深度报告-金融工程

衍生品量化择时系列专题（九）——基于 Prophet 和模式识别的基本面量化模型优化

东方证券 ORIENT SECURITIES | 期货

报告日期： 2024-3-28

谢怡伦 金融工程分析师

从业资格号: F03091687

投资咨询号: Z0019902

Tel: 8621-63325888-1585

Email: yilun.xie@orientfutures.com

★研究背景：

在先前的研究中，我们已经构建了相对成熟的商品基本面量化策略，并且已经在繁微平台上对该策略进行了一段时间的跟踪，在长达两年的跟踪时间内，策略表现整体运行良好，在不加杠杆情况下，年化收益率为 7.32%，夏普值为 1.45，最大回撤为 4.08%，收益风险比 1.8，胜率为 58%，sortino 比率为 2.19。在此基础上，此报告针对原有模型进行改进以达到更好的预测效果。

★模型优化：

针对基础模型进行多方面的优化。首先精简因子数量，使得模型容易跟踪和复现；其次基于 Prophet 预测模型对商品基本面数据进行趋势数据的剥离，并尝试对价格序列本身进行自预测；最后引入模式识别机制，判断当前市场行情是否由基本面数据所控主导，若不是，则采用量价模式下的信号生成机制。

★Prophet 模型：

Prophet 是由 Facebook 于 2017 年推出的时序数据预测模型，相较于传统的时序预测模型具有如下一些优势。**灵活性**：无需对原始数据进行平稳性处理，同时可以对数据进行不同周期的趋势性分解；**可解释性**：模型通过分解出趋势、周期、事件影响这三要素，能够直观反映出各部分对于数据的影响；**效率高**：Prophet 算法运行效率较快，相较于机器学习模型的不断训练，Prophet 只要进行一次拟合即可。

★模型结果：

在采用模式识别机制下的 Prophet 趋势因子优化之后，模型对于单品种的回测表现为：年化收益 31.69%，夏普值接近 1.79，最大回撤控制在 18.71%，胜率为 60%；样本外表现为夏普值接近 1.60，胜率为 60%。

★风险提示

结论基于历史数据分析，未来市场规律的变动可能使结论失效。

扫描二维码，微信关注“东证繁微”小程序

金融工程

重要事项：本报告版权归上海东证期货有限公司所有。未获得东证期货书面授权，任何人不得对本报告进行任何形式的发布、复制。本报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，报告中的信息或意见并不构成交易建议，投资者据此做出的任何投资决策与本公司和作者无关。

有关分析师承诺，见本报告最后部分。并请阅读报告最后一页的免责声明。

目录

1、 前言 5
2、 基础模型 5
3、 Prophet 模型 6
4、 优化模型 8
5、 回测框架以及参数设置 10
6、 单因子检验 11
6.1、 因子概述 11
6.2、 因子表现 14
7、 多因子模型 17
7.1、 多因子模型-方案 1 18
7.2、 多因子模型-方案 2 20
7.3、 样本外验证 21
8、 基于 Prophet 的基本面因子预测 22
8.1、 基于基本面数据趋势进行预测 22
8.2、 基于收益率序列进行 Prophet 自预测 24
8.3、 基于价格序列进行 Prophet 自预测 25
9、 模式识别 26
9.1、 动量趋势信号和 Prophet 趋势信号的对比 26
9.2、 模式识别下的趋势量价信号生成 27
9.3、 模式识别下的 Prophet 趋势因子优化 28
10、 模型应用与展望 29
10.1、 模型的应用 29
10.2、 多品种策略的研发 30

图表目录

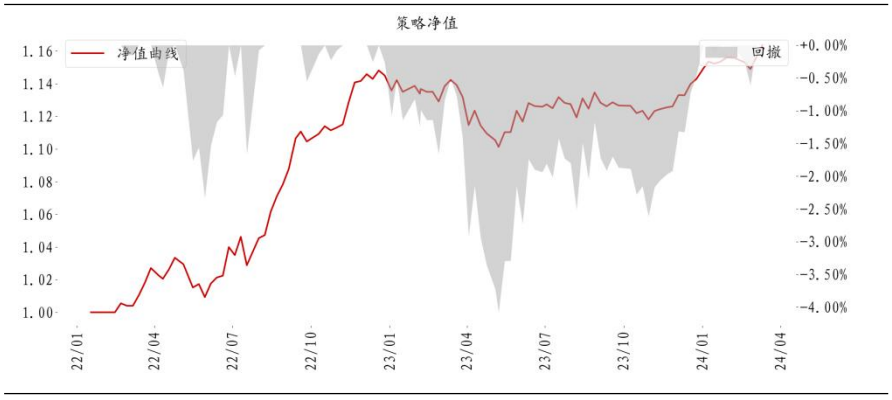
图表 1：全品种基本面量化策略跟踪表现.....5
图表 2：回测指标.....5
图表 3：基本面量化基础模型框架.....6
图表 4：Prophet 模型趋势分解.....7
图表 5：Prophet 预测效果示意.....8
图表 6：基本面量化优化模型框架.....9
图表 7：滚动回归.....10
图表 8：基本面因子分类.....12
图表 9：基本面因子举例.....13
图表 10：因子 DZ02010751 回测曲线.....14
图表 11：回测指标.....14
图表 12：因子 DZ02016959 回测曲线.....14
图表 13：回测指标.....14
图表 14：因子 DZ02050869 回测曲线.....15
图表 15：回测指标.....15
图表 16：因子 DZ02032106 回测曲线.....15
图表 17：回测指标.....15
图表 18：基本面因子回测表现（按夏普值降序）.....16
图表 19：因子选取方案 1.....18
图表 20：因子选取方案 2.....18
图表 21：因子选取方案 1（20 个因子）.....19
图表 22：因子选取方案 1 回测表现.....19
图表 23：回测指标.....19
图表 24：因子选取方案 2（30 个因子）.....20
图表 25：因子选取方案 2 回测表现.....21
图表 26：回测指标.....21
图表 27：因子选取方案 1 回测表现——样本外.....21
图表 28：回测指标.....21
图表 29：因子选取方案 2 回测表现——样本外.....22
图表 30：回测指标.....22
图表 31：回测表现.....23
图表 32：回测指标.....23
图表 33：回测表现——样本外.....23
图表 34：回测指标.....23
图表 35：回测表现.....24
图表 36：回测指标.....24
图表 37：回测表现——样本外.....24
图表 38：回测指标.....24

图表 39：回测表现.....25
图表 40：回测指标.....25
图表 41：回测表现——样本外.....25
图表 42：回测指标.....25
图表 43：回测表现.....26
图表 44：回测指标.....26
图表 45：回测表现——样本外.....27
图表 46：回测指标.....27
图表 47：回测表现.....27
图表 48：回测指标.....27
图表 49：回测表现——样本外.....28
图表 50：回测指标.....28
图表 51：回测表现.....28
图表 52：回测指标.....28
图表 53：回测表现——样本外.....29
图表 54：回测指标.....29

1、前言

在早先的研究中，我们已经构建了相对成熟的商品基本面量化策略，并且已经在繁微平台上对该策略进行了一段时间的跟踪，在长达两年的跟踪时间内，策略表现整体运行良好，在不加杠杆情况下，年化收益率为 7.32%，夏普值为 1.45，最大回撤为 4.08%，收益风险比 1.8，胜率为 58%，sortino 比率为 2.19。策略在 2023 年全年经历了一波较长周期的震荡区间，其余时间段均表现出相对优秀的业绩表现。

图表 1：全品种基本面量化策略跟踪表现



资料来源：东证衍生品研究院

图表 2：回测指标

总收益	16.36%
年化收益	7.32%
年化波动	4.37%
夏普值	1.45
最大回撤	-4.08%
收益风险比	1.8
胜率	0.58
盈亏比	1.32
sortino 比率	2.19

