

本报告系统搭建了“专家行业经验+AI 模型”深度融合的大宗商品价格预测新范式，以海量全维度数据为根基，搭建长中短期全周期研究框架，通过多算法量化筛选与行业专家论证双轨机制提纯核心因子，并构建了涵盖计量模型、集成学习与深度学习的多层预测模型体系，并辅以情景分析应对极端场景。该范式突破传统经验判断的主观性局限，兼具科学性、有效性、解释性与风险可控性，既实现价格预判的精准度提升，又为决策提供可追溯的归因逻辑，为产业企业与投资机构在高波动市场中提升决策效率、抵御尾部风险提供了新一代量化工具。

# 机器学习与量化研究

## 期货市场研究 Machine Learning and Quant Research



本期分析研究员



**高天越**

从业资格号：

F3055799

投资咨询号：

Z0016156



**李逸资**

从业资格号：

F03105861

投资咨询号：

Z0021365



**李光庭**

从业资格号：

F03108562

投资咨询号：

Z0021506

联系人



**黄煦然**

从业资格号：

F03130959



**王博闻**

从业资格号：

F03149658

## 专家+AI：大宗商品价格预测新范式

### 研究院 量化组

#### 研究员

##### 高天越

电话：0755-23887993  
邮箱：gaotianyue@htfc.com  
从业资格号：F3055799  
投资咨询号：Z0016156

##### 李逸资

电话：0755-23887993  
邮箱：liyizi@htfc.com  
从业资格号：F03105861  
投资咨询号：Z0021365

##### 李光庭

电话：0755-23887993  
邮箱：liguangting@htfc.com  
从业资格号：F03108562  
投资咨询号：Z0021506

#### 联系人

##### 黄煦然

电话：0755-23887993  
邮箱：huangxuran@htfc.com  
从业资格号：F03130959

##### 王博闻

电话：0755-23887993  
邮箱：wangbowen@htfc.com  
从业资格号：F03149658

投资咨询业务资格：  
证监许可【2011】1289号

### 摘要

2025年以来，大宗商品市场波动显著加剧，价格中枢虽未形成趋势性上移，但波动率明显抬升，价格走势跳跃性增强，收益率呈现明显左偏与肥尾结构。供需两端扰动、地缘政治冲突、汇率波动及跨市场资金博弈等多重因素叠加，给产业企业成本管理与投资机构决策带来挑战。

在此背景下，我们构建了一套基于“数据+逻辑”的价格预测系统，整合海量宏观、产业链及高频数据，建立多维度研究框架，通过“算法+专家”双验机制筛选核心因子，并采用分层、多模型适配的预测体系，实现对价格走势的高精度、可解释预测。该系统已在华泰天玑平台上线，提供周度与月度预测，支持情景分析与动态优化，显著提升了产业与机构客户的决策效率。

在具体品种实际应用中，我们分别对铜和碳酸锂展开深入分析，建立涵盖供需、库存、价差、进出口、资金流、宏观经济、大类资产、地缘政治等多因子的研究框架，通过动态优选因子机制，实现自适应市场逻辑的变化，精准捕捉新兴驱动因子，价格预测模型在样本外数据集中表现优异，其中铜价的日频预测模型方向准确率达56.6%，盈亏比1.355；碳酸锂的周频预测模型在近两年回测期中MAPE仅1.68%，方向准确率59.68%，盈亏比1.638；月频预测模型MAPE为7.12%，方向准确率56.52%，盈亏比达1.974，验证了系统在实际场景中的有效性与稳健性。

### 核心观点

本报告系统搭建了“专家行业经验+AI模型”深度融合的大宗商品价格预测新范式，以海量全维度数据为根基，搭建长中短期全周期研究框架，通过多算法量化筛选与行业专家论证双轨机制提纯核心因子，并构建了涵盖计量模型、集成学习与深度学习的多层预测模型体系，并辅以情景分析应对极端场景。该范式突破传统经验判断的主观性局限，兼具科学性、有效性、解释性与风险可控性，既实现价格预判的精准度提升，又为决策提供可追溯的归因逻辑，为产业企业与投资机构在高波动市场中提升决策效率、抵御尾部风险提供了新一代量化工具。

## 目录

摘要 .....	2
核心观点 .....	2
一、 市场回顾 .....	6
二、 价格预测 .....	8
2.1、构建价格预测模型的价值与意义 .....	8
2.2、海量数据赋能：模型预测的原料与根基 .....	8
2.3、框架搭建：全周期商品价格预测研究体系 .....	10
2.4、因子筛选：多维度算法与专家双验的筛选体系 .....	11
2.5、预测模型构建：模型选择与“数据+逻辑”双重检验 .....	14
2.6、价格预测：周度高频响应与月度周期把控 .....	16
2.7、情景分析：专家维度与自定义维度的路径推演 .....	18
三、有色金属应用——铜价预测 .....	21
3.1、品种概况 .....	21
3.2、研究框架 .....	24
3.3、数据收集与处理 .....	27
3.4、数据降维 .....	29
3.5、预测效果 .....	34
四、新能源应用——碳酸锂价格预测 .....	37
4.1、品种概况 .....	37
4.2、基本面研究框架 .....	39
4.3、数据收集与处理 .....	42
4.4、数据降维 .....	44
4.5、预测效果 .....	48
总结 .....	51

## 图表

图 1: 主要大宗商品指数价格走势   单位: 无	6
图 2: 主要大宗商品指数 60 日波动率水平   单位: %	6
图 3: 主要大宗商品指数日收益率偏度   单位: 无	6
图 4: 主要大宗商品指数日收益率峰度   单位: 无	6
图 5: 中国:铁矿石价格指数:粉矿(巴西)   单位: 美元/吨	7
图 6: OPEC:价格:一揽子原油   单位: 美元/桶	7
图 7: 因素图谱 (以螺纹钢为例)   单位: 无	9
图 8: 因素图谱 (以螺纹钢为例)   单位: 无	11
图 9: 指标情况 (以螺纹钢为例)   单位: 无	12
图 10: 因子筛选算法逻辑 (以玻璃为例)   单位: 无	12
图 11: 关键因子展示 (以玻璃为例)   单位: 无	14
图 12: 模型评估模块 (以汽油为例)   单位: 无	16
图 13: 模型解析模块 (以汽油为例)   单位: 无	16
图 14: 模型预测模块 (以汽油为例)   单位: 无	17
图 15: 预测验证模块 (以汽油为例)   单位: 无	18
图 16: 专家维度的情景分析模块 (以铁矿石为例)   单位: 无	19
图 17: 专家维度的情景预测模块 (以铁矿石为例)   单位: 无	19
图 18: 自定义维度的情景分析模块 (以铁矿石为例)   单位: 无	19
图 19: 有色金属外盘市场回顾   单位: %	21
图 20: 2025 年 LME 基本金属成交量占比   单位: %	22
图 21: 2025 年 LME 基本金属持仓额占比   单位: %	22
图 22: 2025 年全球三大铜价累计涨跌幅   单位: %	23
图 23: 铜的商品属性: 以供需平衡表举例   单位: 千吨	23
图 24: 2025 年全球三大铜价累计涨跌幅   单位: %	24
图 25: 铜产业链示意   单位: 无	25
图 26: 铜的研究框架   单位: 无	27
图 27: RFE 算法确定最优因子数量   单位: 无	30
图 28: 基于最优因子组合的模型预测效果   单位: 元	34
图 29: 全球不同类型锂资源占比   单位: %	37
图 30: 中国不同类型锂资源占比   单位: %	37
图 31: 全球锂储量   单位: 万吨	38

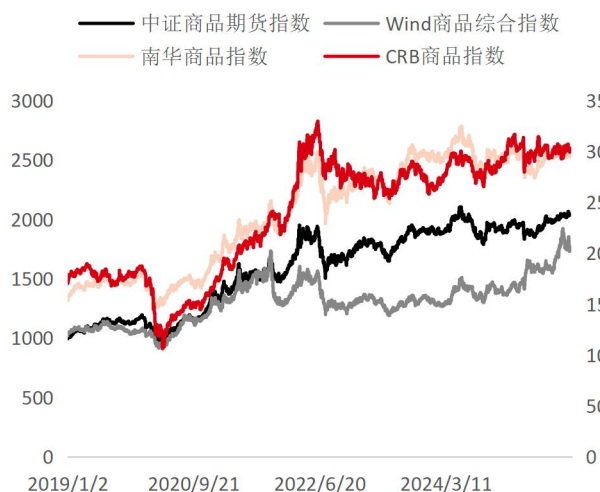
图 32: 全球锂储量分布   单位: % .....	38
图 33: 中国碳酸锂产能分布情况   单位: % .....	38
图 34: 中国碳酸锂产能产量   单位: 吨 .....	38
图 35: 碳酸锂产业链示意   单位: 无 .....	40
图 36: 碳酸锂研究框架   单位: 无 .....	42
图 37: 碳酸锂研究框架   单位: 无 .....	45
图 38: 基于最优因子组合的周频预测效果   单位: 元   % .....	49
图 38: 基于最优因子组合的月预测效果   单位: 元   % .....	50
表 1: 价格预测模型框架 .....	15
表 2: 铜价相关指标梳理   单位: 无 .....	27
表 3: 样本内铜的最优因子集   单位: 无 .....	31
表 4: 不同时期最优因子动态变化   单位: 无 .....	32
表 5: 不同因子组合模型预测效果对比   单位: 无 .....	35
表 6: 碳酸锂相关指标梳理   单位: 无 .....	42
表 7: 样本内碳酸锂的最优因子集 (周频预测)   单位: 无 .....	45
表 8: 样本内碳酸锂的最优因子集 (月频预测)   单位: 无 .....	46
表 9: 不同时期最优因子动态变化   单位: 无 .....	48
表 10: 不同因子组合模型周频预测效果对比   单位: 无 .....	49
表 11: 不同因子组合模型月频预测效果对比   单位: 无 .....	50



## 一、市场回顾

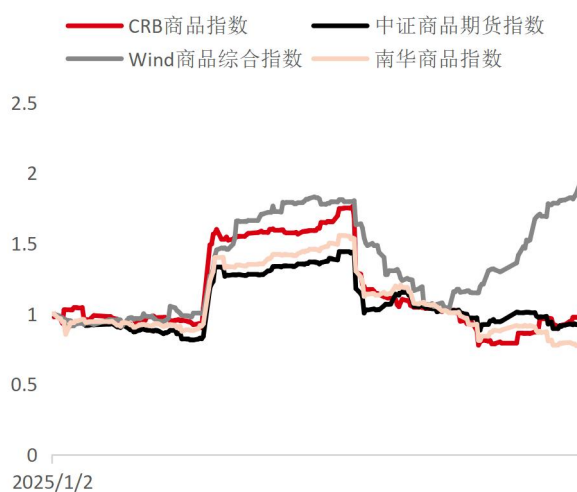
2025 年以来，主要大宗商品指数的价格中枢虽未出现趋势性上移，但波动幅度显著放大。无论从价格走势的跳跃性，还是 60 日滚动波动率的快速抬升来看，市场均呈现出明显高频扰动特征。

图 1：主要大宗商品指数价格走势 | 单位：无



数据来源：朝阳永续, Wind, 华泰期货研究院

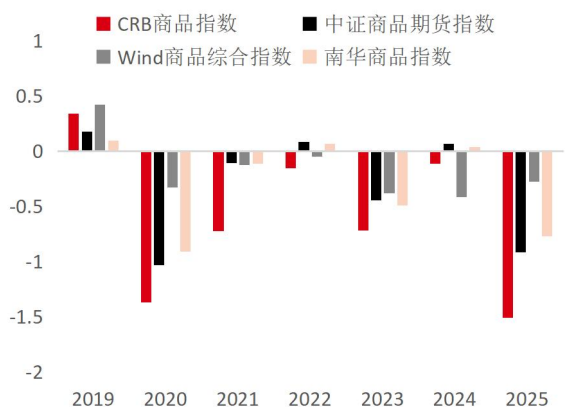
图 2：主要大宗商品指数 60 日波动率水平 | 单位：%



数据来源：朝阳永续, 华泰期货研究院

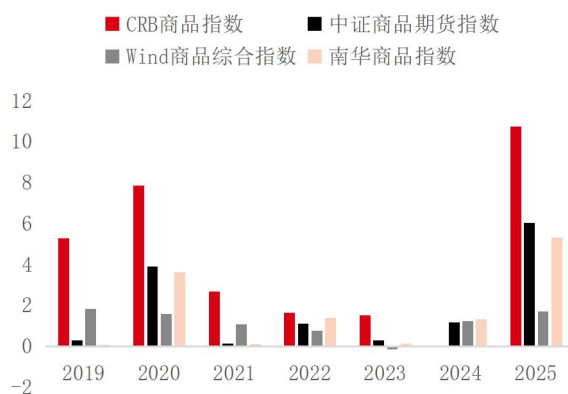
从收益率分布特征来看，2025 年主要大宗商品指数的偏度与峰度同步恶化，价格波动呈现明显的左偏与肥尾结构。

图 3：主要大宗商品指数日收益率偏度 | 单位：无



数据来源：朝阳永续, Wind, 华泰期货研究院

图 4：主要大宗商品指数日收益率峰度 | 单位：无



数据来源：朝阳永续, Wind, 华泰期货研究院

2025 年以来，全球大宗商品市场波动显著放大。国际油价在 70 - 90 美元/桶区间反复拉锯，短周期震荡幅度明显提升。铁矿石受海外供给扰动与预期情绪交替影响，单月涨幅一度突破 10%。铜、铝等工业金属则在新能源链需求推动下持续攀升，阶段性创新高。作为工业体系的关键原料，大宗商品价格的快速变动正沿产业链向下传递，对企业的成本管理、利润可得性与经营策略带来多重压力。

图 5：中国:铁矿石价格指数:粉矿(巴西) | 单位：美元/吨



数据来源：朝阳永续, Wind, 华泰期货研究院

图 6：OPEC:价格:一揽子原油 | 单位：美元/桶



数据来源：朝阳永续, Wind, 华泰期货研究院

推动大宗商品进入高波动区间的力量来自供需两端叠加。供给侧方面，地缘摩擦持续推升能源与矿产资源的波动弹性，中东运输风险抬高原油风险溢价，智利铜矿与澳洲铁矿在环保与监管压力下释放节奏低于预期，使全球供给稳定性下降，中国新能源产业扩张带动锂、钴、镍等金属需求大幅增长，而欧美制造业 PMI 回落又压制传统金属消费，形成“结构性强弱并存”的需求格局。此外，汇率波动与跨市场资金博弈进一步强化金融属性，价格的短期弹性被资金力量显著放大，市场对消息与情绪的敏感度也同步提高。在这种环境下，仅依靠静态供需判断已难以有效把握价格区间，预测模型对短期方向和波动节奏的捕捉能力变得尤为关键。

因此，在市场结构由趋势驱动向波动主导转变的阶段，价格预测的重要性明显上升。无论是产业企业的库存策略，还是投资机构的风险对冲和敞口管理，都需要依赖更及时、更量化的预测工具，以在高波动周期中提升决策效率并降低尾部暴露。

## 二、价格预测

### 2.1、构建价格预测模型的价值与意义

在商品期货交易中，价格判断是贯穿全程的核心逻辑。无论是趋势跟踪、套利交易还是风险对冲，其决策本质都围绕“价格未来走势”展开。将传统的经验判断与我们在商品因子上的研究进行有效结合，通过数学模型的加成，能够突破单一视角的局限，实现对价格走势更精准、更稳健的预判。

