动，短周期方向预测长期面临信噪比低、体制切换频繁与过拟合风险高等挑战。本报告以南华商品综合指数（NHCI）为标的，在严格滚动样本外（Rolling Walk-Forward）约束下，构建覆盖量价、宏观、基本面、风险偏好与事件冲击的多因子机器学习体系，分别对未来5个交易日与20个交易日的涨跌方向进行建模，并形成全样本多空与高置信分层两类交易规则。

➤ 主要结论如下：

第一，5日单模型滚动方向正确率57.65%，AUC约0.565，模型总体合格，接近良好，具备稳定预测增量；可搭配趋势 / 均值回归策略组合使用，适合辅助择时。5日融合模型将方向正确率提升至约61.32%，整体良好，接近优秀；接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，常态化周度级别资产配置和仓位调整。20日方向滚动正确率约为59.37%，距良好线60%仅差约0.63个百分点，但区分度指标AUC达到约0.639，达到优秀水平且显著优于5日任务。显示中周期方向更可分，接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，可作为月度资产配置实盘仓位调整方案。

第二，回测层面，在全样本 $\text{thr}=0.5$ 口径下，5日买入持有净值约1.47，年化6.48%，夏普比率为0.49，盈亏比1.17，最大回撤20.63%。5日融合多空期末净值约2.36，年化15.04%，夏普比率为1.01，盈亏比1.42，均达到合格标准；但是最大回撤为30.5%，高于25%的合格标准，主要由于2022年模型阶段性失效导致的。

5日单模型净值为约1.71，年化收益约9.1%、夏普约0.66、盈亏比约1.24，最大回撤35.13%，虽然获得了超额收益，但是夏普和回撤控制欠佳。5日单模型和融合模型高置信软阈值（0.53—0.54）净值和年化收益略高于5日单模型和融合模型全样本多空、但夏普、盈亏比和回撤控制明显改善；随着阈值进一步收紧，交易次数下降、虽然回撤控制明显提升，但面临着年化收益、夏普、盈亏比大幅下降的问题。对5日任务，更合理的定位或许是：软阈值做收益、紧阈值做风控过滤，而非单一阈值同时满足高频与高置信。

非重叠口径下，20日模型多空略优于持有；高置信条件下，年化收益率小幅提升，但夏普、盈亏比大幅改善和提升至合格和优秀水平，最大回撤明显下降。0.57附近综合最佳，年化11.22%，夏普约0.99，盈亏比2.39，期末净值3.28，最大回撤16.3%，改善非常明显，相关指标已提升至优秀水平。相对于5日，20日高置信在“正确率提升、盈亏比、回撤”三维上更均衡，具备较好的策略投资价值。

本研究仍旧存在诸多不足。一是因子层面，在2022、2025年年末和2026年，模型在商品强趋势反转、政策急转弯、地缘冲击、流动性骤变阶段连续失效，关键因子缺失和因

子质量欠缺，并不能被学习到。二是数据层面，关键因子可能具备，但是因子数据质量、因子数据影响持续时间（特别是事件哑变量）难以界定，导致关键因子信息丢失。三是模型训练方面，单模型仅接近良好，并未达到良好或者优秀的水平，需要提升数据质量和训练方法进一步优化模型AUC、方向准确率。

后续研究主要聚焦以下方面。一是完善实时监控滚动准确率、AUC、分年失效警报与阈值漂移检验，控制风险和回测。二是聚焦单模型有效信息扩充，包括挖掘关键有效因子、提升因子数据质量。三是分体制自适应，优化模型阶段性失效风险。四是重点引入体制自适应与新信息源，继续优化提升5日和20日单模型指标至良好和优秀水平。五是该模型体系推进扩展到可交易的具体商品。

➤ 关键词：南华商品指数；机器学习；方向预测；高置信策略；商品择时。

目录

1. 研究背景与研究目标	
1.1. 研究背景	4
1.2. 研究目标	4
1.3. 研究意义	4
2. 研究原理与方法框架	
2.1. 标签形式	5
2.2. 滚动 Walk-Forward	5
2.3. 模型原理	5
2.4. 验收框架	6
2.5. 指标介绍	6
3. 数据来源与数据处理方法	
3.1. 标的与样本	7
3.2. 因子体系	7
3.3. 标签与价格处理	7
3.4. 缺失值、对齐与防泄漏	8
3.5. 训练、验证、测试集设计	8
4. 模型选择与训练结果	
4.1. 候选模型与选择逻辑	8
4.2. 模型运行结果	8
4.2.1. 主结果总览	8
4.2.2. 模型表现	9
5. 回测规则与结果	
5.1. 回测规则	11
5.2. 回测结果	12
5.2.1. 5 日回测结果	12
5.2.2. 20 日回测结果	14
6. 总结	
6.1. 研究结论	15
6.2. 研究不足	16
6.3. 后续计划	16

1.研究背景与研究目标

1.1.研究背景

商品市场是宏观定价与产业供需的交汇点。对综合商品指数而言，其涨跌既反映能源、工业品与农产品等板块的加权表现，也映射全球增长预期、美元流动性、国内地产与制造业周期以及风险事件冲击。传统技术指标与线性多因子模型在趋势市中往往有效，但在震荡、政策扰动与板块轮动加剧阶段，稳定性显著下降。