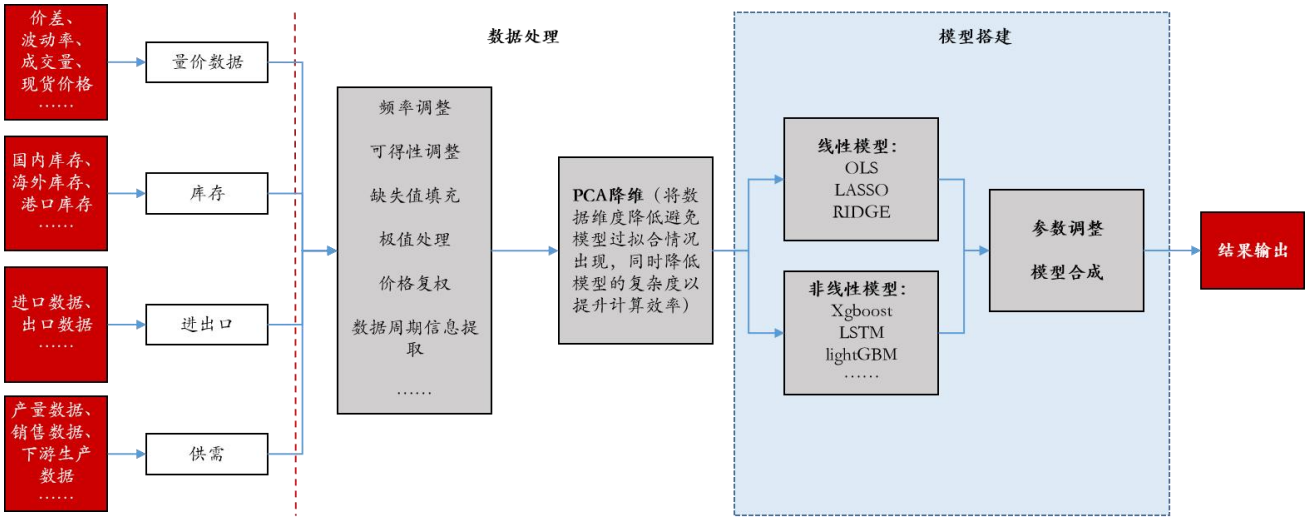
资料来源：东证衍生品研究院

然而在策略的运行过程中，随着对策略的认识不断深入，我们也意识到该策略仍然存在一些值得改进的地方。首先由于该策略的基本面数据基于繁微内部的数据库，导致数据量相对冗余，由于其数据来源相对多样，使得投资者难以复现该策略，我们会在本报告中精简基本面因子的数量，并且我们会通过 Prophet 模型对基本面数据本身进行预测，再利用预测的基本面数据对期货价格进行拟合。其次商品期货的行情走势在一段时间内会出现偏离基本面走势的情况，若不进行调整则会出现相对大幅且长时间的回撤，故在本报告中我们会引入“模式识别”的机制，判断当前行情能否完全由基本面数据所解释，若不能，我们则通过价量因子或者 Prophet 预测模型生成相应信号，这样一来可以在商品基本面偏离价格走势的行情中得到更为有效的因子信号。

2、基础模型

在基础模型中，我们首先对商品基本面数据进行初步的梳理和分类，同时也引入一定的量价数据，根据商品本身的特性，数据分类包括但不限于量价数据、库存数据、进出口和供需数据等基本面数据分类。之后基于梳理过的商品基本面数据进行一系列数据处理，包括频率调整、可得性调整、缺失值填充等，再将处理完的因子通过 PCA 主成分分析法进行降维以降低后续模型的复杂程度，也可以避免模型过拟合的情况出现。在模型搭建层面，我们的做法结合线性模型和非线性模型各自的信号特征，经参数调整后进行信号的合成，最后生成综合信号。在这篇报告的优化部分，我们首先会对原始的基本面数据进行筛选和精简，其次会引入模式识别的机制，通过价量因子和 Prophet 预测模型提升表现。

图表 3：基本面量化基础模型框架



资料来源：东证衍生品研究院

3、Prophet 模型

对于时间序列数据的预测任务，传统的预测方法为自回归预测，例如自回归综合移动平均模型（ARIMA）和自回归条件异方差模型（GARCH），然而该类预测模型需要时间序列数据达到平稳状态，适用于单变量、小规模的预测，对于稳定性较差、非线性的数据，其预测效果并不理想。

此外，还可采用机器学习或者深度学习模型进行训练预测，例如 XGBoost 和 RNN（循环神经网络）等，此类预测模型能够处理相对复杂的训练任务，且预测精度相对较高。然而，机器学习类的预测模型往往缺乏可解释性，其预测结果的输出我们认为是一个“黑箱”，无法知晓其从数据到结果的转化逻辑。

Prophet 预测模型能够很好结合上述两类模型的优势，其预测方式更为简洁，具有逻辑上的可解释性，同时也能达到较高的预测精度。Prophet 预测模型框架最早在 2017 年由 Facebook 工程师提出并且进行开源，该模型适用于各类具有潜在时序特征的数据，并且能够对时序数据的趋势变化、季节性趋势、节假日影响以及突发事件影响具有更好的拟合效果。

Prophet 模型可以选择加法模型和乘法模型，本报告只涉及加法模型。Prophet 算法将时间序列趋势分解为三个主要的组成部分，分别为增长趋势、季节性周期和节假日影响。

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

如上式所示，$g(t)$ 项用于拟合时序数据非周期的趋势变化；

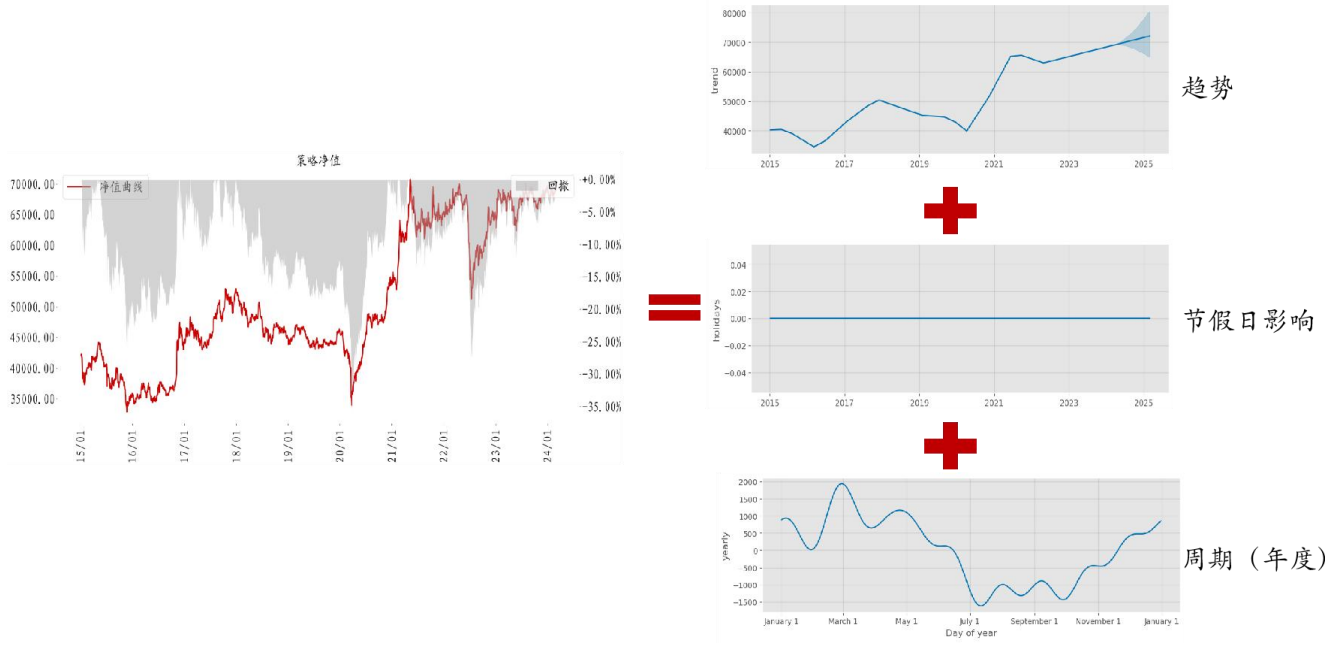
$s(t)$ 项用于拟合时序数据周度、月度和季度的周期性变化趋势；

$h(t)$ 项表示节假日或者突发事件对于时序数据的跳变性影响；

ε_t 项则表示噪声的随机扰动。

为更加直观地展示 Prophet 的运行逻辑，下面以铜期货主连合约价格进行初步尝试。下图显示，Prophet 能够有效地剥离时间序列的周度和年度季节性信号（当然有必要也可进行季度的周期性剥离），以及噪音干扰，在趋势的时间序列中，根据趋势进行后续预测。不难发现，在剥离周期性影响和噪声之后，铜期货主连价格长期来看呈现相对明显的趋势；而在年度周期性信息上，上半年的价格显著高于下半年，且临近年末，铜期货主连价格呈现上涨态势；在节假日影响下，我们在模型中加入了中国国内的节假日参数，而结果显示节假日对价格并无显著影响。