从各方面来看，构建价格预测模型都具有重要的意义。从业务价值来看，模型首先作为辅助决策的工具，能够把我们科学预测的能力进行强化。在传统的经验判断中，个人的跟踪、分析、整合能力在海量指标中始终具有局限性，模型的作用则是用一种更加高效、有逻辑的方式，将各种因素进行结合并提供可量化的决策依据。其次，模型实现了将整个决策流程复用化，并可根据实际情况进行及时调整的目的。将数据采集、预测、复盘的流程打通后，实现事前研判趋势方向，事中动态优化策略，事后复盘追溯成因的决策体系，在某些指标出现异常时，能提前预判可能给价格带来的波动。从管理价值来看，模型推动了管理能力的升级。将零散的决策逻辑整合成标准化模型，将经营理念从“拍脑袋”向“数据驱动”转型，为向“数智化转型”打下基础。

总体来说，价格预测模型的构建有四个不可替代的优势：科学性，模型的科学评估和决策能力可以避免传统经验论里过于主观的判断；有效性，通过海量数据与模型的筛选与整合，挖掘市场规律，提升价格预测的精准度；解释性，每个预测结论都有相应的数据支撑，支持业绩归因；控风险，通过情景分析预判未来可能的风险，提前做好应对。

从落地执行角度来看，价格预测模型的可行性具备明确预期。其一，历史规律存在延续性，大宗商品价格波动并非毫无章法，诸如季节性供需变动引发的价格周期，可通过历史数据有效捕捉；其二，行业需求逻辑具有共识性，上游生产企业与下游采购商等市场主体，均认可“供需主导价格”的核心准则，模型能精准契合行业普遍认知；其三，市场参与者行为具备合理性，各类主体的决策不会陷入无逻辑极端，例如企业不会因短期情绪盲目扩产或减产；其四，市场结构具备稳定性，产业链“供给-需求-库存”的核心逻辑、宏观经济对商品的影响机制不会突然断裂，为模型预测提供了坚实可靠的基础。

### 2.2、海量数据赋能：模型预测的原料与根基

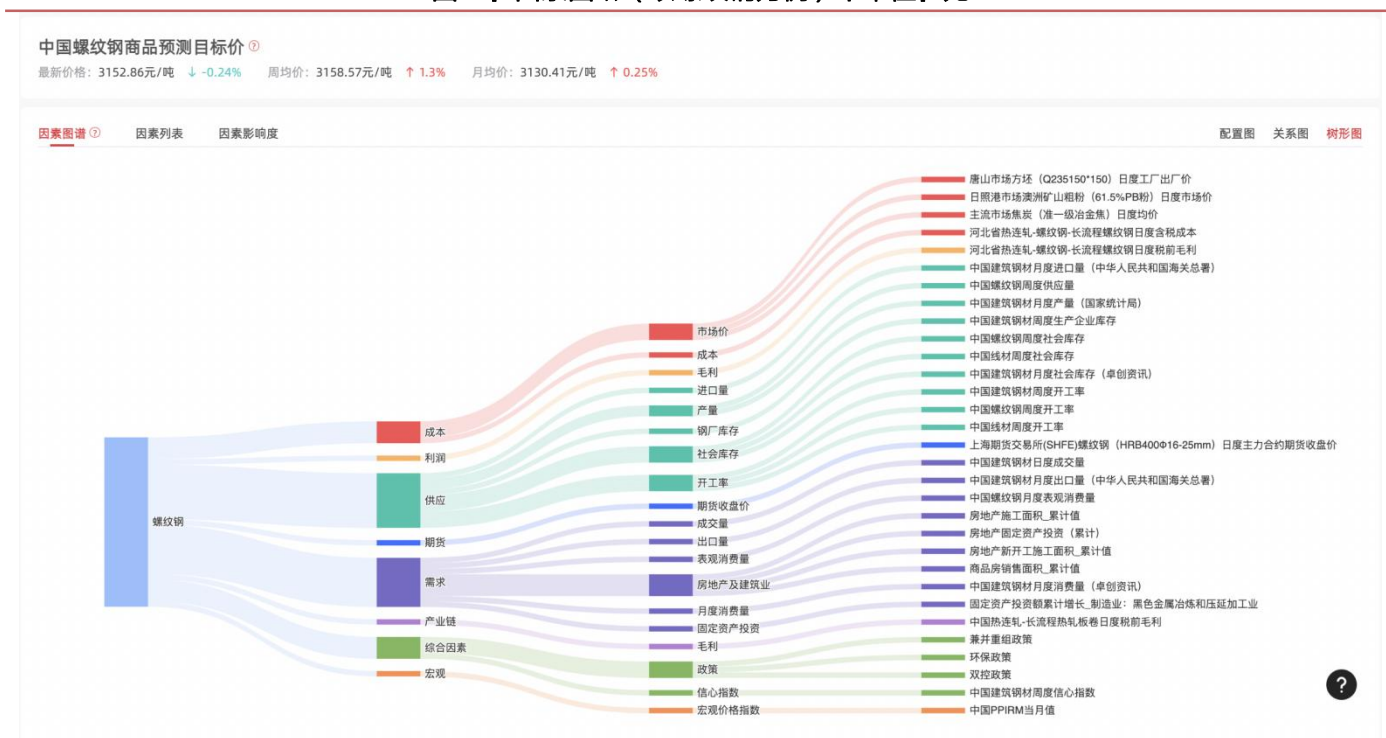
由此可见，构建价格预测模型的必要性和可行性都得到了论证，接下来最重



要的便是如何构建价格预测模型。模型的探索分为四个主要的环节：数据基础、研究体系、算法挖掘、模型优化。

想要训练出一个准确率高的预测模型，数据作为原料与根基，它的质与量都直接影响到最终输出的结果。我们整合构建了全面、专业的数据池，包括七大常见的数据终端、超过 200 万的指标，并已经在华泰天玑系统上线。数据池对全方位、各维度的数据都实现了覆盖，从宏观经济（如 GDP、制造业指数、进出口数据），到金融数据（如期货行情、基差、资金流向），再到产业链数据（如供需格局、库存水平、生产利润、开工率等）。拥有全链条数据集的好处就是，当我们开始研究具体的某一品种时，可以自上而下地分析该品种所处的宏观环境、行业周期与产业链供需矛盾，无需再为维度单一而受限；同时能通过多源数据的交叉验证与关联分析，精准挖掘驱动价格波动的核心因子，为模型提供更全面、扎实的训练基础，从源头提升预测的精准度与稳健性。

图 7：因素图谱（以螺纹钢为例）| 单位：无



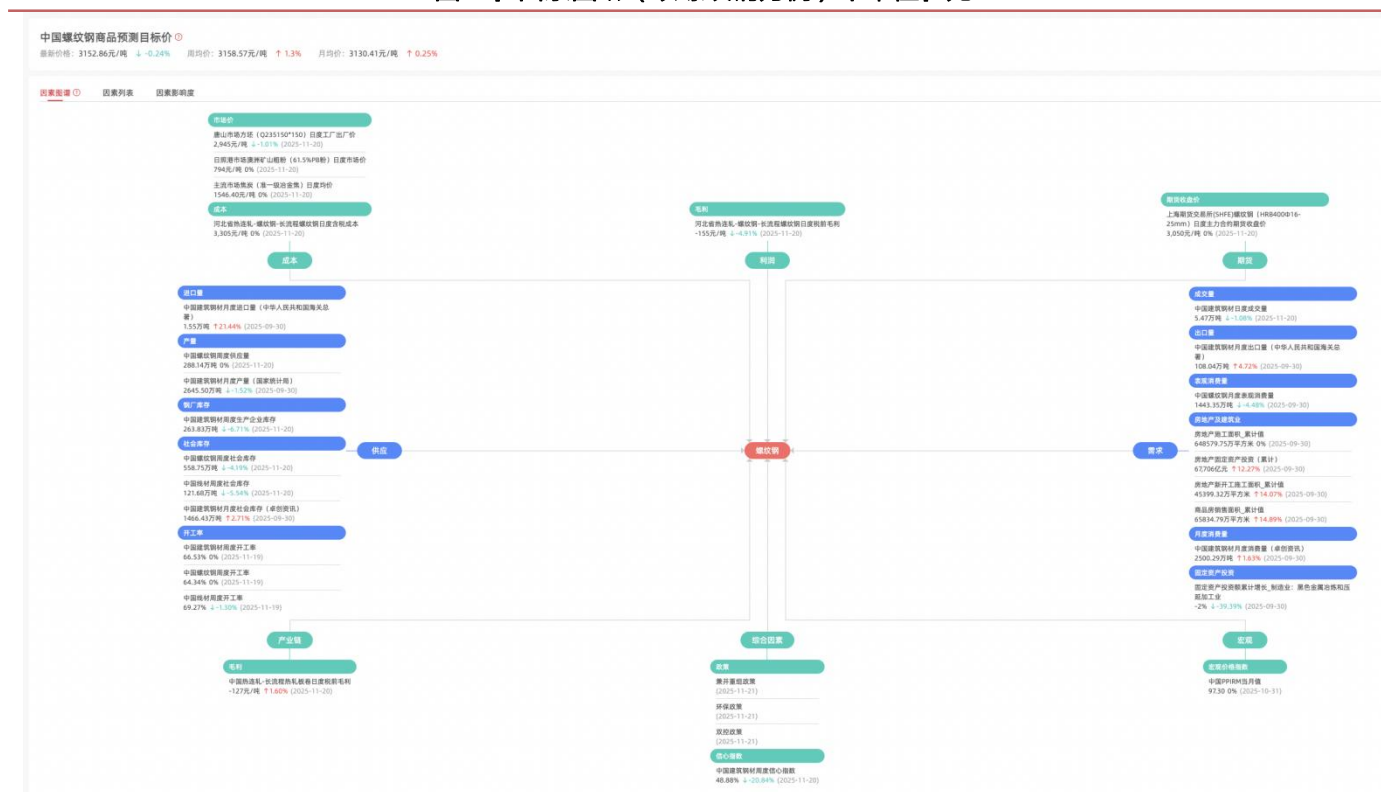
数据来源：卓创资讯，华泰期货研究院

举例，当我们在分析螺纹钢时，我们可以先将其逻辑关系梳理出来，如逻辑链“供给-需求-库存-利润”，而每个环节都可以找到相对应的数据量化该因素的影响，如中国建筑钢材月度进口量、重要省份螺纹钢含税成本价、房地产新开工施工面积等等，这些数据为后续建模都提供了扎实的原材料。

## 2.3、框架搭建：全周期商品价格预测研究体系

在准备好了充足的数据原料之后，研究体系的确定也是模型搭建的关键一步，而分析影响因素、确定逻辑框架、梳理关键指标，是研究体系逻辑框架梳理的三个重要步骤。海量的数据需要清晰的归类与框架才能更好地服务于模型。在商品市场中，“均衡价格”依然是价格变动的核心逻辑，是市场对“供给-需求-库存”等因素博弈后的共识价格。在此基础上，我们将影响价格的因素分为长期、中期、和短期对价格驱动的逻辑进行梳理。长期层面来看，以增长理论、经济转型为框架，聚焦政策、科技树、终端消费及产业链价值链分析，把握商品价格的长期趋势性逻辑。中期层面来说，主要围绕供需均衡与博弈展开。短期层面，则聚焦消息面带来的短期波动与冲击，如关税政策、地缘政治冲突等突发因素，捕捉价格的短期偏离与修复的机会。三层体系相互支撑，从长期趋势到中期均衡再到短期波动，形成了一套覆盖全周期、多维度的商品价格预测研究逻辑。我们在天玑上将研究体系的逻辑与具体指标情况进行了可视化，让整个分析框架更加清晰易读。

图 8: 因素图谱 (以螺纹钢为例) | 单位: 无

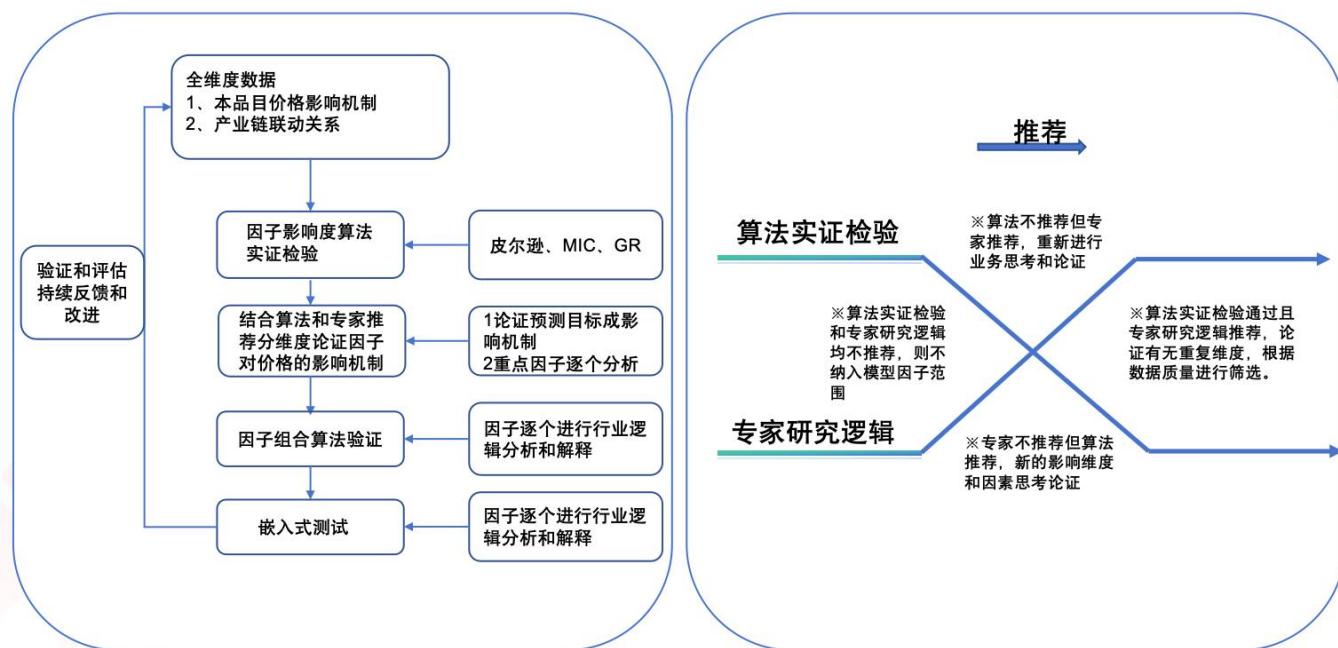


数据来源：卓创资讯，华泰期货研究院

## 2.4、因子筛选：多维度算法与专家双验的筛选体系

在完成各维度数据的整理归类后，因子筛选是后续建模中至关重要的核心环节。模型的预测效果并非取决于因子数量的堆砌，而关键在于筛选出 “影响力强、行业有共识、可量化落地” 的核心因子，为此，我们建立了一套严格的因子筛选标准与流程。

图 9：指标情况（以螺纹钢为例）| 单位：无



数据来源：卓创资讯，华泰期货研究院

首要遵循四大核心原则：一是契合经济学底层逻辑（如供需关系主导价格的核心规律）；二是与商品价格具备强相关性，能实质驱动价格波动；三是获得行业普遍认可，属于市场重点关注的关键指标；四是具备 5 年以上的连续数据积累，满足长期跟踪与模型训练的需求。

其次，我们构建了“建模前多维度筛选+建模后消融测试验证”的因子选择算法逻辑，确保筛选出的因子既具备理论支撑又有实战价值。建模前，我们用了三种算法进行多维度量化筛选：