近年来，梯度提升树（XGBoost、LightGBM）等非线性方法在截面选股与宏观择时中得到广泛应用。然而，若缺乏严格的时间序列切分与信息时点控制，极易因未来函数与重复调参产生虚高样本外表现。商品指数研究更需强调：标签与特征的对齐、宏观数据发布滞后、以及训练窗长度对过拟合的约束。

在此背景下，围绕 NHCI（南华商品指数）构建一套可审计、可滚动更新、可对照基准的机器学习择时框架，具有明确的方法论价值与潜在应用价值。一方面可以作为对未来整体商品价格走势的定量判断；另一方面可以作为量化大类资产配置策略中商品配置和仓位管理的重要依据。

1.2.研究目标

本报告设定如下研究目标：

预测目标：在无未来函数前提下，预测 NHCI 未来 5 个交易日、20 个交易日的涨跌方向，并输出上涨概率。

方法论目标：建立仅允许滚动 Walk-Forward 的训练—验证—测试流程，明确训练集长度、间隔（gap）与仅样本外评级的验收标准。

因子目标：整合量价、宏观、库存基本面、风险偏好、期限结构与事件因子，形成可解释的特征池与折内筛选机制。

策略目标：将模型概率映射为全样本多空与高置信交易规则，并与买入持有、双均线、线性多因子等基准比较净值与风险收益特征。

1.3.研究意义

对研究体系而言，本项目将商品指数短周期预测从“一次性回测”推进到“可复现流水线”，使方向正确率、AUC、高置信提升、盈亏比、夏普与回撤等指标可在统一口径下横向比较。

对资产配置而言，由于基准的收益率且在高置信子集上具备稳定的盈亏比与可控回撤，可以作为资产配置的重要依据，提高大类资产配置的整体收益和回撤控制。

对投资实践而言，即便绝对正确率仅略高于随机，只要在高置信子集上具备稳定的盈亏比与可控回撤，仍可能形成有价值的仓位过滤模块。

对风险管理而言，分年、分体制与 2022 年后子样本的表现揭示了模型失效区间，有助于提示尾部风险与模型监控阈值。

对产品设计而言，5 日与 20 日双期限结果可为不同换手约束的商品增强、CTA 叠加或指数择时产品提供差异化信号频率选择。

2. 研究原理与方法框架

2.1. 标签形式

记交易日为 t ，指数收盘价为 P_t 。定义：

$$r_{t,h} = \ln(p_{t+1+h}/p_{t+1})$$

$$y_{t,h} = 1\{r_{t,h} > 0\}$$

其中 $h=5$ 或 $h=20$ 。即：在决策日 t 给出信号后，于 $t+1$ 收盘开仓，持有 h 个交易日，于 $t+1+h$ 收盘平仓。该定义保证特征使用截至 t 的信息，避免“当日收盘既做特征又做成交价”的前视偏差。

模型输出上涨概率 $\hat{p}_t = P_r(y_{t,h} = 1|F_t)$ ，并据此形成仓位 $s_t \in \{+1, -1, 0\}$ 。策略单期盈亏为 $pnl_t = s_t \cdot r_{t,h}$ ，净值由简单收益复利得到： $NATNAV_T = \prod_{t \leq T} (1 + \expml(pnl_t))$ 。

2.2. 滚动 Walk-Forward

金融市场存在显著的非平稳性。若采用随机 K 折交叉验证，时间顺序被打乱，验证集可能“看见”未来体制，导致指标虚高。滚动 Walk-Forward 要求：每一折仅用过去窗口训练、用紧邻验证窗调参、用其后的测试窗评分，并沿时间向前滚动。该方法更接近真实投研中的“每月/每周更新模型”过程，是本报告的硬性约束。

2.3. 模型原理

本研究主模型为树模型集成：

XGBoost / LightGBM：通过梯度提升拟合非线性交互，适合处理缺失值与异质因子；采用浅树、较强正则与早停，以抑制过拟合。

Logistic 回归：在标准化特征上提供线性可分基准与概率校准参考。

动量先验：短窗动量经 sigmoid 变换作为先验概率，与机器学习概率在验证集上混合，用于捕捉趋势延续成分。

折内特征筛选：每折仅在训练集上按 Rank IC 选择 Top-N 因子，避免用全样本 IC 挑因子。

2.4.验收框架

AUC 和方向准确率分级框架：

图 1 验收框架

任务	指标	淘汰	合格	良好	优秀	顶尖
5日	AUC	<55%	≥55%	≥58%	≥61%	≥65%
	方向准确率	<55%	≥55%	≥58%	≥62%	≥66%
20日	AUC	<57%	≥57%	≥60%	≥63%	≥67%
	方向准确率	<56%	≥56%	≥60%	≥64%	≥68%

资料来源：东海期货研究所，同花顺 iFinD

辅助维度包括：高置信相对全样本提升（≥2 个百分点）、盈亏比（合格≥1.6、优质≥2.0）、夏普（合格≥0.8）、最大回撤（幅度≤25%）、过拟合缺口（训-测<8 个百分点）等。评级仅使用样本外指标。

2.5.指标介绍

方向正确率（Directional Accuracy）：对样本外序列，将概率按阈值（常用 0.5，或折内验证集最优阈值）映射为涨跌预测，与真实方向比较 $\text{DirAcc} = \frac{1}{N} \sum_1^N 1\{\hat{y} - y_i\}$ 。滚动正确率指各测试折正确率的均值；池化正确率指合并全部 OOS 点后的正确率。二者在折大小不均时可能略有差异。