图表 4：Prophet 模型趋势分解

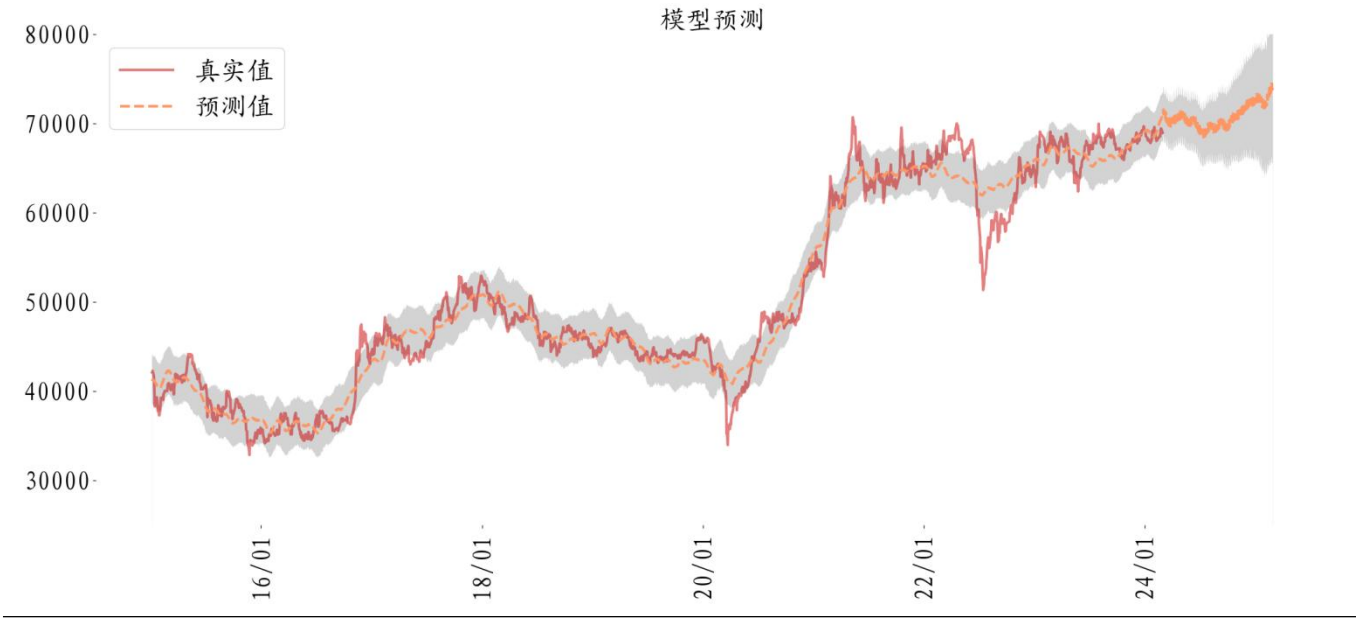


资料来源：东证衍生品研究院

我们进一步利用 Prophet 模型对铜期货主连价格进行预测。下图显示虚线为预测值，灰色区间为预测的上下界，而红色择时真实的价格走势。不难发现 Prophet 模型的预测值相对于真实值更为平滑，在趋势跟踪层面预测地相当不错，然后在短期的波动层面却无

法捕捉，这正是由于模型已经将季节性干扰和噪音干扰进行了剔除，只保留趋势部分，能够更好地捕捉价格在未来一段时间的大体趋势。

图表 5：Prophet 预测效果示意



资料来源：东证衍生品研究院

经过实践我们发现，Prophet 相较于传统的时序预测模型具有如下一些优势：

灵活性：无需对原始数据进行平稳性处理，同时可以对数据进行不同周期的趋势性分解；

可解释性：模型通过分解出趋势、周期、事件影响这三要素，能够直观反映出各部分对于数据的影响；

效率高：Prophet 算法运行效率较快，相较于机器学习模型的不断训练，Prophet 只要进行一次拟合即可。

4、优化模型

在本报告中，主要提出四个不同的优化方向，试图在精简因子数量的基础上，引入模式识别机制，达到更为高效、精准的信号生成。

优化 1：商品因子的精简

基于繁微庞大、完善的商品数据库，且均由相关商品分析师进行维护，我们可以非常高效地将大量的数据倾倒进模型当中，再通过模型的检验，自动给无效因子分配相对较低的权重，给有效因子分配较高的权重，以起到因子筛选的作用。然而这样的做法也会带

来弊端，一方面对于模型后期的逻辑解释性层面很难分辨，另一方面对于相关投资者的复现也存在一定的难度，故我们在这一部分引入商品基本面因子的筛选机制，通过事前对大量基本面因子进行单因子检验，之后再根据相关性进行筛选，使得每个品种的因子数量控制在 20 个-30 个之间。

优化 2：基本面数据的 Prophet 预测

相较于直接拿基本面因子去预测下一期的收益率，这部分我们尝试先采用 Prophet 模型对基本面数据本身进行预测，再以预测后的基本面数据去拟合未来的商品价格，相较于先前的预测方式，该模式能够更为有效地剥离基本面数据的周期性和噪音干扰，且模型自身会对数据进行异常值处理，进一步确保模型的有效性。作为对比，我们同时也会尝试基于商品价格序列自身进行自回归预测，比较两种预测方式的表现差异。

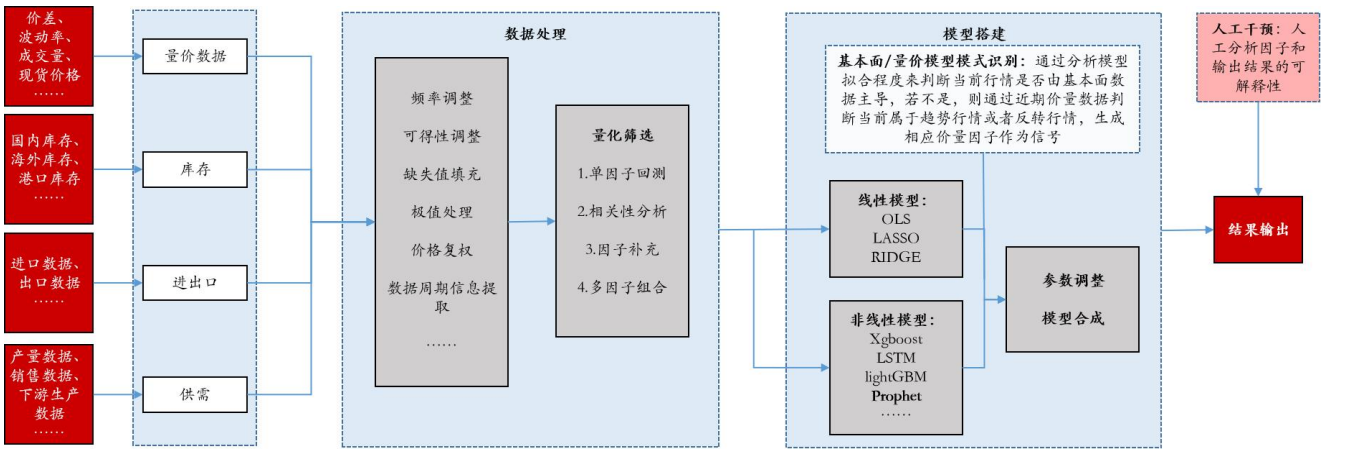
优化 3：信号生成的模式识别

在之前模型的运行过程中，我们发现市场上存在某些特定时间段，在该时间段内商品的价格走势与基本面的走势存在背离，若此时仍然由基本面量化模型生成信号，很容易导致长期的大幅回撤，为了避免这一现象，我们引入信号生成的模式识别机制。具体做法是在每一次生成信号前，根据基本面量化模型的拟合效果，判断当前市场行情能否由基本面数据所解释，若不是，则采用量价因子进行信号生成，由于短期内商品因子的趋势性信号较强，故此处我们用趋势信号进行替换。