**相关性与跨期相关性：**通过皮尔逊相关系数（PCC）衡量因子与预测标的价格的线性关联，同时利用跨期相关系数（CCF）分析因子的滞后特征对预测标的价格时间序列的影响，观测其方向与显著性，捕捉因子在不同时间维度的关联价值。

**信息增益：**弥补相关系数仅能衡量线性关系的局限，通过信息增益量化因子与标的价格的非线性关联，信息增益越大，说明因子对价格的预测效力越强，是常见的机器学习特征选择方法。

**最大信息系数：**用于衡量两个变量之间的互信息，最大信息系数越大，因子的预测效力越强，为前沿的机器学习特征选择方法。

图 10：因子筛选算法逻辑（以玻璃为例）| 单位：无

中国浮法玻璃商品预测目标价 ②				
最新价格: 1155.36元/吨 ↓ -0.41%    周均价: 1162.39元/吨 ↓ -2.42%    月均价: 1186.47元/吨 ↓ -5.19%				
因素图谱 ③	因素列表	因素影响度		
测算频度: 周度 月度				因子数量: 11
不滞后				
因素	频度	① 皮尔逊相关系数	② 信息增益比例	③ 最大信息系数
湖北省浮法玻璃周度企业库存 KF	周度	-0.12	0.14	0.21
中国浮法玻璃浮法工艺日度含税成本(动力煤) KF	日度	0.60	0.34	0.60
中国浮法玻璃浮法工艺日度税前毛利(动力煤) KF	日度	0.76	0.32	0.55
沙河市浮法玻璃周度社会库存(卓创资讯)	周度	-0.15	0.27	0.27
中国浮法玻璃浮法工艺日度税前毛利(管道气) KF	日度	0.85	0.35	0.59
中国浮法玻璃周度企业库存	周度	-0.57	0.31	0.51
中国纯碱(重质纯碱)日度均价(银行承兑汇票180天不贴息支付) KF	日度	0.62	0.30	0.70
中国浮法玻璃周度库存天数 KF	周度	-0.44	0.33	0.47
中国浮法玻璃周度产销率	周度	-0.08	0	0.21
中国浮法玻璃浮法工艺日度含税成本(管道气) KF	日度	0.53	0.31	0.61
郑州商品交易所(ZCE)浮法玻璃日度期货收盘价 KF	日度	0.93	0.48	0.72

数据来源：卓创资讯，华泰期货研究院

建模后，我们则进行嵌入式消融测试：基于预训练模型的特征化因子筛选，将过筛后的因子嵌入已建立的模型中，在模型内评估不同因子组合的预测效果。通过因子的逐次剔除，递归式地剔除像噪比过低、预测能力弱的部分因子，进一步验证并优化因子组合的有效性。

为确保因子的有效性与实用性，最终入选模型的因子还要通过“算法+专家”的双论证：算法推荐且经行业专家认可的因子，直接纳入模型；若两者均不认可，则予以排除；对于算法推荐但专家对业务逻辑存疑的因子，需重新核查其经济合理性；若专家认可但算法未选中，需深挖是否存在数据维度遗漏或算法适配性问题。以玻璃分析为例，我们从 30 个潜在因子中，通过该机制最终筛选出 15 个左右的关键因子，为模型精准度筑牢了因子层的保障，并在天玑上进行标识和展示。



图 11：关键因子展示（以玻璃为例）| 单位：无

中国浮法玻璃商品预测目标价 ②

最新价格：1155.36元/吨 ↓ -0.41% 周均价：1162.39元/吨 ↓ -2.42% 月均价：1186.47元/吨 ↓ -5.19%

因素图谱 ①	因素列表	因素影响度	因子数量：11	
测算频率：	周度	月度	不滞后	
因素	频度	① 皮尔逊相关系数	② 信息增益比例	③ 最大信息系数
湖北省浮法玻璃周度企业库存  KF	周度	-0.12	0.14	0.21
中国浮法玻璃浮法工艺日度含税成本（动力煤）  KF	日度	0.60	0.34	0.60
中国浮法玻璃浮法工艺日度税前毛利（动力煤）  KF	日度	0.76	0.32	0.55
沙河市浮法玻璃周度社会库存(卓创资讯)	周度	-0.15	0.27	0.27
中国浮法玻璃浮法工艺日度税前毛利（管道气）  KF	日度	0.85	0.35	0.59
中国浮法玻璃周度企业库存	周度	-0.57	0.31	0.51
中国纯碱(重质纯碱)日度均价(银行承兑汇票180天不贴息支付)  KF	日度	0.62	0.30	0.70
中国浮法玻璃周度库存天数  KF	周度	-0.44	0.33	0.47
中国浮法玻璃周度产销率	周度	-0.08	0	0.21
中国浮法玻璃浮法工艺日度含税成本（管道气）  KF	日度	0.53	0.31	0.61
郑州商品交易所(ZCE)浮法玻璃日度期货收盘价  KF	日度	0.93	0.48	0.72

数据来源：卓创资讯，华泰期货研究院

2.5、预测模型构建：模型选择与“数据+逻辑”双重检验

在完成了科学严谨的因子筛选与论证后，我们进入了价格预测体系的核心环节——模型构建。核心任务是：将经过验证的有效因子，通过一套适配的、稳健的算法体系，转化为对未来价格点位与趋势的预测。我们的模型并非追求单一的、“终极”的复杂模型，而是强调构建一个分层、互补、且具有持续自我校验能力的模型体系。我们的模型体系不依赖于单一算法，而是根据每个品种，在完成因子筛选与论证的步骤后，所提供因子数据的实际质量情况，如数据规模、特征复杂度和预测周期，选择合适的模型。而在针对不同品种可能提供的因子数据，我们最终选定了以下主流模型进入我们的模型体系。

表 1：价格预测模型框架

模型框架	适用数据规模	适用数据类型	优势	典型场景
ArimaX	中小规模	单变量/多变量时序	简单、可解释性强，支持外生变量	引入外部因素的评估
BVAR	中小规模	多变量时序	处理变量间依赖关系，适合小数据	多因素驱动的评估
Random Forest	中小规模	结构化数据	抗过拟合、可解释性强，能评估特征重要性	高鲁棒性的回归评估
XGBoost	中小规模	结构化数据	高精度、可解释性强，支持非线性关系	基于特征的评估
LSTM	大规模	时序数据	捕捉长期依赖关系，适合复杂时间序列	长时间跨度的评估
Transformer	大规模	时序/多模态数据	捕捉全局依赖关系，适合超长序列	多维度、复杂依赖的评估

数据来源：卓创资讯，华泰期货研究院

上述的五个主流模型可以被划分为以下三种类型：

**经典计量经济模型：**如向量自回归模型，擅长捕捉多变量时间序列之间的动态交互关系，尤其适用于分析宏观经济变量与价格的联动效应。

**树模型与集成学习：**以梯度提升机为代表（如 XGBoost），其核心优势在于能自动处理因子间的复杂非线性关系和交互效应，无需严格的统计假设，对现实世界中复杂数据结构的拟合能力极强。

**深度学习模型：**在数据量足够大、序列依赖性强的场景下，我们会尝试长短记忆网络等循环神经网络，以捕捉价格序列中更深层次的时序依赖模式。

而在选定模型并初步构建后也并非一劳永逸，我们坚持”逻辑+数据“的双驱动运维模式。在因子筛选与论证的步骤中，我们实际上是找出了逻辑稳健和可解释性强的因子作为模型的数据支撑，而模型的作用则是在基本面逻辑框架内，优化每个因子应该被赋予的权重，通过历史数据找到因子与收益的最优映射关系，并在此关系下使用数据驱动模型。而在模型推导出该映射关系后，我们会从数据和逻辑上进行双重检验。其中数据层面上，使用验证集测试模型预测值与真实值之间的精度；而逻辑层面上，我们会通过专家反馈调整，确保结果既符合数据规律，又贴合行业实际。

## 2.6、价格预测：周度高频响应与月度周期把控

有了丰富的因子数据与扎实的模型，下一步我们将它产品化成了华泰天玑平台的“大宗商品价格预测系统”，通过 SaaS 或定制模式服务客户，核心就是让“复杂的模型”变成“人人会用的工具”。由于市场决策存在不同的时间维度和逻辑层次，单一的预测输出无法满足多元化的需求。因此，我们将预测体系具体化为周度预测与月度预测两个核心产品。这并非简单地将同一模型运行于不同时间颗粒度，而是构建了两套目标不同、因子侧重不同、模型微调策略亦不同的、既独立又协同的预测系统。其中周度预测更偏向于快速吸收最新市场信息以此更好地预测短期的价格波动，而月度预测则更偏向于趋势方向地判断和关键价位区域的预测，两者协同验证、相互补充。

下面我们以能源板块的汽油周度预测为例进行展示：

图 12：模型评估模块（以汽油为例）| 单位：无



数据来源：卓创资讯，华泰期货研究院

投资者可以通过模型评估板块，直接看到当前通过因子动态模式识别，筛选出的当前最具解释力的因子，及构成这些因子数据的最新数值和历史走势。而在模型精度界面，我们通过回溯测试的方式，带入各因子历史数据，计算模型在近4期对目标价格的预测值（理论预测值），并与实际价格进行对比。以此评估模型的设置、因子选择以及特征设置的合理性。

图 13：模型解析模块（以汽油为例）| 单位：无

## 模型解析

### 因子贡献度 ①

统计周期: 2025-11-16

数据详情

中国汽油(92#国VI)日度均价(元/吨)	32.68 %
山东市场烷基化油日度市场价	22.48 %
中国蜡油催化裂化日度含税装置成本	6.58 %
中国原油炼油综合生产日度含税装置成本	-3.29 %
汽油基差	-11.15 %
山东市场MTBE日度市场价	-23.83 %

数据来源: 卓创资讯, 华泰期货研究院

而在模型解析板块, 投资者可以进一步观察各因子对于模型预测能力的贡献度水平, 继而掌握各因子近期对未来价格走势的影响程度并有目的性地进行跟踪。

图 14: 模型预测模块 (以汽油为例) | 单位: 无

## 模型预测

### 预测结果 ①

预测周期 2025-11-22 (第47周) 至 2025-12-13 (第50周)

数据详情



数据来源: 卓创资讯, 华泰期货研究院

模型预测板块则是模型最直接的输出结果, 模型将会对未来 1 至 4 期的价格进行预测。

图 15：预测验证模块（以汽油为例）| 单位：无



数据来源：卓创资讯，华泰期货研究院

对于预测结果，我们在预测验证板块会对模型的预测价格与最后实际观察到的真实值进行比较，直观展现模型最近的精度表现，更方便投资者对模型的实际效果进行评估。

## 2.7、情景分析：专家维度与自定义维度的路径推演

标准化的预测模型，无论是周度还是月度，其核心是基于“历史规律延续”与“当前信息外推”的范式。然而，真实世界充满结构性转折点与未知——例如突发的地缘政治冲突、颠覆性的产业政策或不可预见的全球性公共卫生事件，这些情景可能会在短时间内颠覆由历史数据训练出的模型所依赖的基本假设，在这种情况下往往需要人为的介入对模型进行辅导。而除了历史数据之外，个人的主观判断也同样可以作为模型的增量信息，例如通过一线的调研，对部分关键因子未来的走势进行判断。为此我们引入了情景分析的功能，分别从专家维度和自定义维度对未来可能发生的情景进行指标的预设，进而辅导模型对未来价格做出预测，既弥补了标准化预测的静态局限性，又为投资者提供了应对式决策工具。下面我们以太铁矿石为例进行展示：



图 16：专家维度的情景分析模块（以铁矿石为例） | 单位：无



数据来源：卓创资讯，华泰期货研究院

上面是从专家维度出发，针对各关键因子后续可能发生的情景以及对应的数值进行的预测，接下来模型则会根据其余因子的历史数值和专家给出的预测因子值对未来价格做出预测。

图 17：专家维度的情景预测模块（以铁矿石为例） | 单位：无



数据来源：卓创资讯，华泰期货研究院

在给出未来几期的价格预测的绝对点位基础上，模型还会对未来的价格可能的分布概率进行预测，帮助投资者对价格的可能走势有请一个更清晰的把握。而除了专家维度外，投资者还可以根据自己的研究以及主观判断，对未来可能发生的情景以及因子走势，进行自定义数值填写，模型同样可以根据投资者自定义填写的数值做出价格预测。

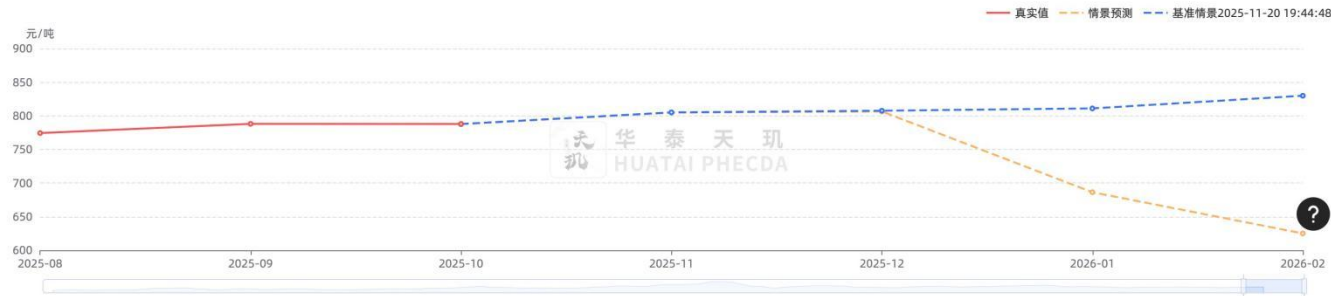
图 18：自定义维度的情景分析模块（以铁矿石为例） | 单位：无

请选择情景： 基准情景 利空情景 利多情景

重置 运算

因子 (数量: 8)	单位	2025-09	2025-10	2025-11	2025-12	2026-01	情景描述	操作
基差		-6.227273	3	-39.9498	-45.0337	-46.7902	预期较差, 现实需求较强	清空
工业增加值当月同比	%	6.5	6.3	6.4	6.5	6.6	宏观情绪转暖	清空
固定资产投资额累计增长_房地产业	%	-14	-14.2	-14.1	-13.7	-13.4	投资信心不足	清空
中国铁矿石原矿月度产量	万吨	8426.7	7628.43	8400	8200	6998.73	天气因素大幅影响局部生产	清空
中国铁矿石周度开工率	%	50.91	50.756	50	49	48	天气因素大幅影响生产	清空
中国铁矿石月度进口量	千吨	116325.77	102211	99437	98156.9	100000	进口量升高后回落	清空
唐山市场方坯(Q235150*150)日度市场价	元/吨	3005	2954.1667	2931.54	2904.26	2894.49	外生因子AI预测	清空
唐山市场热轧板卷(Q235150*150)日度市场价	元/吨	3005	2954.1667	2931.54	2904.26	2894.49	外生因子AI预测	清空

预测结果 预测周期: 2025-11至2026-02



数据来源: 卓创资讯, 华泰期货研究院

自此, 我们通过在华泰天玑系统上构建多层次、多因子的价格预测模型体系, 结合海量数据与专家经验, 实现了对大宗商品价格走势的科学预判与动态优化。未来, 我们将持续完善模型框架、拓展数据维度、深化情景分析能力, 致力于在复杂多变的市场环境中, 为投资者提供更精准、更可靠、更具操作性的决策支持。