AUC：AUC 衡量概率排序能力。随机抽取一正一例，模型给正例更高概率的比例。AUC 对阈值不敏感，适合判断“模型是否学到方向信息”。

IC / ICIR: IC 为预测收益（或概率）与实现收益的截面/时序相关系数；ICIR 为 IC 均值与波动之比。商品指数为单资产时序任务，IC 更多作为因子筛选与辅助监控，而非主评级。

高置信准确率与 Lift: 在 $p \geq up$ 或 $p \leq dn$ 的子集上计算方向正确率，并定义： $Lift = Acc_conf - Acc_all$ 。Lift 显著为正，说明概率两端更可靠，适合作为仓位过滤器。

过拟合缺口: 训练集方向正确率与样本外正确率之差。差值过大表明模型记住了训练噪声。本研究以 8% 作为警戒/淘汰参考。

3. 数据来源与数据处理方法

3.1. 标的与样本

研究对象为南华商品综合指数。日度面板样本约 3493 个交易日（2012-03-15 至 2026-07-31）；由日度抽样得到非重叠 5 日 bar 约 698 个；周度面板约 737 周。5 日标签上涨占比约 52.9%，接近均衡，降低了类别极端不平衡带来的评估扭曲。

3.2. 因子体系

特征按六大类组织，5 日任务特征池约 249 个：

量价技术因子：动量、波动、反转、量价结构等。用于刻画趋势与短期过度反应。

宏观与流动性因子：海外与国内宏观变量。处理上对海外数据采取“+1 交易日并对齐下一 A 股交易日”的时点规则，避免把未发布数据当作已知。

风险偏好与跨资产因子：原油风险、资产比率等，捕捉 risk-on/risk-off 切换。

市场结构因子：按南华权重归一化的合成基差率、合成期限价差，反映基差与近远月结构。

行业基本面因子：库存、产量、开工等，其中钢铁等相关序列在若干阶段显示较高 IC。

事件哑变量：美联储货币政策、鹰鸽表态、地缘冲突、疫情、天气等，用于吸收尾部冲击均值位移。

3.3. 标签与价格处理

价格字段使用 nhci_close。标签收益严格按收盘价计算：

5 日： $r = \ln(P_{t+6}/P_t)$

20 日： $r = \ln(P_{t+21}/P_t)$

方向标签为收益是否为正。该处理对应“次日收盘开仓、持有期满收盘平仓”的可执行近似，而非开盘价或盘中均价成交。

3.4.缺失值、对齐与防泄漏

宏观与基本面数据天然存在发布滞后与缺失。面板构建阶段完成日历对齐与 as-of 处理；模型训练阶段树模型可容忍部分缺失，线性模型则在折内用训练集中位数填充。所有因子筛选、标准化参数、阈值与混合系数仅在当期训练/验证窗内估计，严禁使用测试折信息。

3.5.训练、验证、测试集设计

5 日 bar 任务：

图 2 训练、验证、测试集

集合	长度	用途
训练集	70 bars (约350个交易日)	拟合XGB+LGB+Logit
验证集	12 bars (约60个交易日)	调权重、先验混合、阈值
测试集	1 bar	仅OOS评分
滚动	step=2筛选, step=1加密; gap=1	防止标签重叠污染
OOS规模	307个点 (约2013-12至2026-07)	主评估样本

资料来源：东海期货研究所，同花顺 iFind

20 日周频任务：在周度面板上滚动训练，OOS 约 579 个点；净值图为避免 20 日标签重叠，采用每 4 周取 1 个决策点的非重叠口径。

4.模型选择与训练结果

4.1.候选模型与选择逻辑

研究过程比较了线性 Ridge 多因子、双均线规则、单模型树集成、滞后因子门控、以及多模型概率融合。选择逻辑并非追求训练集拟合最优，而是：1、样本外方向正确率与 AUC，2、2022 年后子样本稳定性，3、高置信分层是否带来 Lift 与盈亏比改善，4、相对基准的净值与回撤，6、是否可解释、可维护、可滚动更新。

4.2.模型运行结果

4.2.1.主结果总览

图 3 模型运行结果总览

类型	方向正确率	AUC	框架评级	说明
5日单模型	57.65%	0.565	合格，接近良好	当前单模型最佳
5日单模型 (lags_gate)	56.00%	0.584	合格，局部良好	AUC较高、方向一般
5日多模型融合	61.32%	0.58	良好，准确率接近优秀	非单模型
20日单模型	59.37%	0.639	良好，局部优秀	准确率距良好差0.63pp

资料来源：东海期货研究所，同花顺 iFinD

4.2.2.模型表现

5 日单模型在 307 个 OOS 点上滚动方向正确率 57.65%，AUC 约 0.565，模型总体合格，接近良好，具备稳定预测增量；可搭配趋势 / 均值回归策略组合使用；适合辅助择时。5 日融合模型将方向正确率提升至约 61.32%，整体良好，接近优秀；接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，常态化资产配置。高置信（0.62/0.38）准确率约 63.6%，盈亏比约 1.99，回撤约-5.7%。相对单模型，融合更像“信号增强器”，通过不同训练设定与跨期先验分散单一规格风险。但其复杂度更高，解释与维护成本上升。