优化 4：模型识别的 Prophet 优化

在上一个优化环节，一旦监测到当前的商品价格走势由价量数据主导，则以趋势因子作为信号生成。为了达到进一步优化，此处继续尝试以 Prophet 模型对商品价格本身进行自预测，再以预测值作为信号替换。

图表 6：基本面量化优化模型框架



资料来源：东证衍生品研究院

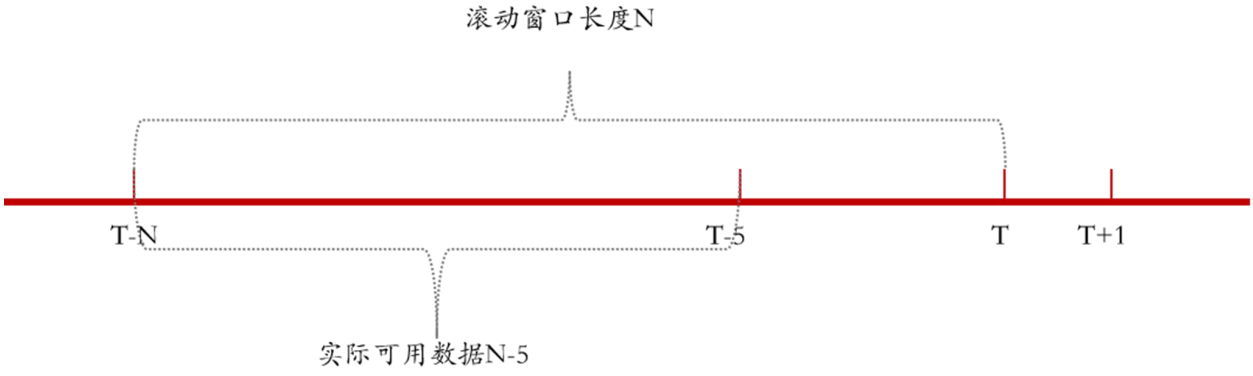
5、回测框架以及参数设置

本报告依据 OLS 进行滚动回归预测。OLS（普通最小二乘法）多元回归的原理为，最优拟合曲线使得各点到直线的距离的平方和（残差平方和 RSS）最小：

$$RSS = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t)^2$$

本报告采用滚动回归的方式进行回测，以避免使用未来数据。以周度预测为例，首先设置相应的滚动回归窗口长度 N，对每一天 T 都截取 T-N 到 T 时间段的基本面数据，由于为周度预测，需要对基本面数据进行 5 天的移动处理，所以实际可用数据点为 N-5 天内的数据，随后根据训练模型得到一列预测值，再根据预测值与真实值的比较去构造回测模型。

图表 7：滚动回归



资料来源：东证衍生品研究院

在滚动回归生成信号的基础上，增加模式识别机制，若判断当前行情处于量价模式下，则不采用滚工回归出来的结果作为信号，而是采用短期趋势因子，或者 Prophet 模型预测出来的结果作为信号。

数据处理层面，由于基本面数据相对量价数据并不完备，故需要进行一些列的数据处理的前期工作，具体做法如下：

频率调整：基本面的原始数据多为低频数据（月频或周频），为便于处理，将所有数据前值填充为日频数据；

标准化：对所有填充后的数据进行 z-score 标准化处理，提高数据之间的可比性；

异常值处理：对所有偏离均值 3 个标准差的数据进行处理；

移仓换月处理：为避免期货展期导致的价格影响，本报告以豆粕复权价格进行回测；

周期性影响：为剔除数据的周期性影响，对数据分别进行环比，同比等处理；

可得性处理：按指标具体可得性进行相应滞后处理；

回测参数：本报告基于日频数据生成周频级别的多空信号，每周更新仓位，手续费设置为双边万三。

为保证模型在样本内外的有效性，本报告在滚动建模的基础划分数据的样本内外。在因子筛选的步骤，数据采用样本内数据，而在最后模型搭建完成后，会利用样本外数据进行进一步验证。样本内数据为 2016 年 1 月至 2022 年 12 月，样本数据为 2023 年 1 月至 2024 年 2 月，即保留将近一年的样本外数据作为模型验证。

6、单因子检验

本报告首先以单品种铜作为示例，展示整个模型框架的搭建过程以及最后的预测结果，因为铜具有较为全面的基本面因子库，数据来源广泛且数据维护相对完善，且铜也是市场活跃度较高、关注度较高的品种之一，但并不代表此模型只适用于铜，我们的根本目的是构建一套具有普适性的商品基本面量化框架，在后期的研究任务中，我们也会基于此框架继续完善多品种策略的构建。

第一步，为了进行对因子数量的精简，首先需要对所有单因子进行单因子检验，本报告采用的检验方式为按照上文提到的回测方式对所有单因子进行回测建模，生成相应回测指标。

在单因子检验之前，我们首先根据因子的可得性进行一次初步筛选，即把上架时间较晚，或已经停更的数据进行剔除，同时为了保证数据频率与模型预测频率相匹配，我们将半年度和年度数据进行剔除，只保留日度、周度、月度（以及极个别季度）数据。在进行初步的筛选之后，基本面因子数据的数量为 126 个。

6.1、因子概述

126 个基本面因子分类如下，在铜自身上下游产业链相关指标和价量指标的基础上，我们同时也引入了宏观因子，其中铜原料和铜冶炼这两类因子占比较高，随后是铜库存类和金融指标类因子。具体各类指标数量如下表所示：

图表 8：基本面因子分类

指标分类	指标数量
铜原料	50
铜冶炼	35
铜库存	17
金融指标	15
铜终端	8
通胀	6
对外贸易及投资	6
铜价格数据	6
景气指标	4
铜交易情绪指标	3
债券市场指标	3
通胀高频	3
海外	2
货币金融	2
汇率	2
消费	1
工业	1
资金面	1
财政	1
高频数据	1

资料来源：东证衍生品研究院

下表对各类因子进行举例，铜原料端包括进口数据和产量数据，冶炼端数据包括精炼铜的进口和产量数据，库存端则涵盖了全球各地的库存和仓单数据，终端包括了下游的企业开工和需求数据，价格数据以各地的现货价格和升贴水数据为主，其余还包括景气指标、对外贸易、通胀指标等金融相关数据。