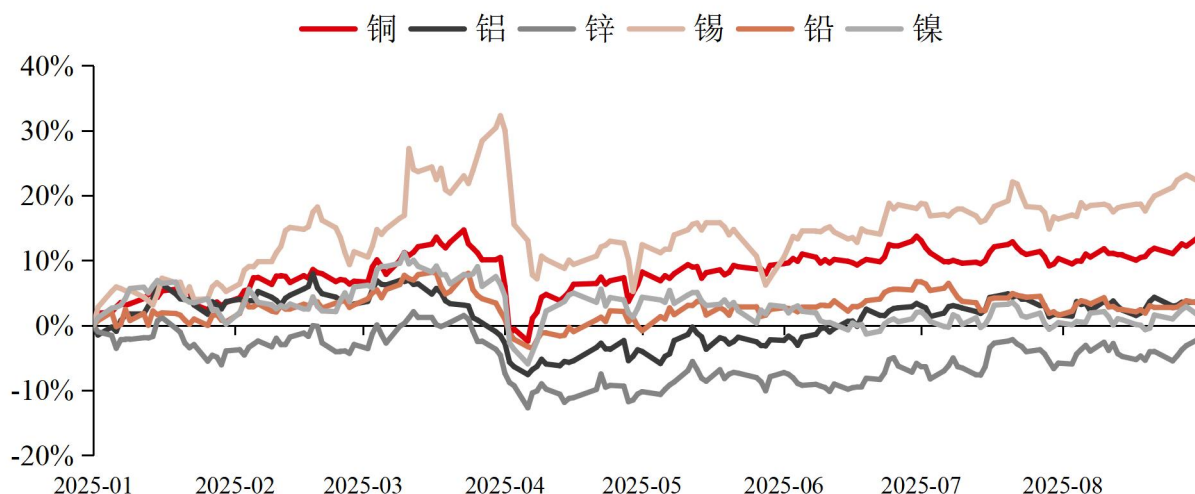
### 三、有色金属应用——铜价预测

#### 3.1、品种概况

铜作为全球工业活动的“晴雨表”，不仅是大宗商品体系中的核心品种，更是一种兼具商品属性与金融属性的重要资产。其商品属性体现在铜广泛用于电力、建筑、交通、机械制造等实体经济领域，是反映全球制造业景气程度的关键原材料；金融属性则表现为铜价易受宏观流动性、汇率变动、投机情绪及资产配置需求的影响，常被视作通胀预期和经济增长的领先指标。这种双重属性决定了影响铜价的因素极为广泛，既包括供需基本面、库存变动、产业链结构等微观变量，也涉及全球货币政策、地缘政治、美元指数、市场风险偏好等宏观因子。因此，对铜价的预测不能仅局限于传统供需分析，还需纳入多市场、多维度、多频段信息及数据，构建融合产业现实与金融逻辑的综合研判框架。

目前，全球范围内开展铜期货交易的主要交易所有：伦敦金属交易所（LME）、纽约商品交易所（COMEX）以及上海期货交易所（SHFE）。LME 作为国际有色金属定价的核心市场，其铜合约（A 级电解铜）采用美元计价，交易活跃、流动性强，被全球公认为基准价格；COMEX 铜期货主要聚焦精炼铜交易，更多反映北美地区的供需状况；SHFE 的沪铜期货则以人民币计价，实行实物交割制度，近年来影响力不断提升，尤其体现中国及亚太区域的实际供需变化。这三家交易所共同构成了全球铜期货的定价体系，彼此价格具备长期联动性的同时，也常因区域基本面、政策差异及资金结构等因素出现明显分化。

图 19：有色金属外盘市场回顾 | 单位：%

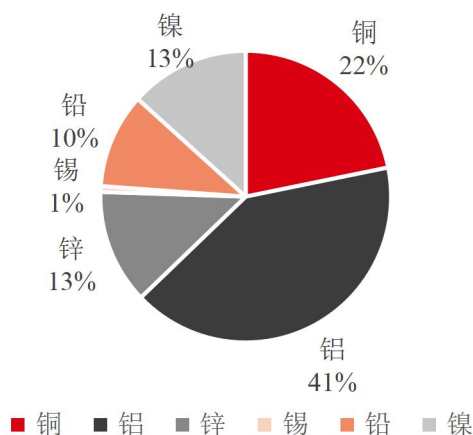


数据来源：SMM，华泰期货研究院

以最具定价影响力的 LME 市场为例，2025 年以来基本金属行情呈现显著结构性分化，其中铜与锡涨幅领先，而锌则表现承压。铜价受益于“弱美元、弱

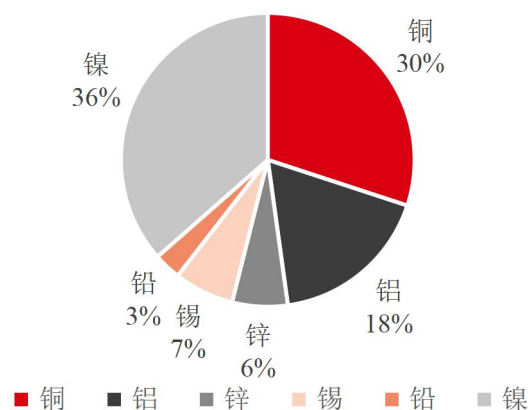
供应、强消费”三重利好推动，截至 2025 年 9 月，LME 铜价累计上涨达 14%，无论是成交规模还是持仓金额，均稳居基本金属前列。这不仅体现了铜在工业金属中的核心地位，也反映了其对宏观与产业变动的高度敏感性。

图 20：2025 年 LME 基本金属成交量占比 | 单位：%



数据来源：SMM，华泰期货研究院

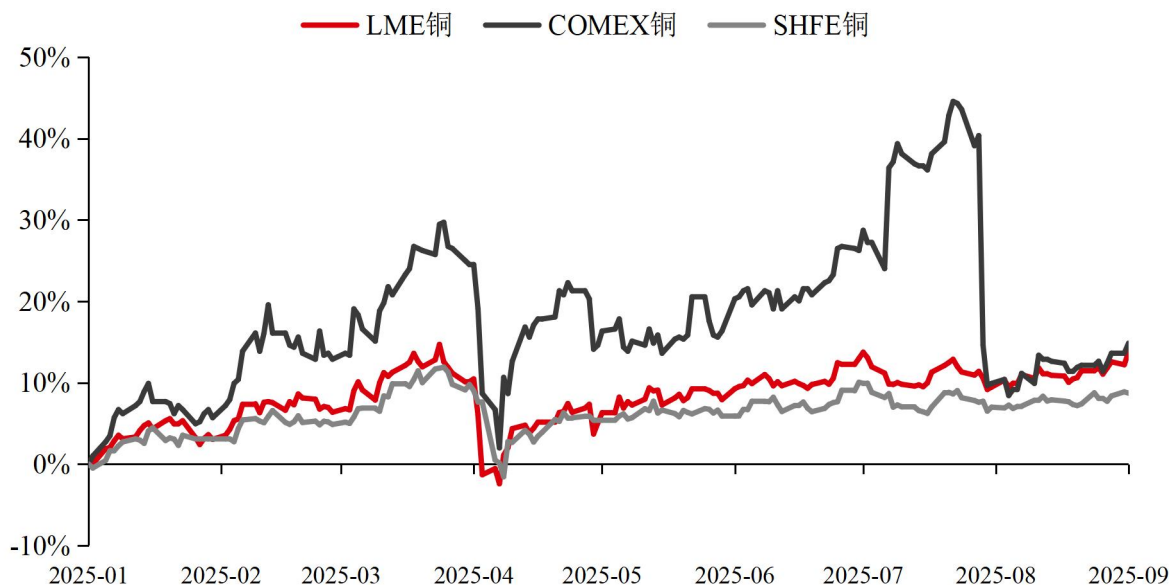
图 21：2025 年 LME 基本金属持仓额占比 | 单位：%



数据来源：SMM，华泰期货研究院

进一步观察全球三大铜期货——LME 伦铜、COMEX 美铜和 SHFE 沪铜在 2025 年以来的价格走势，可发现其整体均呈上行趋势，共性支撑逻辑明显：供应端频现扰动，南美矿山生产受阻、冶炼产能面临瓶颈、地缘因素影响物流，导致全球显性库存持续下降；需求端则受益于全球制造业 PMI 的阶段性回升，以及中国“以旧换新”政策对家电、汽车消费的拉动；叠加宏观层面流动性维持宽松、美元走势疲软，共同助推铜价上涨。

图 22：2025 年全球三大铜价累计涨跌幅 | 单位：%



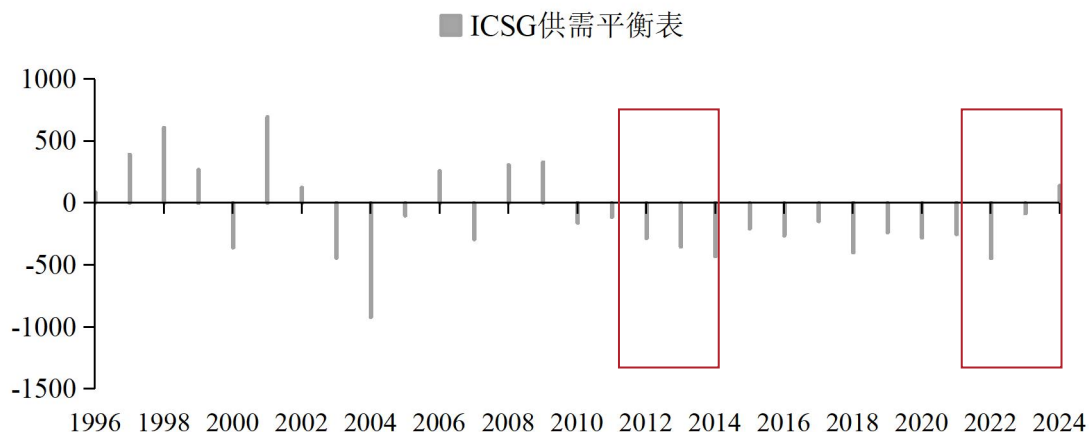
数据来源：SMM，华泰期货研究院

然而，三大铜价之间也多次出现显著分化，尤其表现为美铜涨幅领先，且与伦铜溢价因美国关税政策反复而大幅波动。今年以来美国关税政策无疑一直是铜市场核心议题。特朗普政府于 2 月底启动对铜产品的“232 调查”，随后在 7 月 8 日宣布对进口铜加征 50% 关税，导致美铜价格短期急涨超 15%，伦铜溢价一度逼近 3000 美元/吨。这一结构性分化的背后，主因市场对关税政策的预期引发跨市场套利与库存转移行为，美国进口商提前备货致实物铜集中流向北美，COMEX 库存上升，LME 与 SHFE 库存则持续偏紧，区域供需失衡加剧价格差异。而在 7 月 30 日政策调整，宣布仅对铜成品征税、暂免电解铜和铜精矿关税后，美铜价格迅速回调，与伦铜价差趋于收敛。此类政策反复不仅显著放大价格波动率，也扭曲了跨市场定价结构，对风险监测与预测的前瞻性与稳健性提出更高要求。

在建立起对全球铜市场的基本认知后，我们回归核心目标：如何实现对商品价格波动的准确预测。传统分析框架在实际应用中往往面临解释力不足的困境。例如，从铜的商品属性出发，依据经典“供求决定价格”理论，铜价本应在供不应求时价格上涨、供过于求时下跌，但事实并非如此：2012 - 2015 年间铜市处于持续短缺状态，价格却一路走低；而 2024 年在整体供应过剩、显性库存较前两年上升的背景下，价格反而走高，静态供需分析难以合理解释这些现象。

图 23：铜的商品属性：以供需平衡表举例 | 单位：千吨

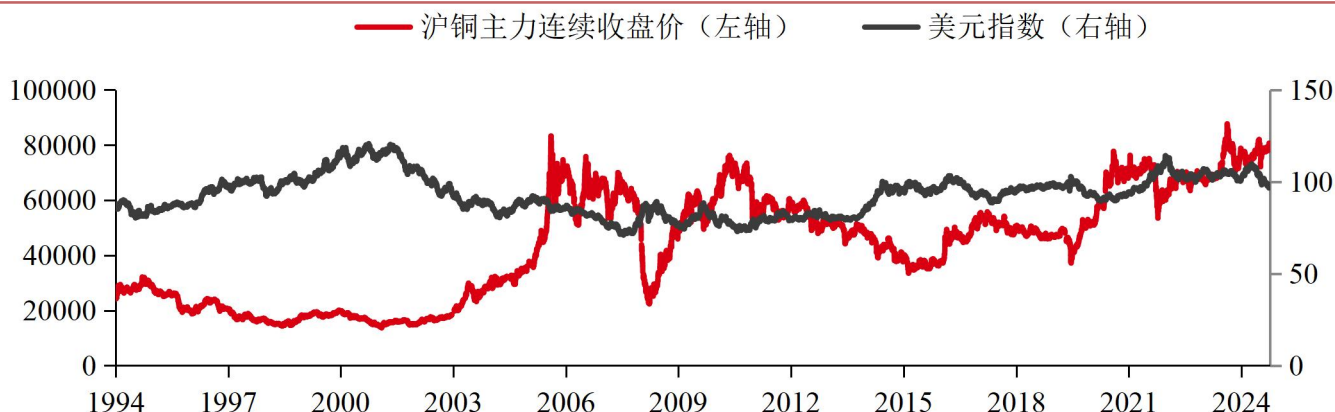




数据来源：SMM，华泰期货研究院

即便换个角度从金融属性切入，以美元指数为例，其与铜价之间的负相关性自2021年起也明显减弱。

图 24：2025 年全球三大铜价累计涨跌幅 | 单位：%



数据来源：SMM，华泰期货研究院

面对这些挑战，我们亟需解决两个关键问题：其一为归因问题，即如何动态识别不同因子对铜价的影响权重，厘清当前市场中的主导驱动逻辑；其二为预测问题，即在确定关键因子的基础上，如何构建能够融合多源信息、适应市场风格切换的智能预测模型，从而提升对价格方向判断的准确性与稳健性。

### 3.2、研究框架

铜的产业链主要由矿产开采、冶炼加工和终端应用三个环节组成，具有原料来源集中、工艺路线分明两大特点。铜精矿是生产粗铜的核心原料，全球约 80% 以上的铜精矿通过火法冶炼工艺（造锍熔炼 - 吹炼 - 火法精炼）转化为粗铜，湿法冶炼则多用于低品位氧化铜矿或难处理矿的处理；而粗铜需进一步通过电解精炼工艺提纯，最终形成纯度达 99.95% 以上的电解铜。

作为铜加工环节的唯一核心原料，电解铜随后通过拉拔、轧制、挤压等不同方式，被加工成铜杆、铜板带、铜管、铜箔等中间产品，并进一步用于制造电线电缆、印制电路板、热交换器及各类精密零部件，最终流入消费终端。电力行业占据铜消费的约 50%，建筑、电子和交通运输分别约占 15%、12%和 10%，其余应用于新能源、光伏等新兴领域。

终端使用所产生的废铜——包括生产过程中产生的新废和消费品淘汰后的旧废——经拆解和火法或湿法再生处理，可重新转化为再生铜原料，返回冶炼或加工环节循环利用。由此，整个铜产业链形成从“铜矿—铜精矿—粗铜—电解铜—铜加工材—终端应用—再生铜”的完整闭环。在这一体系中，电解铜居于承上启下的关键位置，直接连接上游冶炼和下游加工，是影响供需和价格的核心环节。