图 4 5 日单模型 vs 多模型融合回测结果对比

项目	单模型	融合模型
方向正确率	57.65%	61.32%
AUC	0.5646	0.5801
2022 后方向	48.18%	53.64%
高置信默认 (0.62/0.38) 准确率	57.78%	63.64%
高置信夏普	0.32	1.92
高置信盈亏比	1.09	1.99
高置信最大回撤	-14.13%	-5.45%

资料来源：东海期货研究所，同花顺 iFinD

图 5 5 日单模型与融合模型按年份对比

年份	单模型	融合模型
2013	50.00%	0.00%
2014	62.50%	58.33%
2015	58.33%	70.83%
2016	72.00%	72.00%
2017	66.67%	50.00%
2018	44.00%	56.00%
2019	58.33%	83.33%
2020	83.33%	83.33%
2021	60.00%	56.00%
2022	37.50%	37.50%
2023	58.33%	58.33%
2024	50.00%	62.50%
2025	41.67%	50.00%
2026	57.14%	64.29%

资料来源：东海期货研究所，同花顺 iFinD

20 日模型滚动方向正确率 59.37%，AUC 约 0.639，模型整体良好，接近优秀，显示中周期方向更可分，接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，中等仓位常态化实盘。高置信默认阈值下，周频子集正确率约 67.4%，Lift 约 8.0 个百分点，夏普约 1.10，盈亏比约 2.03，回撤约-23.8%，整体更接近“可交易置信策略”的要求。非重叠净值口径下，模型全样本多空期末净值约 2.52，略高于买入持有的 2.40；高置信 0.57 阈值期末净值约 3.28，综合最优。

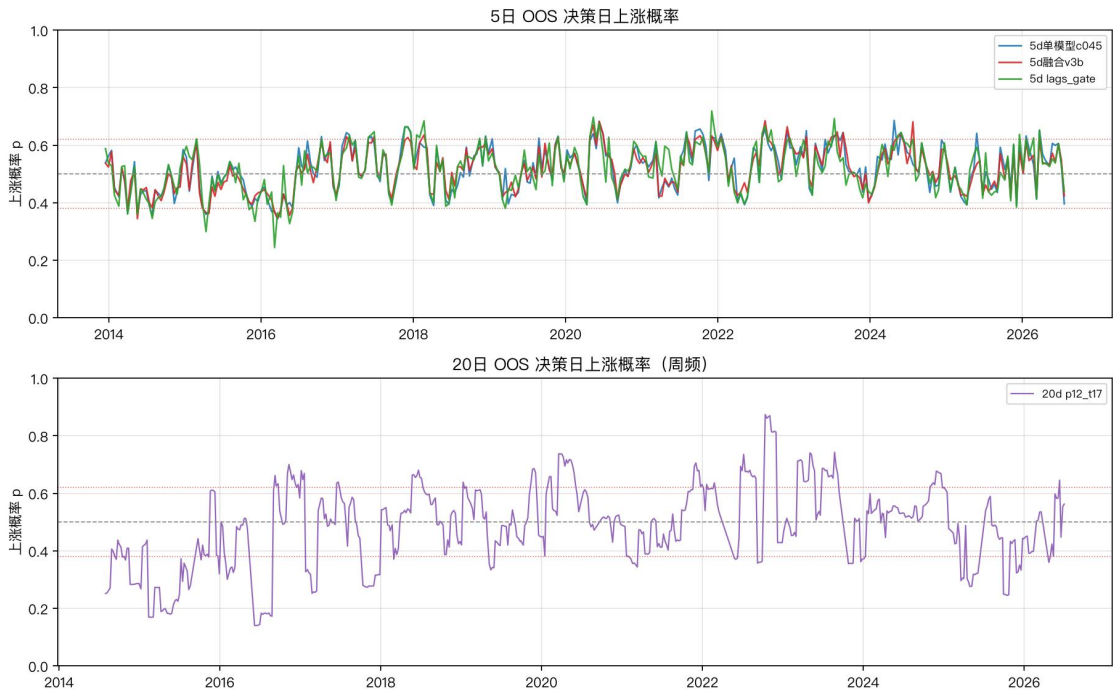
图 6 20 日模型回测结果

指标	数值	框架对照
滚动方向正确率	59.37%	良好 $\geq 60\%$ $\rightarrow$ 合格接近良好 (差 0.63%)
池化方向正确率	59.41%	—
OOS AUC	0.6386	优秀, 相对 5 日更强
高置信方向	67.38%	lift 7.97% ($\geq 2pp \rightarrow$ 通过)
高置信夏普	1.1	合格 $\geq 0.8 \rightarrow$ 通过
高置信盈亏比	2.03	合格 ≥ 1.6 / 优秀 $\geq 2.0 \rightarrow$ 优秀
最大回撤	-23.82%	$\leq 25\% \rightarrow$ 通过

资料来源：东海期货研究所，同花顺 iFinD

OOS 上涨概率序列显示，模型输出多数分布在 0.4—0.6 之间，极端概率占比较低，这与商品短周期低信噪比一致。滚动准确率与 AUC 随时间波动，在 2016—2020 等阶段相对更高，2022 前后出现下台阶，与分年正确率相互印证。对风控而言，可把“滚动准确率跌破阈值”作为模型降权或停用触发器。对于资产配置或者投资而言，可把“滚动准确率跌破阈值”作为停用或者空仓的触发器，以降低回撤。