图表 9：基本面因子举例

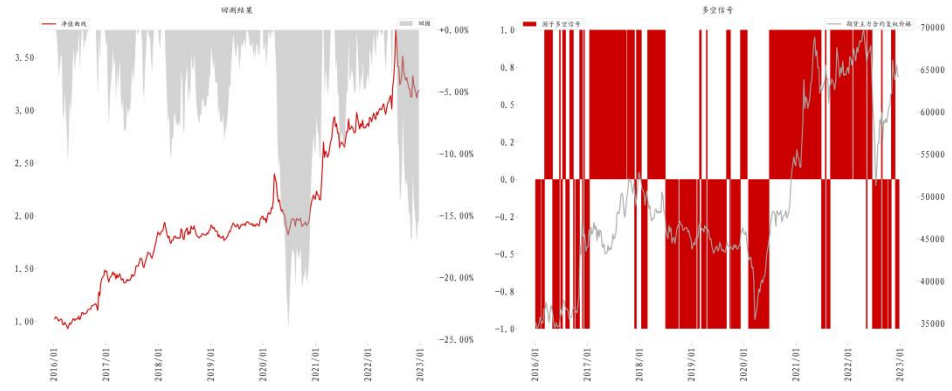
铜原料	铜冶炼	铜库存	铜终端	铜价格数据	景气指标	对外贸易及投资	通胀	其他分类
进口数量:铜矿石及精矿:当月值	进口数量:精炼铜:当月值	LME 铜:库存:合计:全球	中国电线电缆企业开工率	中国铜精矿现货RC（日度）：Cu26%	官方制造业 PMI	进出口贸易总值（美元）	PPI:生产资料:采掘工业:环比	铜：买持量：前 5 名会员合计
进口数量:废铜:当月值	进口数量:精炼铜:智利:当月值	LME 铜:注销仓单:合计:全球	产业在线：家用空调产量	中国地区铜现货升贴水：上海：平水铜	EPMI	FDI	CPI:环比	社会消费品零售总额:当月
智利铜矿产量（月度值）	进口数量：精炼铜：分贸易方式：进料加工贸易：终值	LME 铜:库存:西班牙:毕尔巴鄂	中汽协：汽车产量	中国地区铜现货升贴水：佛山：升水铜	BCI	分国家及地区(美元)	CPI:食品烟酒:畜肉类:猪肉:环比	工业增加值：当月
秘鲁铜矿产量（月度值）	ICSG 全球再生精炼铜产量（月度值）	库存小计:阴极铜:总计	中汽协：汽车经销商库存系数	中国地区铜现货升贴水：天津：平水铜	非制造业 PMI	主要商品(美元)	CPI:衣着:环比	当周日均销量:乘用车:厂家零售
墨西哥铜矿产量（月度值）	ICSG 澳大利亚精炼铜产量（月度值）	可用库容量:阴极铜	光电子器件产量	LME 铜升贴水(0-3)			CPI:交通和通信:环比	社融增量
ICSG 全球铜矿产能利用率（月度值）	ICSG 波兰精炼铜产量（月度值）	国内铜库存：广东	手机产量	进口铜：仓单溢价均价（上海）			CPI:教育文化和娱乐:环比	贷款规模

资料来源：东证衍生品研究院

6.2、因子表现

在这一部分，我们首先对与所有的单因子进行周期性因子扩充，即计算其在不同周期（月度、季度和年度）的比值和差分，以提取单因子在时序上的信息，再将这些衍生因子一并纳入因子库分别进行单因子的检验。以 DZ02017187（进口铜：仓单溢价均价（上海））为例，在原指标的基础上，我们分别纳入了 DZ02017187_mo_delta（仓单溢价均价的月度增量），DZ02017187_qt_delta（仓单溢价均价的季度增量），DZ02017187_yr_delta（仓单溢价均价的年度增量），DZ02017187_mo_ratio（仓单溢价均价的月度比值），DZ02017187_qt_ratio（仓单溢价均价的季度比值），DZ02017187_yr_ratio（仓单溢价均价的年度比值），扩充之后的因子库数量为 1134 个（162*7）。此处虽然在数量上大幅扩充了因子库，但在在后续的筛选过程后我们只会保留 20-30 个因子的数量，而且这些因子彼此之间必然存在一定的相关性，这一问题在建模的后续过程中我们也会进行相应处理。下列图表举例展示部分单因子的表现：

图表 10：因子 DZ02010751 回测曲线



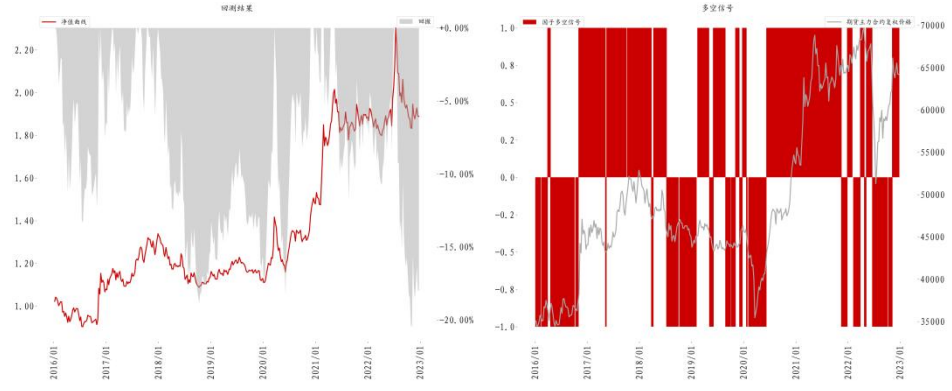
资料来源：东证衍生品研究院

图表 11：回测指标

总收益	217.93%
年化收益	17.74%
年化波动	17.64%
夏普值	0.95
最大回撤	-24.13%
收益风险比	0.74
胜率	0.54
盈亏比	1.30
sortino 比率	1.86

资料来源：东证衍生品研究院

图表 12：因子 DZ02016959 回测曲线



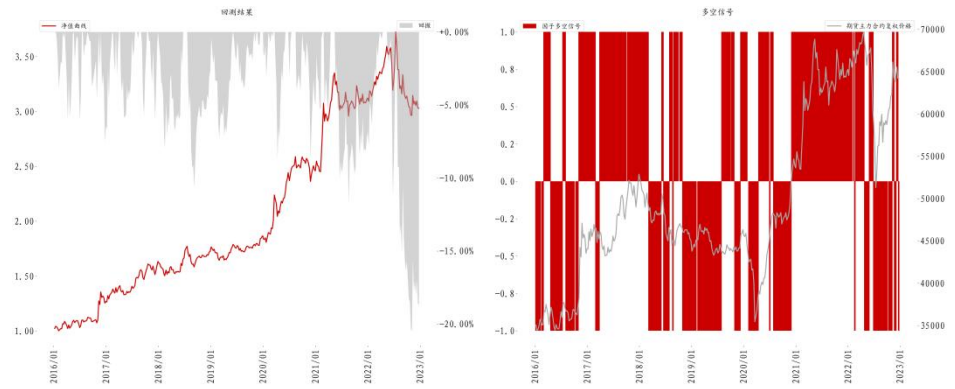
资料来源：东证衍生品研究院

图表 13：回测指标

总收益	88.86%
年化收益	9.39%
年化波动	17.82%
夏普值	0.47
最大回撤	-20.46%
收益风险比	0.46
胜率	0.51
盈亏比	1.24
sortino 比率	0.81

资料来源：东证衍生品研究院

图表 14：因子 DZ02050869 回测曲线



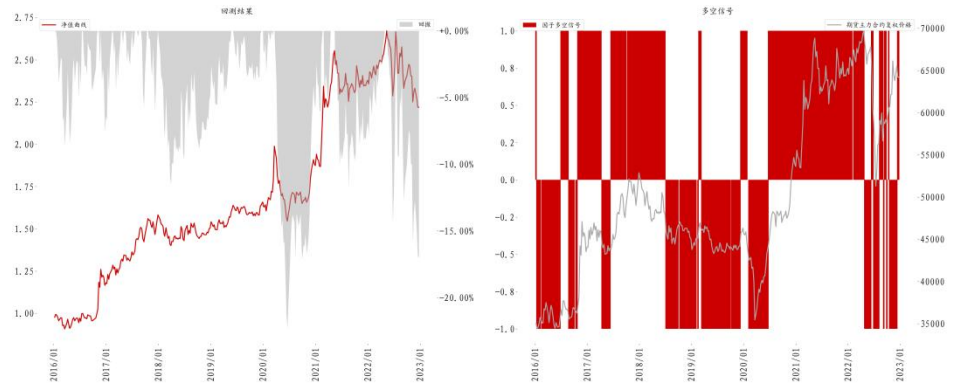
资料来源：东证衍生品研究院

图表 15：回测指标

总收益	203.00%
年化收益	16.94%
年化波动	17.68%
夏普值	0.90
最大回撤	-20.46%
收益风险比	0.83
胜率	0.53
盈亏比	1.30
sortino 比率	1.70