图 25：铜产业链示意 | 单位：无



数据来源：华泰期货研究院

铜的产业链贯穿了从资源开发到终端消费的全过程，形成了“矿山—冶炼—加工—应用”的完整体系。这一体系不仅涉及实体物料的流转，更与全球资本流动和宏观经济情绪紧密相连，因此构建其研究框架需同时关注实体供需与金融博弈的双重逻辑。

上游原料供应构成价格波动的起点。全球铜矿资源分布极不均衡，智利、

秘鲁、刚果（金）等少数国家占据绝大部分储量，其矿山生产状况直接牵动全球供给弹性。矿产端需持续关注一系列变量：资本开支周期影响中长期产能，矿石品位下降趋势推高边际成本，而劳资纠纷、社区冲突、政策变动乃至极端天气，则共同构成生产干扰率，频发扰动短期供应预期。这些因素最终汇聚为矿产端的紧张程度，集中体现于年度长协粗炼/精炼费用（TC/RC）的谈判结果。该费用走低通常反映铜精矿供应趋紧，被视为铜价潜在的看涨信号。

中游冶炼与加工环节是连续上下游的关键枢纽。该环节的核心在于精炼铜的产出与流通，其变量集中在中国等主要生产国的冶炼产能开工率、新增项目投资节奏及企业利润水平。此外，废铜作为重要替代原料，其与精铜价差直接影响再生铜供给弹性——价差收窄将抑制废铜使用，转而提振电解铜消费。另一关键观测维度在于跨市场贸易流。沪伦比值（SHFE/LME）直接决定精炼铜进出口盈亏，引导套利资金和实物铜的跨境流动，成为调节区域间供需平衡的重要风向标，也是研判短期价格结构的重要依据。

下游终端消费是决定铜价长期方向的根本力量。铜需求广泛分布于电力基建、建筑、电子、交通及新能源等领域，其中电力行业占比约一半，是核心需求部门。因此，中国电网投资计划、新能源输电工程建设进度、房地产竣工周期、家电出口数据以及新能源汽车产销表现等，均须纳入高频跟踪体系。海外需求则与全球制造业 PMI、欧美工业生产、住房开工数据等宏观指标密切相关。所有供需力量角逐的实物结果，最终体现为全球显性库存水平（如 LME、SHFE 库存）的变化。库存持续去化或累积，是判断市场紧平衡或过剩最直接的证据。

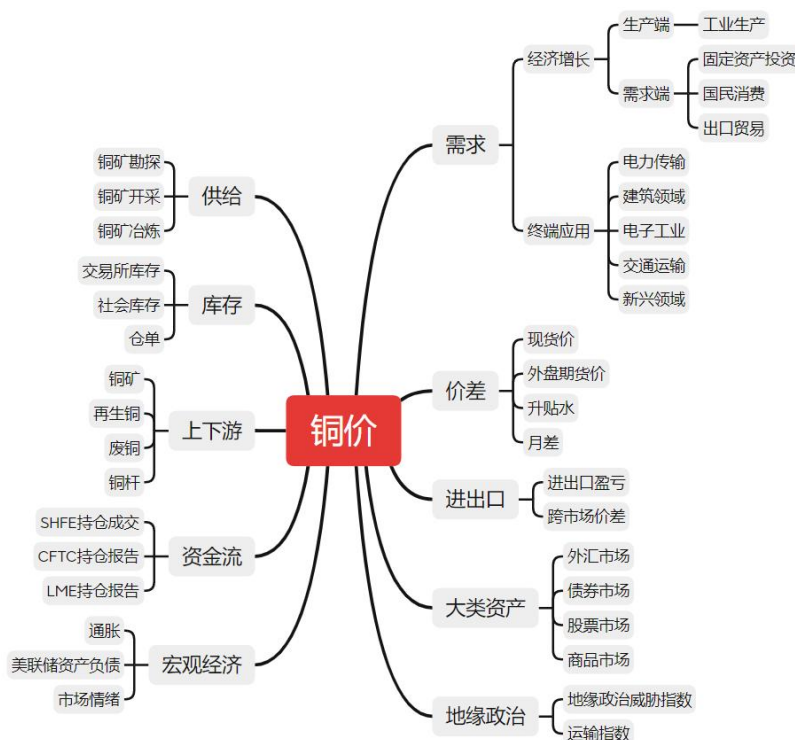
除产业链本身的微观逻辑外，铜价还在极大程度上受到宏观金融因子的深刻塑造。作为强金融属性商品，铜价对全球流动性、利率环境、汇率变动及资产偏好极为敏感。美元指数与铜价历来呈现显著负相关，美元走弱通常提振铜的实物采购与投资需求。实际利率水平则直接影响持有零息资产铜的机会成本，成为预判中长期价格压力的关键。通胀预期同样重要，铜被广泛视为对抗通胀的实物资产，盈亏平衡通胀率的抬升往往助推铜价走强。

在大类资产维度，铜与权益市场、债券及能源商品之间存在显著联动。全球主要矿业股指数可作为铜价的领先指标，因其反映了市场对行业盈利前景的预期；原油等大宗商品龙头的价格波动则会从成本推动和通胀预期两个渠道间接影响铜市。市场情绪与资金流向同样关键：VIX 恐慌指数攀升往往触发风险资产抛售，包括铜在内的大宗商品通常难以独善其身；而 CFTC 等持仓报告所揭示的基金净多头头寸变化，则清晰展现投机资金对铜市的情绪与推动力。

不可忽视的还有突发性事件因子的冲击。地缘政治冲突、国际贸易政策突变、全球公共卫生危机等，会通过改变供给预期、中断贸易流或触发避险情绪，为价格注入显著风险溢价。例如 2025 年美国对铜关税政策的反复，不仅直接引发跨市场价差剧烈波动，更导致全球库存流向重构，成为近年来诠释“政治溢价”的典型事件。

因此，构建铜的研究框架，需系统整合从矿山到终端、从库存到资金、从微观供需到宏观预期的多重变量，形成能够同时响应产业现实与金融逻辑的分析体系。

图 26：铜的研究框架 | 单位：无



数据来源：华泰期货研究院

### 3.3、数据收集与处理

在确立融合基本面、宏观环境与量价信号的综合研究框架之后，为系统捕捉铜价波动的核心影响要素，我们对潜在驱动指标进行了全景式梳理，并设定三重筛选标准以保障指标有效性与研究可靠性：（1）更新频率需满足日频或周频，确保对价格变动的及时响应；（2）数据覆盖周期需早于 2017 年，为长周期回测与因子挖掘提供充足样本支撑；（3）剔除曾出现连续中断超过一个月的指标，以维持样本的连贯性与建模的稳健性。经上述筛选步骤，最终形成包含 200 余个指标的初始清单，受篇幅所限，下表仅展示具有代表性的部分指标：

表 2：铜价相关指标梳理 | 单位：无

一级分类	二级分类	指标名称
库存	交易所库存	LME 库存、COMEX 库存、SHFE 库存
	社会库存	社会库存、保税区库存、仓库库存、总库存
	仓单	LME 注册仓单、LME 注销仓单、SHFE 仓单
	现货价	电解铜平均价、LME 现货价
价差	升贴水	电解铜升贴水平均价
	月差	LME 升贴水 0-3、当月与次月月差
	期货价	LME 铜价、COMEX 铜价
上下游	铜精矿	铜精矿价格、铜精矿入港量、铜精矿出港量、铜精矿港口库存
	冶炼厂	铜冶炼厂现货粗炼费 TC、铜冶炼厂现货精炼费 RC、火法冶炼成本
	再生铜	再生铜价格、精废价差、进口废铜盈亏平衡
	铜杆	电解铜杆加工费、制杆开工率、电解铜杆产量
进出口	进出口盈亏	现货进口盈亏
	跨市场价差	现货沪伦比值、提单洋山铜溢价、仓单洋山铜溢价
资金流	成交持仓	持仓量、成交量
	CFTC 持仓报告	各类报告头寸持仓数量、
	LME 持仓报告	各投资类型持仓数量、各投资类型交易者数量
大类资产	外汇市场	美元兑人民币汇率、美元指数
	债券市场	美国国债收益率（2 年期 10 年期）、中国国债收益率（2 年期 10 年期）
	股票市场	上证综合指数
	商品市场	LBMA 金价、LBMA 银价、LME 铝价、布伦特原油、WTI 原油、OPEC 一揽子原油、螺纹钢价格指数
宏观经济	通胀预期	美国 5 年通胀率预期、美国盈亏平衡点通胀率
	美联储资产负债表	美联储各类资产及负债当周值
	市场情绪	VIX 指数、OFR 金融压力指数、期权隐含波动率、花旗经济意外指数
地缘政治	地缘政治	地缘政治威胁指数
	海运指数	波罗的海运输指数
量价	量价因子	Curve、动量、期限结构、偏度、波动率、价值、持仓、流动性、均价突破、宏观 Beta 因子

数据来源：华泰期货研究院



在完成指标初选后，为进一步提升数据质量与模型适用性，我们对所有纳入的指标进行统一预处理，具体包括：（1）平稳性检验：对非平稳序列进行差分处理，直至通过单位根检验，确保时间序列的平稳性，规避伪回归问题；（2）异常值处理：对超出历史 1% 与 99% 分位数的极端值进行 Winsorize 缩尾处理，以降低异常观测对模型估计的干扰；（3）样本划分：将 2017 年 1 月至 2021 年 12 月设定为样本内区间，用于因子挖掘与模型训练；2022 年 1 月以后的数据作为样本外区间，用于模型预测效果的验证与优化。

### 3.4、数据降维

尽管初始指标库涵盖了铜价波动相关的多维度信息，但 200 余个指标的高维特征也带来了多重研究挑战，使得数据降维成为模型构建前的必要环节。首先，指标间普遍存在多重共线性问题，冗余信息会放大模型计算负荷，降低分析效率。其次，高维数据易引发“维度灾难”，导致模型过度拟合样本内数据，削弱对样本外新数据的预测稳健性，尤其难以应对极端事件或政策突变引发的价格结构性转折。此外，过多指标会掩盖核心驱动逻辑，使因子的经济意义模糊化，不利于后续风险传导机制的解读与策略落地。因此，通过科学的降维方法提炼关键信息、剔除冗余变量，既能简化模型结构、提升运算效率，也能增强模型的泛化能力与经济可解释性，为精准识别铜价波动的核心驱动因子提供支撑。

#### （1）组内选优：基于线性评估模型的初步筛选

为在保留核心信息的同时有效降低数据维度，我们首先采用“组内选优”策略。具体而言，我们依据经济学含义将全部指标划分为一级、二级和三级分类。在同一个三级分类下，各指标间不仅经济含义相近，且通常存在较高的统计相关性。为避免信息重叠和多重共线性，我们在每个三级分类内部，采用多种线性模型评估各指标与铜价之间的线性关联强度，仅保留得分最高的一个指标，所使用的线性评估模型包括：

**相关性分析：**计算每个指标与铜价之间的 Pearson 相关系数，直接衡量指标与铜价的线性相关方向（正相关 / 负相关）与强度（系数绝对值介于 0-1 之间），系数绝对值越接近 1，线性关联越显著。

**线性回归：**对数据进行归一化处理，消除量纲影响后，建立一元线性回归模型，以铜价为因变量，单个指标为自变量。通过回归系数的大小衡量线性相关强度。

支持向量机：采用线性核的支持向量机模型，通过模型拟合优度（ $R^2$ ）评估指标对于铜价涨跌的解释力， $R^2$  越高则指标对铜价波动的解释比例越高。

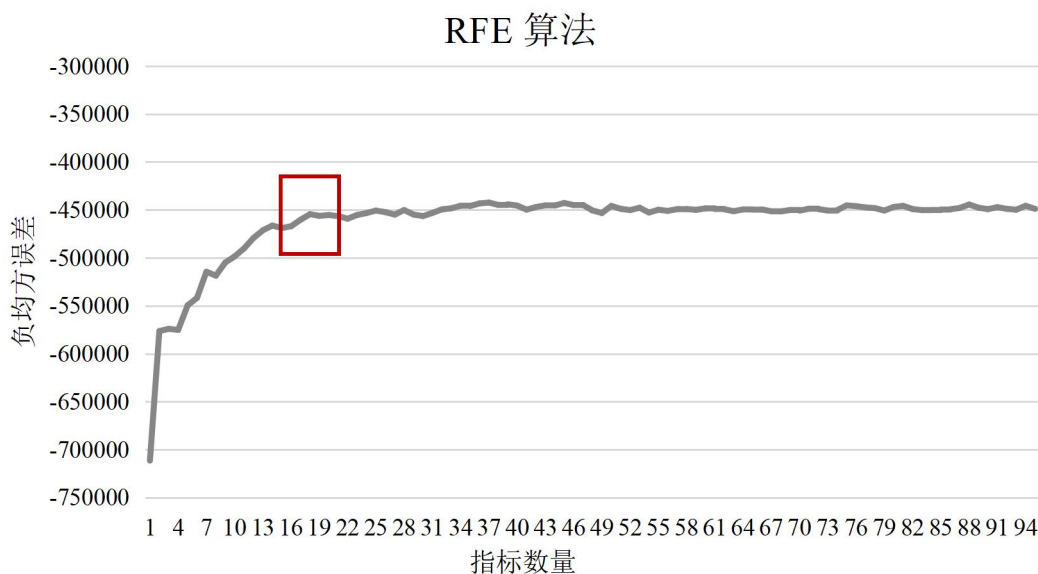
综合上述三种模型的评估结果（如相关系数绝对值、回归系数绝对值、SVM 的  $R^2$  等）进行加权打分，每组仅保留综合得分最高的指标，最终将初始 200 余个指标精简至 102 个，既剔除了冗余信息，又保留了各细分逻辑下的核心线性关联指标。

## （2）最优因子数量确定：递归特征消除的应用

初步筛选后的 102 个指标仍可能包含冗余信息。如何确定一个既能充分捕捉铜价波动信息，又能避免过度拟合的“最优”因子数量，是接下来的关键问题。我们采用递归特征消除（Recursive Feature Elimination, RFE）算法来解决这一问题。

RFE 是针对高维数据的特征选择算法，其核心逻辑是“递归迭代 + 特征排序”：首先以全部指标为输入训练基模型，通过模型输出的特征重要性对指标排序；随后剔除排名最末的若干个“最不重要”指标，基于剩余指标重新训练模型并排序；重复上述“训练 — 排序 — 剔除”过程，直至指标数量减少至 1 个。该算法的优势在于并非孤立评估单个特征，而是通过递归过程捕捉特征间的交互效应，确保保留的特征组合具有协同解释力，避免遗漏“单个重要性低但与其他特征联动后关键”的指标。

图 27：RFE 算法确定最优因子数量 | 单位：无



数据来源：华泰期货研究院

结果显示（上图横轴为引入模型的指标数量，纵轴为负均方误差）：当指标

数量为 18 时，负均方误差达到最大值（即 MSE 最小）；继续增加指标数量，负均方误差增幅趋于平缓（边际效果递减），表明新增指标多为噪音或重复信息，反而可能导致模型过拟合；若减少指标数量，负均方误差则显著下降，说明信息覆盖不足。

### （3）核心因子挖掘：线性与非线性关系的综合评估

目标明确后，即需从 102 个指标中精准挖掘出上述 18 个核心因子。为确保所选因子能全面捕捉铜价波动的驱动力，在沿用相关性分析、线性回归和支持向量机评估指标线性关系的基础上，额外引入了随机森林（Random Forest）和极端梯度提升树（XGBoost）两种强大的集成学习模型，以捕捉指标与铜价间的非线性关联及特征交互效应。