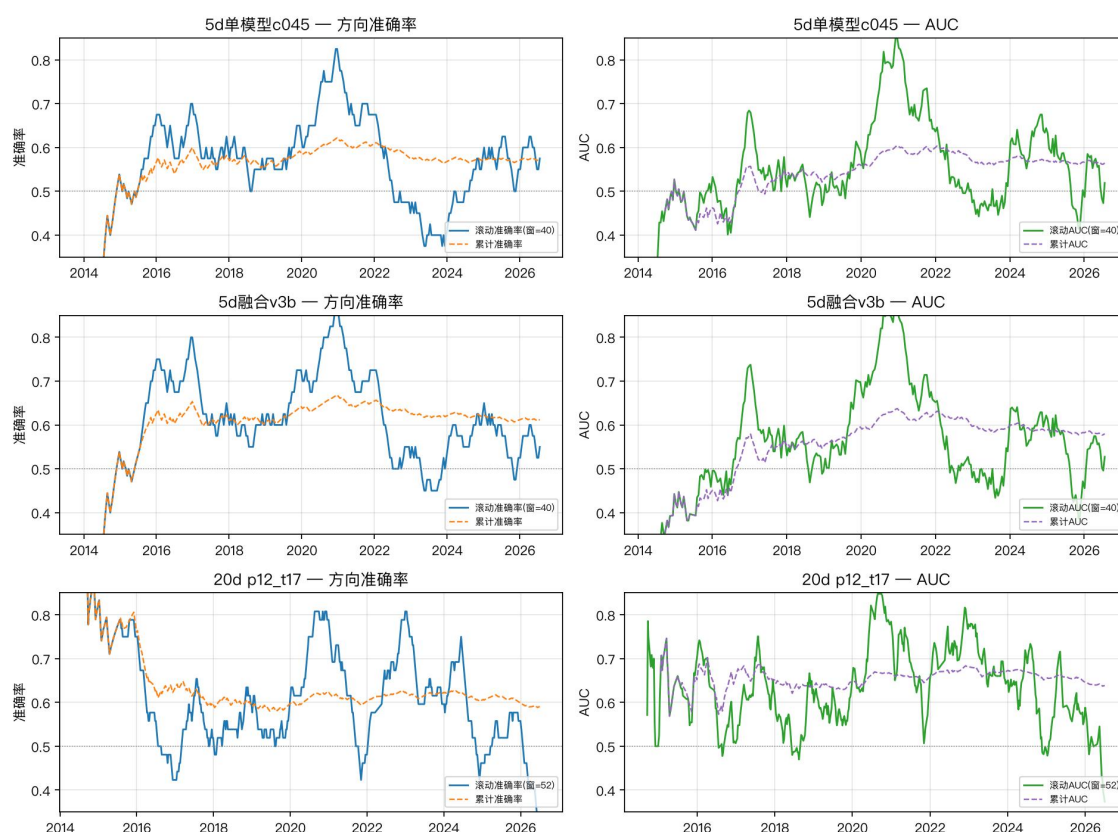
图 7 OOS 上涨概率曲线



资料来源：东海期货研究所，同花顺 iFinD

图8 OOS 准确率与 AUC 曲线

OOS 决策序列：准确率与 AUC（滚动 / 累计）



资料来源：东海期货研究所，同花顺 iFind

5.回测规则与结果

5.1.回测规则

回测标的与仓位：回测交易的是 NHCI 指数方向，仓位取+1（做多）、-1（做空）、0（空仓）。这是指数点位层面的信号回测，不等价于某一具体期货主力合约账户的成交回报。

买卖价格：本回测买入与卖出均按南华商品指数收盘价计算，不使用开盘价、最高最低价或盘口价：5日开仓价为 T+1 收盘、平仓价为 T+6 收盘；20日开仓价为 T+1 日收盘、平仓价为 T+21 收盘。未建模手续费（主要因为中低频，成本不到 1%，高置信成本更低）、滑点、冲击、保证金成本。

信号规则：全样本多空： $p \geq 0.5$ 做多，否则做空，始终满仓。

高置信： $p \geq up$ 做多， $p < 1-up$ 做空，中间空仓。默认 $up=0.62$ ；并系统扫描 0.53—0.59 等软阈值。

基准策略：买入持有，每期持有多头指数收益；双均线，短长期均线交叉，信号滞后一期；多因子线性 Ridge，前 70% 样本拟合，后段按预测符号交易。所有策略在同一决策日历上比较，以保证公平性。