资料来源：东证衍生品研究院

图表 16：因子 DZ02032106 回测曲线



资料来源：东证衍生品研究院

图表 17：回测指标

总收益	121.88%
年化收益	11.91%
年化波动	17.77%
夏普值	0.61
最大回撤	-22.33%
收益风险比	0.53
胜率	0.51
盈亏比	1.30
sortino 比率	1.08

资料来源：东证衍生品研究院

上图以四个不同基本面单因子回测表现作为举例，分别是进口数量:铜矿石及精矿（DZ02010751）、进口数量：废铜：分国别：阿联酋-中国（DZ02016959）、中国:出口金额:服装及衣着附件（DZ02050869）、社会融资规模:新增未贴现银行承兑汇票（DZ02032106）。部分因子的表现相对不错，也就是基本面单因子本身具有一定的预测能力，但不足之处在于单因子预测的稳健性不足，部分回测曲线的最大回撤接近 30%，且波动较大。

下表展示按夏普值排序，表现较好的部分因子。除了部分原始因子外，相关衍生因子占比较大，且同一周期的比值和差分因子的表现几乎相同，比如 DZ02017187_mo_ratio 和 DZ02017187_mo_delta 表现一致，同样像 DZ02010847_mo_ratio 和 DZ02010847_mo_delta 也具有一致的表现，相同的例子还有很多，在后续操作中，我们会对这类因子根据相关性和表现进行去重。

在表现较好的因子类别中，各大类因子具有涉及，包括价格数据、冶炼数据、原料数据、库存数据以及金融相关指标，在因子筛选部分，我们尽可能保留较多的因子类别以达到因子类别的多样化。

单因子中表现最好的是仓单溢价均价（上海）的月度周期数据，其年化收益率为 27.54%，夏普值达 1.52，最大回撤为 13.27%。单看夏普值的话，该表现在可以接受的范围内，但由于单指标稳定性不佳，策略的最大回撤也较大，超过 13%，且由于是单因子预测，平仓持仓时间较长，很难及时应对短期的价格趋势变化。

图表 18：基本面因子回测表现（按夏普值降序）

	总收益	年化收益	夏普值	最大回撤	收益风险比	胜率	分类	因子描述(原始指标)
DZ02010842	460.12%	27.54%	1.52	-13.27%	2.07	56.62%	铜冶炼	进口数量:精炼铜:巴西:当月值
DZ02017187_mo_ratio	398.55%	25.46%	1.4	-18.46%	1.38	55.21%	铜价格数据	进口铜:仓单溢价均价（上海）
DZ02017187_mo_delta	398.55%	25.46%	1.4	-18.46%	1.38	55.21%	铜价格数据	进口铜:仓单溢价均价（上海）
DZ02017116_mo_ratio	358.68%	23.99%	1.31	-16.24%	1.48	55.49%	铜价格数据	中国铜精矿现货RC（日度）：Cu26%
DZ02017116_mo_delta	358.68%	23.99%	1.31	-16.24%	1.48	55.49%	铜价格数据	中国铜精矿现货RC（日度）：Cu26%
DZ02016910_qt_ratio	347.06%	23.54%	1.28	-16.31%	1.44	55.21%	铜原料	进口数量：铜精矿：分国别：赞比亚-中国：终值
DZ02016910_qt_delta	347.06%	23.54%	1.28	-16.31%	1.44	55.21%	铜原料	进口数量：铜精矿：分国别：赞比亚-中国：终值
DZ02010990	329.16%	22.83%	1.24	-18.20%	1.25	55.21%	铜库存	LME 铜:库存:韩国:光阳
DZ02010836_qt_ratio	310.84%	22.08%	1.2	-16.51%	1.34	54.08%	铜冶炼	进口数量:精炼铜:赞比亚:当月值
DZ02010836_qt_delta	310.84%	22.08%	1.2	-16.51%	1.34	54.08%	铜冶炼	进口数量:精炼铜:赞比亚:当月值
DZ02016904	304.31%	21.80%	1.18	-20.78%	1.05	57.18%	铜原料	进口数量：铜精矿：分国别：菲律宾-中国：终值
DZ02011021_mo_ratio	305.47%	21.85%	1.18	-18.46%	1.18	56.90%	铜库存	可用库容量:阴极铜

DZ02011021_mo_delta	305.47%	21.85%	1.18	-18.46%	1.18	56.90%	铜库存	可用库容量:阴极铜
DZ02017101_yr_ratio	305.64%	21.86%	1.18	-16.31%	1.34	54.08%	铜库存	国内铜库存:广东
DZ02017101_yr_delta	305.64%	21.86%	1.18	-16.31%	1.34	54.08%	铜库存	国内铜库存:广东
DZ02017011_qt_ratio	290.09%	21.19%	1.15	-18.46%	1.15	54.08%	铜冶炼	ICSG 澳大利亚精炼铜产量（月度值）
DZ02017011_qt_delta	290.09%	21.19%	1.15	-18.46%	1.15	54.08%	铜冶炼	ICSG 澳大利亚精炼铜产量（月度值）
DZ02025181_qt_ratio	287.50%	21.07%	1.14	-12.48%	1.69	54.37%	景气指标	PMI:新出口订单
DZ02025181_qt_delta	287.50%	21.07%	1.14	-12.48%	1.69	54.37%	景气指标	PMI:新出口订单
DZ02050871	284.74%	20.95%	1.13	-18.20%	1.15	51.55%	对外贸易及投资	中国:出口金额:成品油(海关口径):当月值
DZ02010847_mo_ratio	283.71%	20.91%	1.13	-11.52%	1.81	52.96%	铜冶炼	出口数量:精炼铜:韩国:当月值

资料来源：东证衍生品研究院

7、多因子模型

在进行单因子检验之后，我们按照因子表现、因子类别、因子之间的相关性进行筛选，将因子数量控制在 20 个至 30 个的区间范围内。具体做法为：首先在各大类内部对原始因子及其衍生因子进行相关性筛选，由于因子数量较多，我们设置相关性筛选阈值为 0.5，即若存在彼此之间相关性高于 0.5 的因子，则保留表现较好的因子，删除表现较差的因子，此处以夏普值作为因子表现好坏的度量；其次，设置各大类保留的因子数量，根据各大类因子整体表现好坏和大类内部原始因子的数量进行相应数量的设置，这部分我们尝试两套因子保留数量的方案，分别尝试保留 20 个和保留 30 个因子（方案 1 为 20 个，方案 2 为 30 个），考察其各自效果。

因子保留数量的选择很大程度上取决于主观的考量，当因子数量较少时，无法涵盖全部的因子大类，且预测精度相对单因子并无明显提升；而当因子数量过多时，则提升了模型过拟合的风险。此外，在我们之前的研究过程中，通过对因子 PCA 降维发现，在保留 95%的信息度时，降维后的因子数量基本维持在 20 到 30 个之间，故基于上述原因，我们认为因子的水量保留在 20 到 30 个是一个比较保险的选择，具体选取方案如下表：

图表 19：因子选取方案 1

方案 1（保留 20 个）	
指标分类	保留个数
铜原料	4
铜冶炼	3
铜库存	3
金融指标	2
铜终端	2
通胀	2
对外贸易及投资	1
铜价格数据	1
景气指标	1
铜交易情绪指标	1

资料来源：东证衍生品研究院

图表 20：因子选取方案 2

方案 2（保留 30 个）	
指标分类	保留个数
铜原料	5
铜冶炼	5
铜库存	4
金融指标	4
铜终端	3
通胀	2
对外贸易及投资	1
铜价格数据	1
景气指标	1
铜交易情绪指标	1
债券市场指标	1
通胀高频	1
海外	1