在特征重要性评估方面，两类模型均能输出相应指标以量化各变量对预测铜价的贡献程度。随机森林通常采用“平均 Gini 不纯度减少量”作为重要性度量，即统计某特征在所有树的分裂过程中带来的节点不纯度下降的平均幅度，下降越多表明该特征对区分样本铜价差异的作用越大。XGBoost 则常用“平均均方误差减少百分比”与“平均节点不纯度下降量”两项指标：前者反映某特征在所有决策树中，每次分裂所带来的均方误差（MSE）减少量占总 MSE 减少量的百分比均值，直接反映该特征对降低预测误差的贡献比例；后者则衡量其在分裂时降低节点数值离散程度的平均效能，数值越高代表特征对模型决策的影响越关键。这些重要性指标为从大量候选因子中筛选核心驱动变量提供了量化依据。

最终，我们为每个备选指标计算其在相关系数、信息比率、线性回归系数、同向变化率、Random Forest 的特征重要性、XGBoost 的平均均方误差减少百分比、XGBoost 的平均节点不纯度下降量等共计 7 个评估指标上的得分。对这些得分进行标准化并赋予相应权重加总，得到每个指标的综合得分。按综合得分从高到低排序，排名前 18 位的指标即被确定为样本内数据（2017-2021 年期间）最终的核心驱动因子组合，如下表所示：

**表 3：样本内铜的最优因子集 | 单位：无**

指标名称	相关系数	信息比率	线性回归系数	同向变化率	特征重要性	平均均方误差减少百分比	平均节点不纯度下降量	最终得分	综合排名
LME 现货价格	0.75	1.00	0.66	0.62	0.94	0.18	0.57	16.72	1
LME 铜收盘价	0.76	0.49	0.16	0.56	1.00	0.01	0.60	11.47	2

COMEX 铜收盘价	0.75	0.42	0.31	0.50	0.98	0.30	0.59	11.38	3
COMEX 铜报告头寸非商业多头持仓数量合计	0.55	0.25	1.00	0.72	0.39	0.40	0.27	10.65	4
LME 铝收盘价	0.76	0.27	0.27	0.45	0.74	0.18	0.44	8.94	5
美国 10 年期国债收益率	0.73	0.40	0.29	0.47	0.39	0.08	0.24	8.56	6
均价突破因子	0.78	0.46	0.19	0.59	0.11	0.09	0.07	8.27	7
LME 其他机构空头持仓总计	0.44	0.10	0.42	0.78	0.48	0.03	0.38	8.09	8
LME 投资基金多头持仓总计	0.43	0.11	0.32	0.32	0.76	0.56	0.60	7.69	9
美国通胀率预期：5 年	0.73	0.18	0.19	0.47	0.57	0.21	0.36	7.64	10
LME 其他机构多头持仓总计	0.42	0.03	0.17	0.84	0.58	0.07	0.46	7.52	11
按地区划分现货总计	0.76	0.38	0.20	0.58	0.06	0.18	0.03	7.48	12
LME 注册仓单	0.30	0.15	0.10	0.18	0.79	0.06	1.00	7.32	13
按地区划分升贴水总计	1.00	0.20	0.11	0.70	0.12	0.29	0.06	7.29	14
美国:美元指数	0.75	0.22	0.13	0.43	0.45	0.25	0.31	7.20	15
Curve 因子	0.84	0.18	0.13	0.67	0.26	0.20	0.14	7.19	16
CFTC 报告头寸非商业套利总计	0.92	0.17	0.18	0.60	0.33	0.02	0.17	7.17	17
进口废铜盈亏平衡	0.77	0.46	0.17	0.43	0.05	0.00	0.03	7.17	18
...	...	...	...	...	...	...	...	...	...

数据来源：华泰期货研究院

进一步的，考虑到铜价驱动逻辑可能随市场环境动态演变——某些因子在常态下贡献度有限，但在特定结构性变化时期（如地缘冲突、政策转向）可能成为关键驱动力——我们引入动态因子筛选机制作为优化。具体而言，在每个月的最后一个交易日，将截至该日的全部最新数据纳入训练集，重新运行前述的七维度评估流程（包括线性与非线性指标），为所有备选指标进行新一轮综合打分，并选取排名前 18 的因子作为下一个月的预测组合，旨在使因子体系能够自适应市场风格的切换，捕捉动态演变的定价逻辑。

表 4：不同时期最优因子动态变化 | 单位：无

2017-2021 年样本内 最优因子	综合得分	综合排 名	最新一期 最优因子	综合得分	综合排 名
LME 现货价格	16.72	1	LME 铜收盘价	17.37	1
LME 铜收盘价	11.47	2	LME 商业企业交易者数量	13.99	2
COMEX 铜收盘价	11.38	3	LME 铝收盘价	13.62	3
COMEX 铜报告头寸非商业多头持仓数量合计	10.65	4	COMEX 铜收盘价	12.97	4
LME 铝收盘价	8.94	5	偏度因子	12.00	5
美国 10 年期国债收益率	8.56	6	平水铜升贴水	11.74	6
均价突破因子	8.27	7	均价突破因子	11.28	7
LME 其他机构空头持仓总计	8.09	8	铜精矿港口库存	10.97	8
LME 投资基金多头持仓总计	7.69	9	期权隐含波动率	10.90	9
美国通胀率预期：5 年	7.64	10	湿法铜升贴水	10.55	10
LME 其他机构多头持仓总计	7.52	11	美元兑人民币汇率	10.40	11
按地区划分现货总计	7.48	12	LME 注册仓单	10.30	12
LME 注册仓单	7.32	13	现货沪伦比值	10.11	13
按地区划分升贴水总计	7.29	14	全国焦化企业开工率	10.04	14
美国:美元指数	7.20	15	布伦特原油期货价	9.87	15
Curve 因子	7.19	16	美国通胀率预期：5 年	9.34	16
CFTC 报告头寸非商业套利总计	7.17	17	螺纹钢价格指数	9.27	17
进口废铜盈亏平衡	7.17	18	美国:10 年期盈亏平衡通胀率	9.27	18
...	...	...	...	...	...

数据来源：华泰期货研究院

对比 2017 - 2021 年样本内期间与最新一期的最优因子组合可以发现,核心驱动因子构成发生了显著变化,这深刻反映了市场主导逻辑的演变。样本内期占据重要地位的因子,如 COMEX 铜报告头寸非商业多头持仓数量合计、LME 其他机构空头持仓总计等资金流指标,在最新组合中已被偏度因子、期权隐含波动率等更具实时性的市场情绪与风险指标所替代。同时,铜精矿港口库存、平水铜/湿法铜升贴水等贴近国内现货供需基本面的指标重要性显著提升,而螺纹钢价格指数、全国焦化企业开工率等关联品种指标的入选,则暗示了跨商品市场的传导



效应在近期定价中作用增强。这一对比充分说明，驱动铜价的核心因素并非一成不变，动态识别机制能够有效捕捉不同时期主导逻辑的切换，避免因依赖历史静态因子而错过结构性变化带来的预测信号。

### 3.5、预测效果

基于前文“组内选优 — RFE 提纯 — 多模型评分”层层筛选得到的 18 个核心驱动因子组合，为科学评估所构建因子体系在铜价预测中的实际效能，我们分别采用 Random Forest 与 XGBoost 两种机器学习模型进行日频预测，进而测算预测准确率。

每日在更新全部指标数据后，以最新数据为输入，通过训练完成的模型预测下一交易日铜价的涨跌幅；次日更新数据后，再滚动预测后续交易日走势，形成“数据更新 — 滚动预测”的循环机制。为降低单一模型偏差对结果的干扰，设定预测方向判定规则：若两类模型均预测涨跌幅为正数，则标记“预测上涨”；均预测涨跌幅为负数，则标记“预测下跌”；若模型间预测方向出现分歧，则将该日视为“无明确信号”，不纳入后续统计。通过将预测结果与铜期货在样本外期间（2022 年至今）的实际涨跌走势进行比对，可直观检验模型的有效性。

下图展示了预测结果与真实行情的对比情况。其中红色曲线代表铜期货价格的实际走势，黑色柱体表示模型预测正确的交易日，灰色柱体则表示预测错误的交易日，柱体高度对应当日价格波动幅度。从图中可以清晰观察到，黑色柱体在数量上明显多于灰色柱体，表明基于前期五年样本内数据严格筛选出的 18 个核心因子组合，在样本外期间能够持续提供有效的方向性预测信号，模型整体具备良好的泛化能力

图 28：基于最优因子组合的模型预测效果 | 单位：元



数据来源：SMM, 华泰期货研究院

为进一步量化核心因子组合的预测价值，排除“随机有效性”干扰，设计四组对照实验，通过控制模型参数、训练方法、回测周期等其他变量不变，仅改变输入因子组合以开展对比分析：

固定因子组合（实验组 1）：基于 2017 - 2021 年样本内数据筛选出的 18 个核心因子，因子组合在整个样本外期间保持固定 —— 此设定旨在验证“静态筛选的核心因子是否具备持续预测能力”；

动态因子组合（实验组 2）：以月度为频率，动态优选最新一批 18 个核心因子 —— 此设定旨在验证“筛选因子的方法是否能自适应市场风格的切换，捕捉动态演变”；

量价因子组合（对照组 1）：仅纳入 11 个量价类指标（如期限结构、动量因子、波动率因子等），不含任何基本面（产业链供需）与宏观因子 —— 用于检验“基本面与宏观因子的增量预测价值”；

所有因子组合（对照组 2）：采用组内选优后的 102 个指标，未经过 RFE 算法进一步提纯 —— 用于验证“数据降维对剔除冗余噪音、提升模型稳健性的作用”。

表 5：不同因子组合模型预测效果对比 | 单位：无

因子组合	方向准确率	预测正确期间价格波动（元）	预测错误期间价格波动（元）	盈亏比
所有因子	52.60%	170430	144480	1.180
量价因子	53.06%	177380	147810	1.200

固定因子	55.66%	197480	146990	1.343
动态因子	56.60%	195520	144270	1.355

数据来源：华泰期货研究院

通过严谨的回测与对照实验，我们验证了从数据降维到因子筛选整套方法的有效性。首先，“固定因子”组合的预测表现显著优于仅使用量价因子的对照组，也优于使用全部 102 个指标的对照组，这表明筛选流程能够在保留核心信息的同时，有效剔除冗余噪声，提炼出真正具有预测力的驱动因子。其次，引入月度频率的动态因子重组机制后，模型的方向预测准确性与盈亏比均获得进一步提升，证明动态识别机制能够有效适应市场逻辑的变化，及时捕捉新兴主导因子，是提升预测模型稳健性与适应性的关键优化。

本部分通过动态挖掘核心驱动因子，结合多算法筛选与专家经验双重验证，构建了以机器学习为核心的铜价预测体系。我们采用 XGBoost、Random Forest 等模型，有效捕捉了铜价在宏观、产业与金融等多维度影响下的非线性关系与时序依赖。模型在测试中展现出较高的预测精度，不仅验证了因子驱动的有效性，也提供了可靠的价格预判与决策支持。

## 四、新能源应用——碳酸锂价格预测

### 4.1、品种概况

碳酸锂作为全球新能源产业的核心原料，不仅处于锂产业链的定价中心位置，也同时具备显著的商品属性与金融属性。其商品属性来自于在动力电池、储能、电动车、3C 电子等领域的广泛应用，尤其在新能源汽车与储能体系中，碳酸锂是生产正极材料与电解液的重要原料。与此同时，其价格又强烈受供给弹性有限、地缘构成复杂的上游资源结构影响，使其在实际交易中逐渐体现出金融属性：成本曲线预期、库存周期、产业补库节奏与资金风险偏好都会对价格形成显著扰动，成为投资者判断新能源周期的关键资产。

全球锂资源储量丰富，根据美国地质调查局（USGS）数据，2021 年全球已探明的锂资源量达到 8900 万吨，折合 2200 万金属吨，1.12 亿吨 LCE。但分布非常不均，主要集中在南美、澳大利亚、中国等地。主要存在形式有盐湖卤水、锂辉石、锂云母等，从占比来看，卤水矿占比较大，全球占比达到 64%，中国锂资源中卤水矿占比达到 79%。

图 29：全球不同类型锂资源占比 | 单位：%

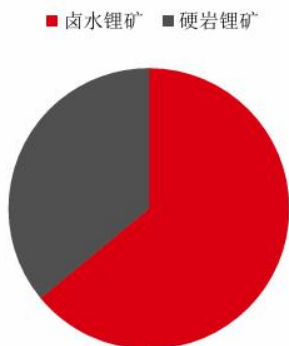
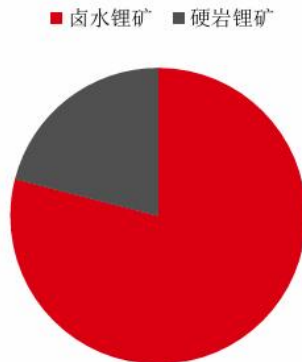


图 30：中国不同类型锂资源占比 | 单位：%

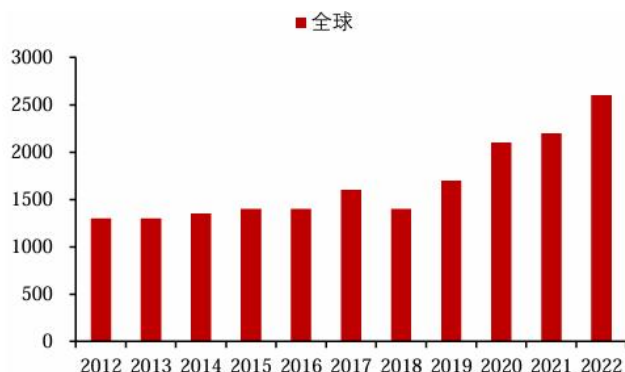


数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

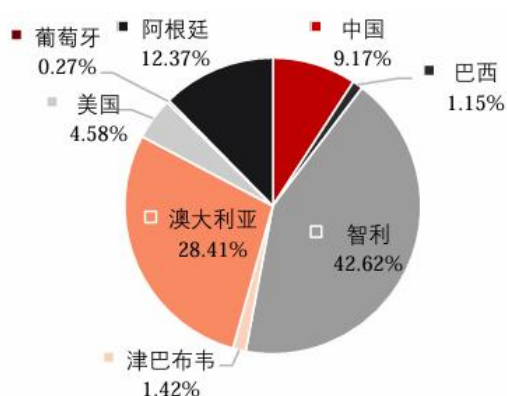
从全球供应体系来看，碳酸锂的生产结构多样化。其上游原料主要包括锂辉石、锂云母与盐湖卤水三类，其中南美“锂三角”和澳大利构成全球供给基础，中国则在云母提锂与盐湖资源方面占据优势。锂辉石适用于硫酸法等成熟工艺，成本可控、产品稳定。锂云母工艺路径更为复杂，需依赖硫酸盐焙烧、熟化等流程，能耗高、杂质多，但在资源端具有稀缺优势。盐湖卤水技术则受制于镁锂比、气候与工艺成熟度，沉淀法、吸附法、膜法、煅烧法等路线并存，盐湖资源储量虽大，但真正可工业化释放的产能增长仍受限。