5.2. 回测结果

5.2.1.5 日回测结果

在全样本 $\text{thr}=0.5$ 口径下，买入持有净值约 1.47，年化 6.48%，夏普比率为 0.49，盈亏比 1.17，最大回撤 20.63%。5 日融合多空期末净值约 2.36，年化 15.04%，夏普比率为 1.01，盈亏比 1.42，均达到合格标准；但是最大回撤为 30.5%，高于 25% 的合格标准，主要由于 2022 年模型阶段性失效导致的。5 日单模型净值为约 1.71，年化收益约 9.1%、夏普约 0.66、盈亏比约 1.24，最大回撤 35.13%，虽然获得了超额收益，但是夏普和回撤控制欠佳。双均线与线性 Ridge 显著落后。结果表明：即便方向正确率仅接近良好线，单模型和融合模型的概率信号仍可能通过多空对称持仓转化为超额净值，但这部分超额在 2022 年后方向正确率仅约 48.2%，暴露体制切换下的脆弱性，同时也伴随着较大回撤，不宜直接等同于“稳健 alpha”。分年看，2016、2017、2020 等年份较强，2018、2022、2025 明显偏弱。这与商品市场在贸易摩擦、地缘冲突、疫情后通胀冲顶回落、国内需求预期摆动等阶段的结构变化相一致，提示单一静态因子权重难以穿越全部体制。

图 9 5 日全样本多空净值曲线



资料来源：东海期货研究所，同花顺 iFind

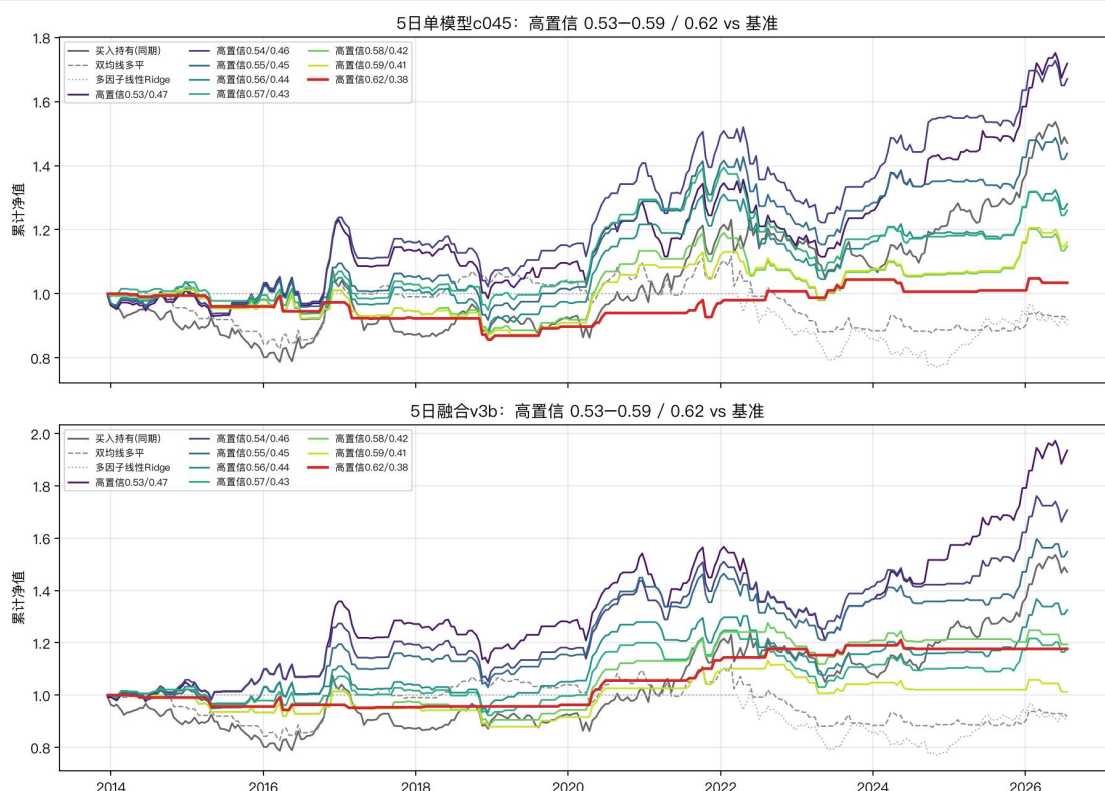
图 10 5 日全样本多空净值对比

策略	年化收益	夏普	盈亏比	最大回撤	期末净值
买入持有(同期)	6.48%	0.49	1.17	-20.63%	1.47
双均线多空	-8.68%	-0.53	0.8	-55.50%	0.57
多因子线性Ridge	-1.67%	-0.16	0.88	-24.57%	0.9
5d单模型多空	9.14%	0.66	1.24	-35.13%	1.71
5d融合多空	15.04%	1.01	1.42	-30.50%	2.36

资料来源：东海期货研究所，同花顺 iFinD

5 日单模型高置信软阈值 (0.53—0.54) 降低交易频率，净值和年化收益略高于 5 日单模型全样本多空、但夏普、盈亏比和回撤控制明显改善；但随着阈值进一步收紧，交易次数下降、虽然回撤控制明显提升，但面临着年化收益、夏普、盈亏比大幅下降的问题。5 日融合高置信软阈值 (0.53—0.54)，净值和年化收益明显低于 5 日融合全样本多空，但高于 5 日单模型高置信水平；随着阈值进一步收紧，交易次数下降、虽然回撤控制明显提升，但与单模型一样，也面临着收益、夏普、盈亏比大幅下降的问题。因此，对 5 日任务，更合理的定位或许是：软阈值做收益、紧阈值做风控过滤，而非单一阈值同时满足高频与高置信。