资料来源：东证衍生品研究院

7.1、多因子模型-方案 1

下表显示方案 1 中的指标选取，其中铜原料 4 个，铜冶炼和铜库存 3 个，金融指标、铜终端和通胀数据各 2 个，其余各 1 个。

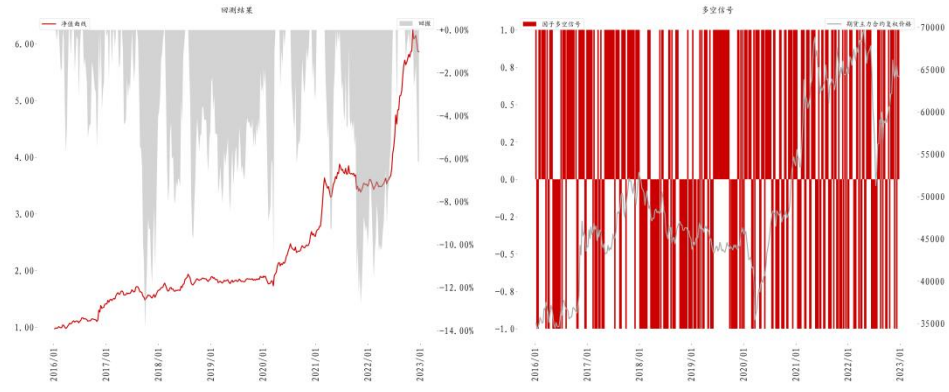
策略整体在收益端表现较好，年化收益达到 28.26%，夏普值达到 1.60，且策略在风险控制端的表现也相对稳定，其最大回撤为 13.91%，发生在 2022 年初这段时间，导致收益风险比达到 2.04。

图表 21：因子选取方案 1（20 个因子）

指标分类	指标代码	指标名称
铜原料	DZ02016904	进口数量：铜精矿：分国别：菲律宾-中国：终值
	DZ02016910_qt_delta	进口数量：铜精矿：分国别：赞比亚-中国：终值
	DZ02016924_mo_delta	智利铜矿产量（月度值）
	DZ02016934_qt_delta	进口数量:阳极铜:刚果（金）:当月值
铜冶炼	DZ02010836_qt_ratio	进口数量:精炼铜:赞比亚:当月值
	DZ02010842	进口数量:精炼铜:巴西:当月值
	DZ02017011_qt_delta	ICSG 澳大利亚精炼铜产量（月度值）
铜库存	DZ02010990	LME 铜:库存:韩国:光阳
	DZ02011021_mo_ratio	可用库容量:阴极铜
	DZ02017101_yr_ratio	国内铜库存：广东
金融指标	DZ02032075_yr_delta	金融机构:新增人民币贷款:票据融资:当月值
	DZ02032108_qt_ratio	社会融资规模:非金融企业境内股票融资:当月值
铜终端	DZ02017050	中国电线电缆企业开工率
	DZ02017078_yr_ratio	中汽协：汽车经销商库存系数
通胀	DZ02019963_qt_ratio	CPI:交通和通信:环比
	DZ02019970_mo_ratio	CPI:教育文化和娱乐:环比
景气指标	DZ02025181_qt_delta	PMI:新出口订单
对外贸易及投资	DZ02050871	中国:出口金额:成品油(海关口径):当月值
铜价格数据	DZ02017187_mo_delta	进口铜：仓单溢价均价（上海）
铜交易情绪指标	DZ02017169_qt_ratio	期货成交量:阴极铜

资料来源：东证衍生品研究院

图表 22：因子选取方案 1 回测表现



资料来源：东证衍生品研究院

图表 23：回测指标

总收益	486.15%
年化收益	28.36%
年化波动	17.10%
夏普值	1.60
最大回撤	-13.91%
收益风险比	2.04
胜率	0.60
盈亏比	1.27
sortino 比率	3.47
平均持仓时间	13.5

资料来源：东证衍生品研究院

7.2、多因子模型-方案 2

在方案 2 的因子选取中，因子数量达到 30 个，具体因子选取方案如下表：

其中铜原料和铜冶炼 5 个，铜库存、金融指标 4 个，铜终端 3 个，价格数据通胀数据 2 个，其余各 1 个，其中标粗部分为方案 2 相对于方案 1 新增的指标。

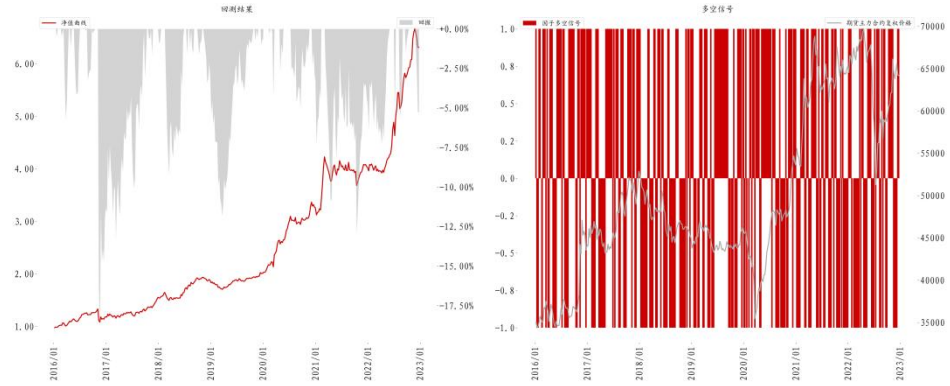
图表 24：因子选取方案 2（30 个因子）

指标分类	指标代码	指标名称
铜原料	DZ02016904	进口数量：铜精矿：分国别：菲律宾-中国：终值
	DZ02016910_qt_delta	进口数量：铜精矿：分国别：赞比亚-中国：终值
	DZ02016924_mo_delta	智利铜矿产量（月度值）
	DZ02016934_qt_delta	进口数量:阳极铜:刚果（金）:当月值
	DZ02010788_qt_delta	**进口数量:阳极铜:南非:当月值**
铜冶炼	DZ02010836_qt_ratio	进口数量:精炼铜:赞比亚:当月值
	DZ02010842	进口数量:精炼铜:巴西:当月值
	DZ02017011_qt_delta	ICSG 澳大利亚精炼铜产量（月度值）
	DZ02010847_mo_ratio	**铜冶炼**
	DZ02010847_yr_delta	**铜冶炼**
铜库存	DZ02010990	LME 铜:库存:韩国:光阳
	DZ02011021_mo_ratio	可用库容量:阴极铜
	DZ02017101_yr_ratio	国内铜库存：广东
	DZ02010990_qt_ratio	**LME 铜:库存:韩国:光阳**
金融指标	DZ02032075_yr_delta	金融机构:新增人民币贷款:票据融资:当月值
	DZ02032108_qt_ratio	社会融资规模:非金融企业境内股票融资:当月值
	DZ02032075	**金融机构:新增人民币贷款:票据融资:当月值**
	DZ02032093_mo_ratio	**金融机构:新增人民币存款:财政存款:当月值**
铜终端	DZ02017050	中国电线电缆企业开工率
	DZ02017078_yr_ratio	中汽协：汽车经销商库存系数
	DZ02017089_qt_ratio	**电子计算机产量**
铜价格数据	DZ02017187_mo_delta	进口铜：仓单溢价均价（上海）
	DZ02032198_qt_ratio	**平均批发价:7 种重点监测水果**
通胀	DZ02019963_qt_ratio	CPI:交通和通信:环比
	DZ02019970_mo_ratio	CPI:教育文化和娱乐:环比
景气指标	DZ02025181_qt_delta	PMI:新出口订单
对外贸易及投资	DZ02050871	中国:出口金额:成品油(海关口径):当月值
铜交易情绪指标	DZ02017169_qt_ratio	期货成交量:阴极铜
债券市场指标	**DZ02032149**	**国债到期收益率:1 年**
海外	**DZ02032190_yr_delta**	**美国:国债收益率:10 年**

资料来源：东证衍生品研究院

相较于方案 1，方案 2 的因子选取方式中，策略的收益表现小幅提升，年化收益达到 29.71%，夏普值为 1.67，在最大回撤层面，策略虽然在 2017 年初依旧经历了一波明显的回撤，最大回撤达到 18.89%，策略胜率为 61%。

图表 25：因子选取方案 2 回测表现



资料来源：东证衍生品研究院

图表 26：回测指标

总收益	531.26%
年化收益	29.71%
年化波动	17.17%
夏普值	1.67
最大回撤	-18.89%
收益风险比	1.57
胜率	0.61
盈亏比	1.23
sortino 比率	2.93
平均持仓时间	13.95