图 31：全球锂储量 | 单位：万吨



数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

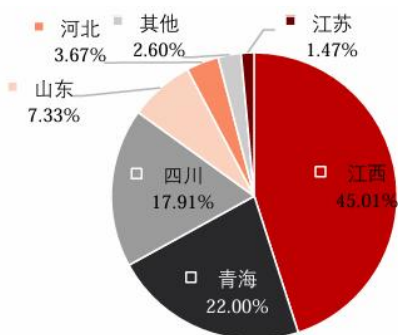
图 32：全球锂储量分布 | 单位：%



数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

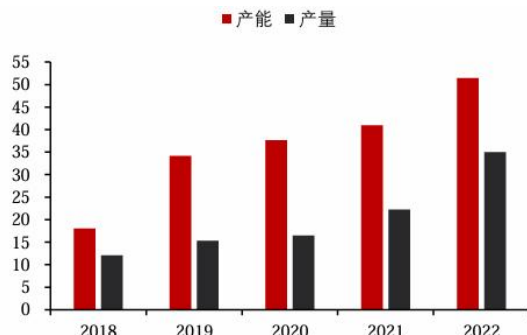
产业链供需格局决定了碳酸锂价格的高度波动性。近年来，中国碳酸锂产能持续扩张，2022 年产能已达 54 万吨，产量约 34 万吨，同比大幅增长 39.1%，2023 年仍有大规模新增投放。全球层面，锂资源储备集中度极高，智利、澳大利亚和中国占据绝对主导地位。其中澳洲锂辉石品位高，南美盐湖资源储量大但开发周期长，使得短期供给弹性偏弱，需求端的任何阶段性波动都会放大价格波动幅度。尤其在新能源汽车与储能装机快速扩张背景下，碳酸锂需求呈现趋势性增长，光伏配储能政策与整车销量扩张共同强化锂盐的中长期景气预期。

图 33：中国碳酸锂产能分布情况 | 单位：%



数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

图 34：中国碳酸锂产能产量 | 单位：吨



数据来源：华泰期货贵金属与有色专题 20230306《鹏程万“锂”系列报告-碳酸锂产业链概况》，华泰期货研究院

在价格形成机制上，碳酸锂兼具成本驱动与金融交易属性。一方面，上游锂辉石（6%品位）价格、硫酸等化工辅料成本及能源价格会直接推动碳酸锂成本曲线变化，使行业利润在景气周期中大幅波动。另一方面，盐湖提锂项目的开发



进度、锂云母资源政策调整、龙头企业扩产节奏、进出口变化等因素使市场情绪与风险偏好周期化反复。例如，硫酸法路线中硫酸与天然气的成本占比高，锂辉石价格上行会迅速推高行业边际成本，从而带动碳酸锂价格快速反弹。而在供应宽松或终端需求走弱阶段，库存累积会导致价格短期回落幅度加剧。

## 4.2、基本面研究框架

碳酸锂产业链覆盖资源开采、盐类冶炼、正极材料制造和终端电池消费，是新能源产业链中纵深最长、成本结构最复杂的关键环节。相较于传统金属，碳酸锂具有资源集中度高、工艺路线分化明显、区域供给结构分化、技术路径快速演化等特征，其价格高度敏感于供需差值、库存结构、政策预期以及上游矿石价格波动。产业链上游由锂辉石、锂云母和盐湖卤水构成，其中锂辉石与锂云母主要通过硫酸法、硫酸盐法等硬岩工艺生产，盐湖则依据卤水特性采用沉淀法、吸附法、膜法等路线。原料端差异巨大，使碳酸锂的边际成本呈现显著的多峰结构：澳洲锂辉石构成全球成本曲线中段，江西锂云母构成成本曲线右端，而青海盐湖与南美盐湖则构成低成本锚点。这种结构性供给的并存，使得碳酸锂在每一轮周期中对矿价变化、政策扰动以及季节性生产条件极为敏感，而这些因素最终决定冶炼厂的排产节奏、开工率以及有效产能的兑现程度。

中游产能是供给端的核心枢纽。中国是全球最大的碳酸锂生产国和贸易定价中心，2022 年国内碳酸锂产能达到 54 万吨，产量约 34 万吨，硫酸法与硫酸盐法依旧占据主导份额。在碳酸锂的生产体系中，不同原料路线带来完全不同的成本逻辑：锂辉石路线高度依赖澳洲矿价，锂云母路线受限于江西地区环保、能耗与资源配置政策，产量波动具有显著区域属性。而盐湖卤水路线则具有产能释放慢、回收率改善慢、季节性强（受气温、蒸发条件影响）的典型特征。中游企业需要同时处理杂质控制、能耗成本、酸耗碱耗、蒸发浓缩效率等问题，生产成本在不同阶段出现明显分化。当价格跌破云母成本时，高成本产能将被迫退出，形成产业链自发去产能的供给弹性，而当价格过高时，中游冶炼厂会加速排产，矿端也会加速扩产。

下游需求是碳酸锂长期趋势的决定力量，主要由动力电池、储能电池、3C 消费电子与两轮电动车构成。动力电池中，磷酸铁锂和三元材料对碳酸锂的需求结构差异明显，不同电池技术路线、车型结构以及整车产销节奏，会直接改变碳酸锂的边际需求路径。在过去三年中，中国新能源汽车对碳酸锂的需求保持高速增

长，电池总装机量持续上升，同时储能产业快速扩容。需求端真实消费的最终体现在正极材料产量、动力电池装机量、库存周转天数等指标上，不同环节的链条反馈对价格有着非线性作用。

图 35：碳酸锂产业链示意 | 单位：无



数据来源：华泰期货研究院

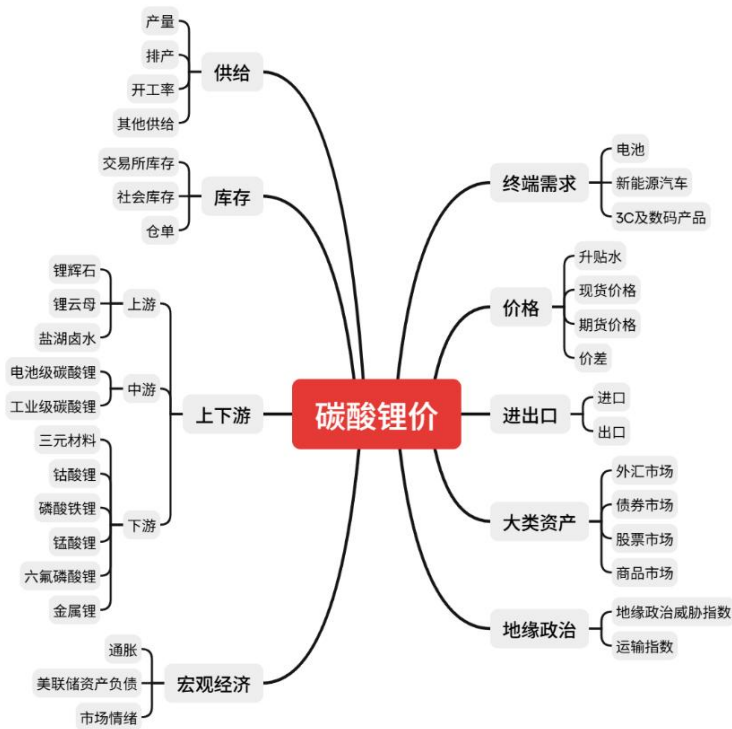
库存结构是碳酸锂现货价格的核心短周期指标。碳酸锂作为化工属性较强的商品，其显性库存主要分布在冶炼厂、正极材料厂、电池厂以及交易所和社会环节。库存的结构分布，决定了现货的紧张程度和议价权归属。当产业链库存快速去化时，冶炼厂与贸易商的议价能力增强。反之，则出现需求不及预期一库存被动累库一现货贴水扩大的反应。期货上市后，交易所仓单体系为市场新增了一套可观察、可交割的显性库存指标，但仓单规模相较产业链实际库存仍然有限，它更多承担的是价格发现与跨期结构展示的功能。基差、月间价差与升贴水等结构性价格信号，逐渐成为市场判断供需紧张度、资金仓位结构以及贸易流向变化的重要工具，使碳酸锂在保留化工品现货属性的同时，也呈现出更典型的大宗商品金融化特征。

进出口流向是碳酸锂市场的国际定价通道。中国既是最大的生产国，也是最大的进口国，进口以南美盐湖一级产品为主，而出口则多为冶炼后的电池级产品。进出口利润决定了贸易商是否进行跨境套利，而这种套利行为又直接改变国内供需格局，使得中国价格与国际价格通过贸易通道紧密联动。

宏观经济、美元流动性与大类资产情绪共同塑造碳酸锂的金融属性。碳酸锂虽不具备金属属性，但其产业链高度资本化且依赖外矿，使其显著受美元指数、美国实际利率、汇率波动以及资金风险偏好影响。在大宗商品维度，碳酸锂与黑色系、能源品、贵金属并非同周期，但交易资金具有明显的横向迁移效应：当市场整体风险偏好上升时，碳酸锂等高波动品种更容易获得增量资金，价格弹性也更高。此外，新能源汽车产业链股票、锂矿上市公司股价、大型锂盐企业的资本支出节奏，也可以作为市场对锂产业景气度的领先指标。

地缘政治与突发事件则为碳酸锂价格提供显著的风险溢价。澳洲矿山劳资纠纷、南美盐湖政策变化、智利/阿根廷的矿权制度改革、国内江西地区环保督察、青海盐湖极端天气扰动，都可能改变供给预期。而国际贸易政策变化、全球疫情、港口卡点等因素则可能影响碳酸锂的跨境物流成本和供给节奏。在碳酸锂这种供给链条长、环节复杂的商品中，给定一个高成本、长周期的供应体系，任何扰动都可能在短期内放大价格波动，加剧供需错配。

图 36：碳酸锂研究框架 | 单位：无



数据来源：华泰期货研究院

### 4.3、数据收集与处理

针对碳酸锂价格波动大、影响因素复杂的特征，我们在构建预测模型时，为系统捕捉碳酸锂价格波动的核心影响要素，我们系统整合了涵盖资源端、冶炼端、消费端与金融市场等多个维度的海量数据。构建了覆盖资源储量、产能投放、开工率、库存结构、进出口流向、成本曲线、技术路线演变及政策动向的全链条数据库，最终形成包含 300 余个指标的清单，下表仅展示具有代表性的部分指标：

表 6：碳酸锂相关指标梳理 | 单位：无

一级分类	二级分类	指标名称
库存	交易所库存	交割厂库库存、交割仓库库存
	社会库存	社会库存、海外企业库存、仓库库存、冶炼厂库存、下游库存
	仓单	交易所期初、期末库存
价格	现货价	电池级工业级均价、市场价、贸易散单均价
	升贴水	不同原料端升贴水、升贴水区间
	期货价	当月合约、次月合约
	价差	电池级-工业级价差

上下游	上游	锂辉石、锂云母、盐湖卤水、磷锂铝石、运费、库存、上游进出口、上游产量
	中游	电池级碳酸锂、工业级碳酸锂
	下游	三元材料、钴酸锂、锰酸锂、磷酸铁锂、六氟磷酸锂、金属锂
进出口	进口	进口均价、数量、金额
	出口	出口均价、数量、金额
供给	产量	按原料划分产量
	排产	按原料划分排产
	开工率	各原料端开工率、开工负荷率
	供给其他	产能利用率、可交割产量占比
终端需求	电池	电池产量、装车量、库存
	新能源汽车	新能源汽车产量、销量、库存、进出口数量、保有量
	3C 及数码产品	手机出货量、上市数量、产量、平板电脑进出口、微型计算机设备产量
成本利润	成本	生产成本、加工成本、现金成本、提纯飞
	利润	生产利润、进口利润、毛利、理论交割利润
大类资产	外汇市场	美元兑人民币汇率、美元指数
	债券市场	美国国债收益率（2 年期 10 年期）、中国国债收益率（2 年期 10 年期）
	股票市场	上证综合指数
	商品市场	工业硅、镍、铜、原油
	通胀预期	美国通胀率预期、美国盈亏平衡点通胀率
宏观经济	美联储资产负债表	美联储各类资产及负债当周值
	市场情绪	VIX 指数、OFR 金融压力指数、期权隐含波动率、花旗经济意外指数
地缘政治	地缘政治	地缘政治威胁指数
	海运指数	波罗的海运输指数、上海出口集装箱运价指数
量价	量价因子	Curve、动量、期限结构、偏度、波动率、价值、持仓、流动性、均价突破、宏观 Beta 因子

数据来源：华泰期货研究院

在完成指标初选后，为进一步提升数据质量与模型适用性，我们对所有纳入的指标进行统一预处理，具体包括：（1）统一频率（2）平稳性检验（3）异常值缩尾处理（4）样本划分：将 2020 年 1 月至 2023 年 12 月设定为样本内区间，用于因子挖掘与模型训练；2024 年 1 月以后的数据作为样本外区间，用于模型



预测效果的验证与优化。

在此基础上，为更精准地平衡短期盘面异动与中期趋势判断，我们将预测体系细分为周度预测与月度预测两个维度。这并非简单地将同一模型运行于不同时间颗粒度，而是构建了两套目标不同、因子侧重不同、模型微调策略亦不同的、既独立又协同的预测系统。其中，周度预测更侧重于快速响应最新市场信息，捕捉短期价格波动；月度预测则更聚焦于趋势方向的判断与关键价位区域的把握。两者协同验证、相互补充，共同提升模型在全周期中的预测能力。

## 4.4、数据降维

### 1) 组内选优：主要基于线性评估模型的初步筛选

为在保留核心信息的同时有效降低数据维度，我们首先采用“组内选优”策略。具体而言，我们依据经济学含义将全部指标划分为一级、二级和三级分类。在同一个三级分类下，各指标间不仅经济含义相近，且通常存在较高的统计相关性。为避免信息重叠和多重共线性，我们在每个三级分类内部，采用多种线性模型评估各指标与碳酸锂价格之间的线性关联强度，仅保留得分最高的一个指标，所使用的评估维度包括：

**相关系数：**通过皮尔逊相关系数（PCC）衡量因子与预测标的价格的线性关联程度。

**信息比率：**通过皮尔逊相关系数（PCC）衡量因子的变化率与预测标的变化率的线性关联程度。

**同向变化率：**因子与预测标的同向变化的比率。

**回归系数 Beta：**标准化处理后，基于拟合因子与预测标的价格拟合线性回归模型（OLS），通过回归模型系数来衡量因子与预测标的的线性相关程度。

**回归模型 MAE：**通过平均绝对误差衡量用因子预测下一期标的价格的准确程度。

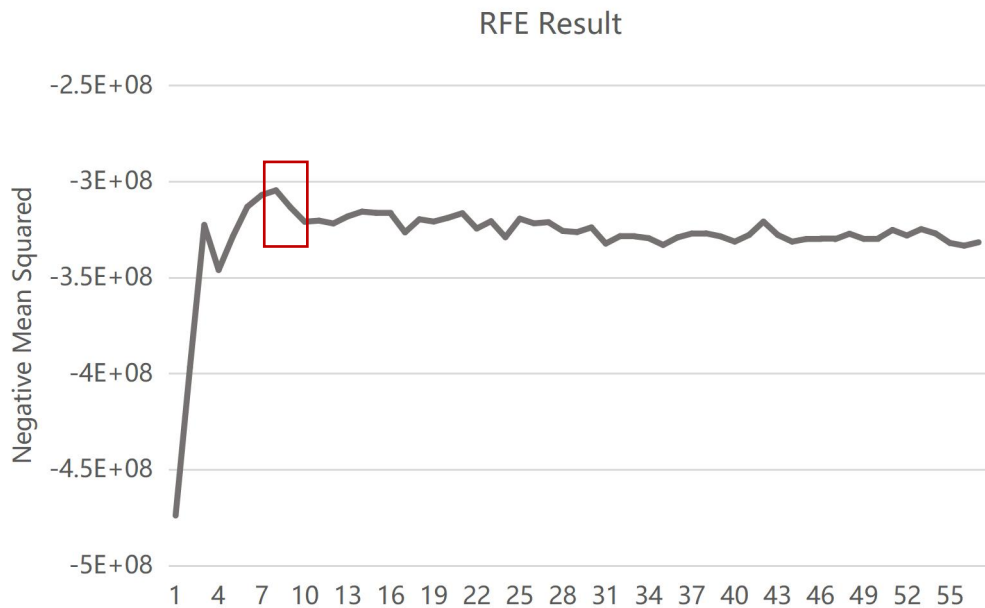
**最大信息系数 MIC：**用于衡量两个变量之间的互信息，最大信息系数越大，因子的预测效力越强。