图 11 5 日高置信阈值净值曲线



资料来源：东海期货研究所，同花顺 iFinD

图 12 5d 单模型净值对比

策略	年化收益	夏普	盈亏比	最大回撤	期末净值	交易次数
买入持有(同期)	6.48%	0.49	1.17	-20.63%	1.47	307
高置信0.53/0.47	9.22%	0.76	1.34	-19.96%	1.72	234
高置信0.54/0.46	8.72%	0.75	1.36	-19.60%	1.67	213
高置信0.55/0.45	6.09%	0.6	1.3	-19.14%	1.44	189
高置信0.56/0.44	4.11%	0.44	1.21	-20.21%	1.28	168
高置信0.57/0.43	3.84%	0.42	1.22	-23.05%	1.26	146
高置信0.58/0.42	2.28%	0.29	1.15	-17.69%	1.15	123
高置信0.59/0.41	2.47%	0.32	1.18	-13.67%	1.16	106
高置信0.62/0.38	0.55%	0.12	1.09	-14.32%	1.03	45

资料来源：东海期货研究所，同花顺 iFinD

图 13 5 日融合模型净值对比

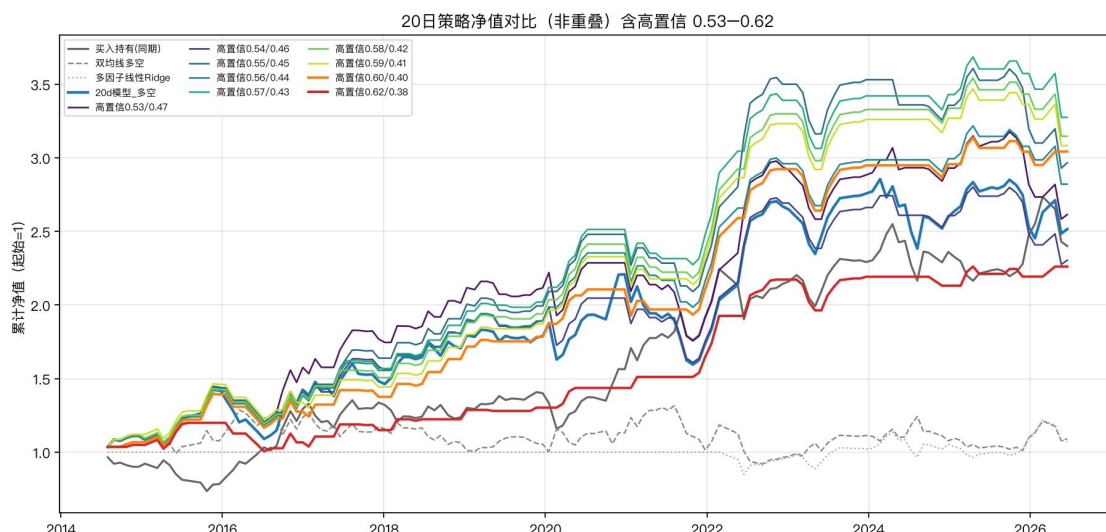
策略	年化收益	夏普	盈亏比	最大回撤	期末净值	交易次数
买入持有(同期)	6.48%	0.49	1.17	-20.63%	1.47	307
高置信0.53/0.47	11.35%	0.92	1.45	-22.80%	1.94	223
高置信0.54/0.46	9.10%	0.8	1.41	-17.33%	1.71	199
高置信0.55/0.45	7.38%	0.7	1.38	-17.36%	1.55	178
高置信0.56/0.44	4.70%	0.51	1.27	-17.32%	1.33	150
高置信0.57/0.43	2.72%	0.34	1.18	-17.35%	1.18	126
高置信0.58/0.42	2.93%	0.39	1.27	-12.34%	1.19	92
高置信0.59/0.41	0.20%	0.06	1.02	-13.36%	1.01	75
高置信0.62/0.38	2.69%	0.62	1.99	-5.67%	1.18	33

资料来源：东海期货研究所，同花顺 iFinD

5.2.2.20 日回测结果

非重叠口径下，20 日模型多空略优于持有；高置信 0.53—0.62 曲线显示，年化收益率小幅提升，但夏普、盈亏比大幅改善和提升至合格和优秀水平，最大回撤也明显下降。其中，0.57 附近综合最佳，年化 11.22%，夏普约 0.99，盈亏比 2.39，期末净值 3.28，最大回撤下降至 16.3%，指标改善非常明显，相关指标已提升至优秀水平。在 0.58—0.60 一带接近，整体较为优秀。相对于 5 日，20 日高置信在“正确率提升、盈亏比、回撤”三维上更均衡，具备较好的策略投资价值。

图 14 20 日非重叠口径下净值



资料来源：东海期货研究所，同花顺 iFinD

图 15 20 日非重叠口径下净值对比

策略	年化收益	夏普	盈亏比	最大回撤	期末净值	交易次数
买入持有(同期)	8.17%	0.58	1.43	-23.89%	2.4	145
双均线多空	0.78%	0.13	1.04	-30.01%	1.09	145
多因子线性Ridge	0.61%	0.11	1.09	-15.21%	1.07	50
20d模型_多空	8.62%	0.6	1.46	-27.76%	2.52	145
高置信0.53/0.47	9.00%	0.68	1.65	-23.20%	2.62	115
高置信0.54/0.46	7.77%	0.61	1.58	-21.41%	2.3	107
高置信0.55/0.45	10.24%	0.83	1.95	-18.79%	2.97	98
高置信0.56/0.44	9.75%	0.85	2.03	-16.53%	2.82	90
高置信0.57/0.43	11.22%	0.99	2.39	-16.53%	3.28	83
高置信0.58/0.42	10.82%	0.96	2.36	-16.53%	3.15	80
高置信0.59/0.41	10.62%	0.96	2.41	-16.53%	3.08	76
高置信0.60/0.40	10.49%	0.99	2.57	-16.53%	3.04	73
高置信0.62/0.38	7.59%	0.83	2.51	-16.21%	2.26	57

资料来源：东海期货研究所，同花顺 iFinD

6.总结

6.1.研究结论

本研究经过因子构建、数据处理和利用机器学习建模和回测，我们可以得出大致结论。

第一，5 日单模型滚动方向正确率 57.65%，AUC 约 0.565，模型总体合格，接近良好，具备稳定预测增量；可搭配趋势/均值回归策略组合使用，适合辅助择时。5 日融合模型将方向正确率提升至约 61.32%，整体良好，接近优秀；接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，常态化周度级别资产配置和仓位调整。20 日方向滚动正确率

约为 59.37%，距良好线 60% 仅差约 0.63 个百分点，但区分度指标 AUC 达到约 0.639，达到优秀水平且显著优于 5 日任务。显示中周期方向更可分，接近主流头部 CTA 商品机器学习模型水平，可独立构建择时模块，可作为月度资产配置实盘仓位调整方案。

第二，回测层面，在全样本 $\text{thr}=0.5$ 口径下，5 日买入持有净值约 1.47，年化 6.48%，夏普比率为 0.49，盈亏比 1.17，最大回撤 20.63%。5 日融合多空期末净值约 2.36，年化 15.04%，夏普比率为 1.01，盈亏比 1.42，均达到合格标准；但是最大回撤为 30.5%，高于 25% 的合格标准，主要由于 2022 年模型阶段性失效导致的。

5 日单模型净值为约 1.71，年化收益约 9.1%、夏普约 0.66、盈亏比约 1.24，最大回撤 35.13%，虽然获得了超额收益，但是夏普和回撤控制欠佳。双均线与线性 Ridge 显著落后。5 日单模型和融合模型高置信软阈值（0.53—0.54）降低交易频率，净值和年化收益略高于 5 日单模型和融合模型全样本多空、但夏普、盈亏比和回撤控制明显改善；但随着阈值进一步收紧，交易次数下降、虽然回撤控制明显提升，但面临着年化收益、夏普、盈亏比大幅下降的问题。对 5 日任务，更合理的定位或许是：软阈值做收益、紧阈值做风控过滤，而非单一阈值同时满足高频与高置信。

非重叠口径下，20 日模型多空略优于持有；高置信条件下，年化收益率小幅提升，但夏普、盈亏比大幅改善和提升至合格和优秀水平，最大回撤明显下降。0.57 附近综合最佳，年化 11.22%，夏普约 0.99，盈亏比 2.39，期末净值 3.28，最大回撤 16.3%，改善非常明显，相关指标已提升至优秀水平。相对于 5 日，20 日高置信在“正确率提升、盈亏比、回撤”三维上更均衡，具备较好的策略投资价值。

6.2. 研究不足

本研究仍旧存在诸多不足。一是因子层面，在 2022、2025 年年末和 2026 年，模型在商品强趋势反转、政策急转弯、地缘冲击、流动性骤变阶段连续失效，关键因子缺失和因子质量欠缺，并不能被学习到。二是数据层面，关键因子可能具备，但是因子数据质量、因子数据影响持续时间（特别是事件哑变量）难以界定，导致关键因子信息丢失。三是模型训练方面，单模型仅接近良好，并未达到良好或者优秀的水平，需要提升数据质量和训练方法进一步优化模型 AUC、方向准确率。

6.3. 后续计划

针对本研究的诸多不足，后续研究主要聚焦以下方面。一是完善实时监控滚动准确率、AUC、分年失效警报与阈值漂移检验，控制风险和回测。二是聚焦单模型有效信息扩充，包括挖掘关键有效因子、提升因子数据质量。三是分体制自适应，优化模型阶段性失效风险。四是重点引入体制自适应与新信息源，继续优化提升 5 日和 20 日单模型指标至良好和优秀水平。五是该模型体系推进扩展到可交易的具体商品。

重要声明

本报告由东海期货有限责任公司研究所团队完成，报告中信息均源于公开可获得资料。东海期货力求报告内容的客观、公正，但对这些信息的准确性及完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的观点、结论和建议等全部内容只提供给客户做参考之用，并不构成对客户投资建议，也未考虑个别客户特殊的投资目标、财务状况或需要，客户不应单纯依靠本报告而取代个人的独立判断。在任何情况下，本公司不对任何人因使用本报告中的任何内容所导致的任何损失负任何责任，交易者需自行承担风险。本报告版权仅为东海期货有限责任公司研究所所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制发布，如引用、转载、刊发，须注明出处为东海期货有限责任公司。

东海期货有限责任公司研究所

地址：上海浦东新区峨山路505号东方纯一大厦10楼

联系人：贾利军

电话：021-68757181

网址：www.qh168.com.cn

E-MAIL:Jialj@qh168.com.cn