综合上述评估维度进行加权打分，每组仅保留综合得分最高的指标，最终将初始 300 余个指标精简至 86 个，既剔除了冗余信息，又保留了各细分逻辑下的核心关联指标。

(2) 最优因子数量确定：递归特征消除的应用

采用递归特征消除 (Recursive Feature Elimination, RFE) 算法来确定最优因子数量，既确保给模型带入足量的信息用于预测下一期价格变化，又不会带模型带去过多噪音。结果显示当指标数量为 8 时，负均方误差达到最大值 (即 MSE 最小)，此时为最佳预测效果。

图 37：碳酸锂研究框架 | 单位：无



数据来源：华泰期货研究院

目标明确后，即需从 86 个指标中精准挖掘出最终的 8 个核心因子。为确保所选因子能全面捕捉碳酸锂价格波动的驱动力，在沿用相关性分析、线性回归模型的基础上，额外引入了随机森林 (Random Forest) 和极端梯度提升树 (XGBoost) 两种机器学习模型，以捕捉指标与碳酸锂价格间的非线性关联及特征交互效应。

最终，我们为每个备选指标计算其在相关系数、信息比率、线性回归系数、同向变化率、最大信息系数、平均绝对误差、Random Forest 的特征重要性、XGBoost 的平均均方误差减少百分比、XGBoost 的平均节点不纯度下降量等共计 9 个评估指标上的得分。对这些得分进行标准化并赋予相应权重加总，得到每个指标的综合得分。按综合得分从高到低排序，排名前 8 位的指标即被确定为样本内数据 (2020-2023 年期间) 最终的核心驱动因子组合，如下表所示：

表 7：样本内碳酸锂的最优因子集 (周频预测) | 单位：无

指标名称	相关	信息	线性回	同向	最大信	平均绝	特征	平均均方	平均节点	最终得	综
------	----	----	-----	----	-----	-----	----	------	------	-----	---

	系数	比率	归系数	变化率	息系数	对误差	重要性	误差减少百分比	不纯度下降量	分	合排名
华东市场碳酸锂(电池级)日度市场价	0.08	0.97	0.77	0.94	0.76	0.70	1.00	1.00	0.45	20.59	1
生产利润_总计	0.07	0.60	1.00	0.50	1.00	1.00	0.13	0.42	0.77	15.03	2
工业级碳酸锂 - 平均价	0.11	0.91	0.71	1.00	0.65	0.51	0.23	0.43	0.50	12.56	3
磷酸铁锂价格_总计	0.11	0.59	0.73	0.78	0.59	0.87	0.03	0.03	0.45	10.31	4
工业硅_总计	0.38	0.10	0.29	0.13	0.68	0.56	0.01	0.00	1.00	8.97	5
锰酸锂(动力型)-平均价	0.12	0.79	0.58	0.73	0.57	0.47	0.02	0.02	0.45	8.59	6
碳酸锂--电渗析法-碳酸锂-毛利润	0.19	0.27	0.45	0.62	0.64	0.60	0.01	0.01	0.60	8.38	7
三元正极材料价格: 动力单晶型_平均	0.18	0.40	0.56	0.43	0.28	0.53	0.01	0.03	0.71	8.19	8
...	...	...	...	...			...	...	...	...	...

数据来源：华泰期货研究院

对于周频预测与月频预测两个维度划分，虽然标的价格是一样的，但原始的日频数据会经过不同的降频处理，选拔出来的前 8 名核心因子也会有所区别：

表 8：样本内碳酸锂的最优因子集（月频预测） | 单位：无

指标名称	相关系数	信息比率	线性回归系数	同向变化率	最大信息系数	平均绝对误差	特征重要性	平均均方误差减少百分比	平均节点不纯度下降量	最终得分	综合排名
三元前驱体生产成本_总计	0.02	0.62	1.00	0.86	0.36	0.75	1.00	0.72	0.81	20.27	1
国内现货锂矿提炼碳酸锂工业级中	0.20	0.45	0.73	0.77	0.29	0.87	0.59	1.00	0.45	16.31	2

国月度生产装置											
成本											
生产利润_总计	0.21	0.41	0.93	0.43	1.00	0.95	0.16	0.00	0.94	13.94	3
六氟磷酸锂利润_平均	0.09	0.46	0.63	0.69	0.43	0.46	0.65	0.17	0.72	13.54	4
金属锂-有效产能-月度	0.25	0.34	0.86	0.14	0.82	0.54	0.31	0.03	0.93	12.60	5
新能源汽车：产量：中国（月）	0.36	0.27	0.46	0.54	0.93	0.64	0.23	0.04	0.54	10.61	6
碳酸锂--硫酸法-碳酸锂-成本	0.23	0.14	0.45	0.74	0.31	0.68	0.08	0.00	0.95	9.83	7
现金成本_总计	0.41	0.32	0.99	0.51	0.80	1.00	0.06	0.00	0.00	9.58	8
...	...	...	...	...			...	...	...	...	...

数据来源：华泰期货研究院

进一步的，考虑到价格驱动逻辑可能随市场环境动态演变，我们引入动态因子筛选机制作为优化。具体而言，在每个月的最后一个交易日，将截至该日的全部最新数据纳入训练集，重新运行前述的两轮评估流程，为所有备选指标进行新一轮综合打分，并选取排名前 8 的因子作为下一个月的预测组合，旨在使因子体系能够自适应市场风格的切换，捕捉动态演变的定价逻辑。

表 9：不同时期最优因子动态变化 | 单位：无

2020-2023 年样本内 最优因子	综合得分	综合排名	最新一期 最优因子	综合得分	综合排名
华东市场碳酸锂(电池级)日度 市场价	20.59	1	华东市场碳酸锂(工业级)日度市场 价	23.19	1
生产利润_总计	15.03	2	工业级碳酸锂 - 平均价	16.46	2
工业级碳酸锂 - 平均价	12.56	3	锰酸锂(动力型)-平均价	9.49	3
磷酸铁锂价格_总计	10.31	4	碳酸锂生产利润：外购锂云母精矿 (Li <sub>2</sub> O:2.5%)	9.38	4
工业硅_总计	8.97	5	碳酸锂：生产毛利：中国（周）	9.37	5
锰酸锂(动力型)-平均价	8.59	6	磷酸铁锂（高端动力型） - 平均价	8.62	6
碳酸锂--电渗析法-碳酸锂- 毛利润	8.38	7	现金成本_总计	8.57	7
三元正极材料价格：动力单晶 型_平均	8.19	8	三元正极材料：动力多晶型：622 型： 市场价：中国（日）	8.29	8
...	...	...	...	...	...

数据来源：华泰期货研究院

4.5、预测效果

为科学评估所构建因子体系在碳酸锂价格预测中的实际效能，我们采用 Random Forest 模型分别进行周频和月频预测，并通过计算样本外期间（2024 年至今）的平均偏差率（MAPE）、预测方向准确率和盈亏比来综合评估预测准确性。

为进一步量化核心因子组合的预测价值，排除“随机有效性”干扰，设计三组对照实验，通过控制模型参数、训练方法、回测周期等其他变量不变，仅改变输入因子组合以开展对比分析：

固定因子组合：基于 2020 - 2023 年样本内数据筛选出的 8 个核心因子，因子组合在整个样本外期间保持固定；

滚动因子组合：以月度为频率调整，滚动两年历史数据为训练集，重复因子筛选流程；

动态因子组合：以月度为频率调整，累积所有历史数据为训练集，重复因子



筛选流程。

表 10：不同因子组合模型周频预测效果对比 | 单位：无

因子组合	近期平均偏差率	历史平均偏差率	方向准确率	盈亏比
固定因子	1.26%	1.54%	48.91%	0.948
滚动因子	1.71%	1.79%	54.35%	1.409
动态因子	1.23%	1.68%	59.78%	1.638

数据来源：SMM， 华泰期货研究院

图 38：基于最优因子组合的周频预测效果 | 单位：元 | %



数据来源：卓创资讯，华泰期货研究院

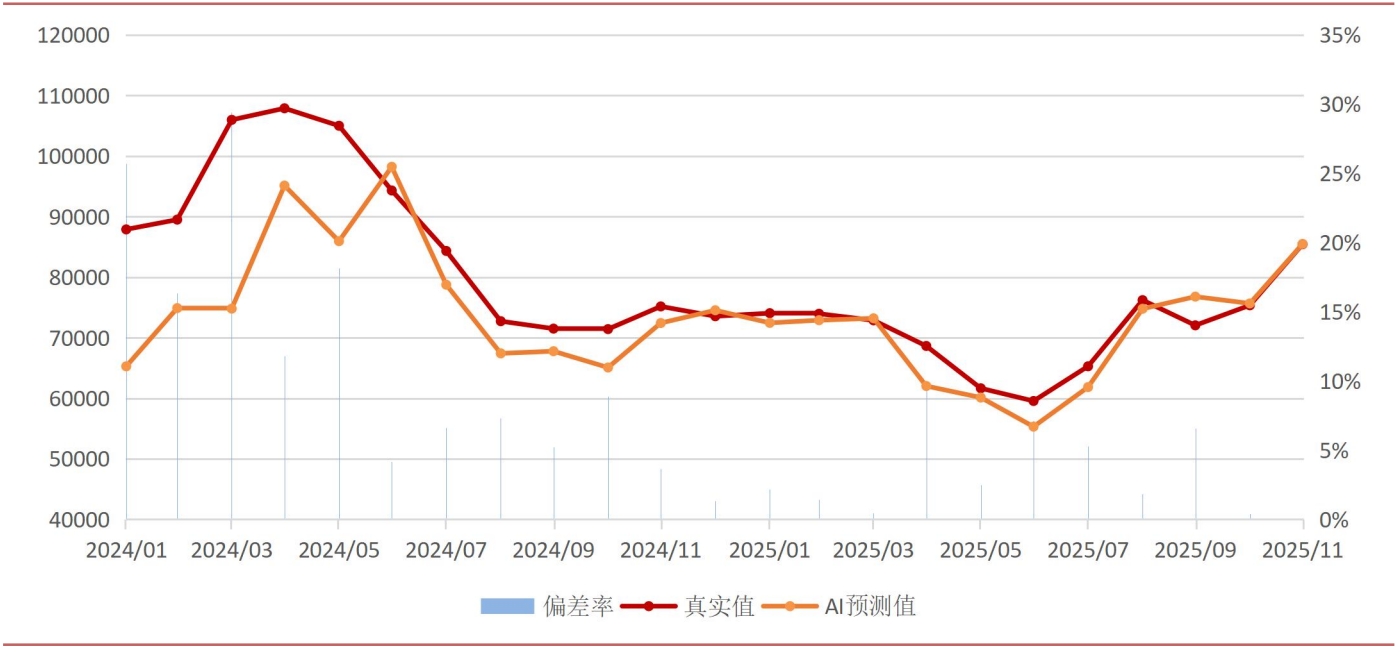
在周频预测中，模型凭借对高频数据的快速响应与因子动态赋权，准确捕捉了价格的主要波动方向与幅度，历史回测偏差率持续处于较低水平，尤其在识别短期市场情绪转折引发的价格异动方面表现突出。

表 11：不同因子组合模型月频预测效果对比 | 单位：无

因子组合	近期平均偏差率	历史平均偏差率	方向准确率	盈亏比
固定因子	3.67%	7.65%	52.17%	0.966
滚动因子	4.35%	12.36%	39.13%	0.761
动态因子	1.97%	7.12%	56.52%	1.974

数据来源：SMM， 华泰期货研究院

图 39：基于最优因子组合的月预测效果 | 单位：元 | %



数据来源：卓创资讯，华泰期货研究院

在月频预测中，则依托对供需结构与成本边际的深入分析，精准预判了中期价格的关键支撑与压力区间，随着回测期的推移，训练集的数据点足够多之后，趋势判断与真实走势高度一致，验证了模型在结构性行情中的稳健性。

本部分通过构建双频预测体系，结合动态因子挖掘与机器学习建模，有效应对了碳酸锂价格的高波动特征。模型在周度与月度维度均展现出优异的预测能力，不仅精准捕捉短期波动，也稳健把握中期趋势，为产业链企业与投资者提供了兼具时效性与前瞻性的价格决策工具。

## 总结

本报告基于 2025 年商品市场波动率持续抬升，大宗市场不确定性加强的背景，围绕“专家知识+ AI 模型”的商品价格预测体系展开系统化梳理。我们以覆盖宏观、产业链与高频量价在内的全量指标池为基础，通过相关性、信息增益、MIC 等多算法手段构建核心因子组合，并在三层研究框架（长期趋势—中期供需—短期行情）中将行业逻辑与统计显著性结合，形成可迭代的因子体系。模型端以 ARIMAX、XGBoost、LSTM、Transformer 等多层结构为主体，通过验证集反馈与专家复核实现逻辑与模型间的双向校准，并在华泰天玑平台上完成产品化，包括周度/月度预测、模型解析、预测验证与情景推演模块。

在铜价预测案例中，我们从矿山、冶炼、加工到终端与再生构建全链条指标体系，并通过动态优选因子机制，精准捕捉新兴驱动因子，以适应不同阶段的定价主导力量。样本外结果显示，动态因子版本在方向性准确率与盈亏比上均显著优于静态固定因子与全因子模型，日频预测方向准确率达 56.6%，盈亏比 1.355。

在碳酸锂部分，我们将资源端产能扩张节奏、成本曲线、库存周期与新能源终端变量整合成多维指标体系。为更精准地平衡短期盘面异动与中期趋势判断，我们分别构建周频与月频预测模型。其中周度预测更侧重于快速响应最新市场信息，捕捉短期价格波动，周频模型在样本外统计期间 MAPE 仅 1.68%，方向准确率 59.68%，盈亏比 1.638；月度预测则更聚焦于趋势方向的判断与关键价位区域的把握。月频模型在样本外统计期间 MAPE 为 7.12%，方向准确率 56.52%，盈亏比达 1.974。

整体来看，本体系以数据驱动与专家逻辑双核协同，在高波动、强非线性的大宗商品市场中提供了可量化、可解释、可推演的价格预测框架，并在实际应用中显示出良好的稳健性与扩展潜力。

## 免责声明

本报告基于本公司认为可靠的、已公开的信息编制，但本公司对该等信息的准确性及完整性不作任何保证。本报告所载的意见、结论及预测仅反映报告发布当日的观点和判断。在不同时期，本公司可能会发出与本报告所载意见、评估及预测不一致的研究报告。本公司不保证本报告所含信息保持在最新状态。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。

本公司力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，投资者并不能依靠本报告以取代行使独立判断。对投资者依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华泰期货研究院”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权力。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

华泰期货有限公司版权所有并保留一切权利。

## 公司总部

广州市天河区临江大道 1 号之一 2101-2106 单元 | 邮编：510000

电话：400-6280-888

网址：www.htfc.com



客服热线：400-628-0888

官方网址：[www.htfc.com](http://www.htfc.com)

公司总部：广州市南沙区横沥镇明珠三街 1 号 10 层