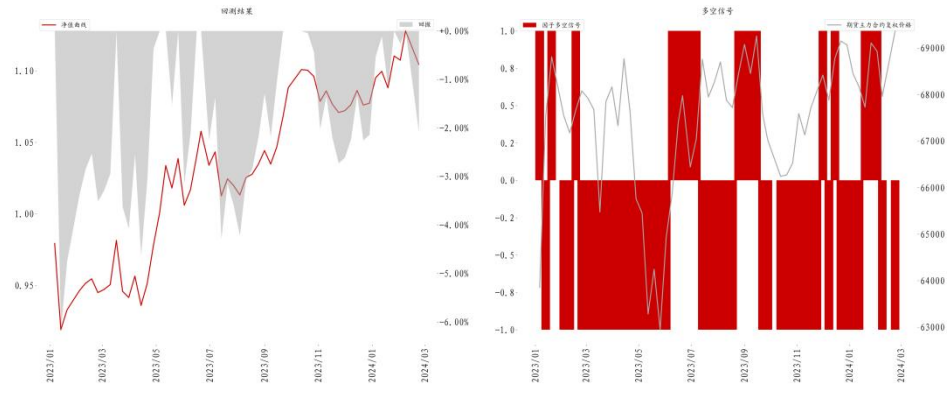
资料来源：东证衍生品研究院

综合对比方案 1 和方案 2 的因子选取方式，选择后续的建模以 30 个基本面因子（方案 2）作为基础。首先方案 2 的选取方式囊括了更多的因子大类（包括债券市场指标和海外数据），另外从收益端的角度来看，方案 2 的年化收益和夏普值都略高于方案 1。故我们选取方案 2 的因子选取方式作为后续优化的基础，即共选取 13 类，30 个基本面因子。

7.3、样本外验证

上文的比较均基于样本内回测，即 2016 年至 2023 年的回测结果，得到方案 2 的因子选取方式好于方案 1，由于该选取方式偏向主观筛选，为了进一步验证结果的一致性，需要在样本外进行进一步验证。前文提到过，我们保留了 2023 年至 2024 年这将近一年的数据作为样本外数据，我们将两种因子选取方式在样本外进行回测，得到的结果如下：

图表 27：因子选取方案 1 回测表现——样本外



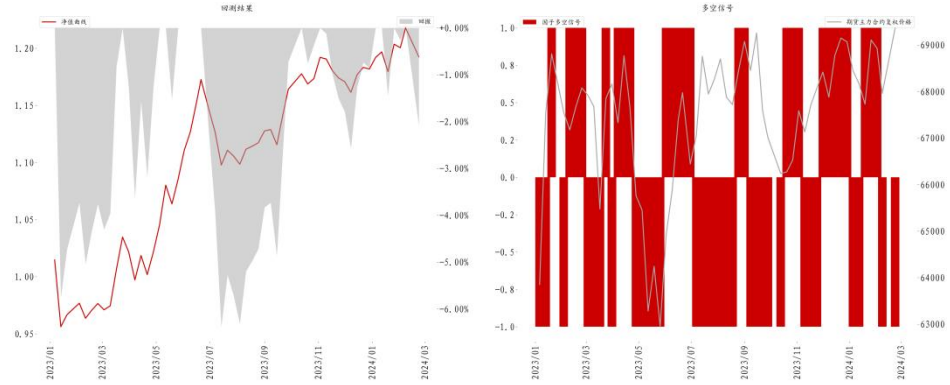
资料来源：东证衍生品研究院

图表 28：回测指标

总收益	10.44%
年化收益	9.18%
年化波动	12.65%
夏普值	0.65
最大回撤	-6.16%
收益风险比	1.49
胜率	0.61
盈亏比	0.82
sortino 比率	0.82
平均持仓时间	19.0

资料来源：东证衍生品研究院

图表 29：因子选取方案 2 回测表现——样本外



资料来源：东证衍生品研究院

图表 30：回测指标

总收益	19.25%
年化收益	16.84%
年化波动	11.94%
夏普值	1.33
最大回撤	-6.38%
收益风险比	2.64
胜率	0.60
盈亏比	1.09
sortino 比率	2.05
平均持仓时间	15.85

资料来源：东证衍生品研究院

结果显示，方案 1 和方案 2 的因子选取方式在样本外均存在一定程度的表现衰减，但依然在可接受的范围之内。方案 1 的夏普值从样本内的 1.60 下降至 0.65，方案 2 的夏普值从样本内的 1.67 下降至 1.33，根据样本外的结果显示，依然是方案 2 的因子选取方式具有更好的表现。这一现象一方面验证了模型在样本内外表现的一致性，另一方面也符合我们的推测：由于方案 2 的因子数量多于方案 1，能够使得模型的表现更为稳健，更具有一致性。当然这并不意味着因子数量越多越好，因子数量的不断增多势必会增加模型过拟合的风险，同时也提升了模型复现的难度，这与本报告的初衷相违背。综上所述，基于模型给出的分析，同时也基于我们过往的经验性判断，选择将因子数量控制在 30 个是一个相对稳妥的选择。

8、基于 Prophet 的基本面因子预测

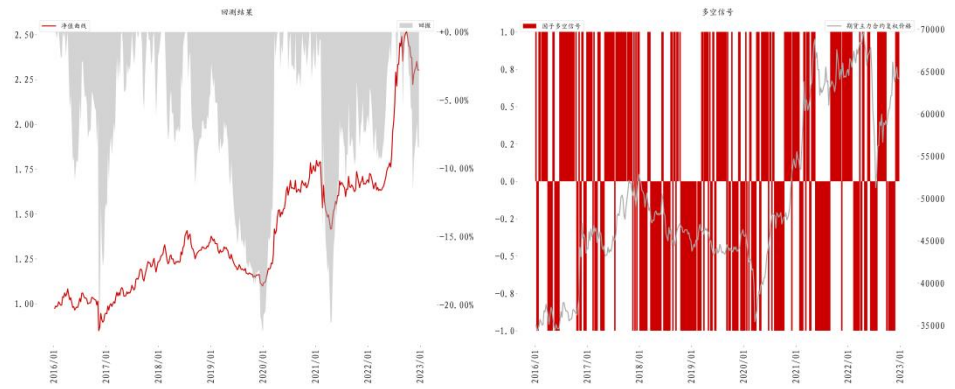
在前文部分，我们已经介绍过 Prophet 预测模型的优势，在这一章节，尝试三种不同方式基于 Prophet 模型生成预测信号，第一种是提取基本面数据的趋势信息，并以此作为基本面信号进行预测；第二种是以期货主连合约的收益率价格序列进行预测；第三种是直接以期货的主连价格序列本身进行趋势的自预测。

8.1、基于基本面数据趋势进行预测

在前文中，我们从原始基本面因子的梳理开始，一直进展到基本面因子的筛选。并且根据不同选择方式的比较，最终设置基本面因子数量为 30 个，再以 30 个因子进行多元线性回归，模型整体可以取得相对不错的效果，但由于基本面数据本身相对容易收到季节性因素的影响，我们在利用基本面数据进行预测时，只希望保留其趋势性的信息，而尽可能剔除周期性影响以及噪声扰动。Prophet 预测模型可以很好地满足该需求，在前述的预测过程中，我们通过当期的基本面数据去预测下一期的收益率情况。而在这一部分，我们对该环节进行修改：首先我们对基本面数据和价格数据统一进行剔除周期性影响的

处理；其次，我们先对基本面数据本身进行趋势预测，再以基本面数据的预测值对同一期的价格数据进行拟合；其次，我们尝试以期货的收益率序列进行预测；最后，我们不再以收益率作为预测目标，而是以价格本身作为预测目标，一方面由于已经对价格数据作了去周期性影响和噪音扰动的处理，另一方面也因为相较于预测收益率而言，预测价格可以拿到更多的预测序列 y 值（以周度预测为例，如果以收益率作为预测目标，训练时最新一周的 y 值以 nan 值填充；而如果以期货价格作为预测目标，训练时则可以拿到全部日期的价格数据）。

图表 31：回测表现



资料来源：东证衍生品研究院

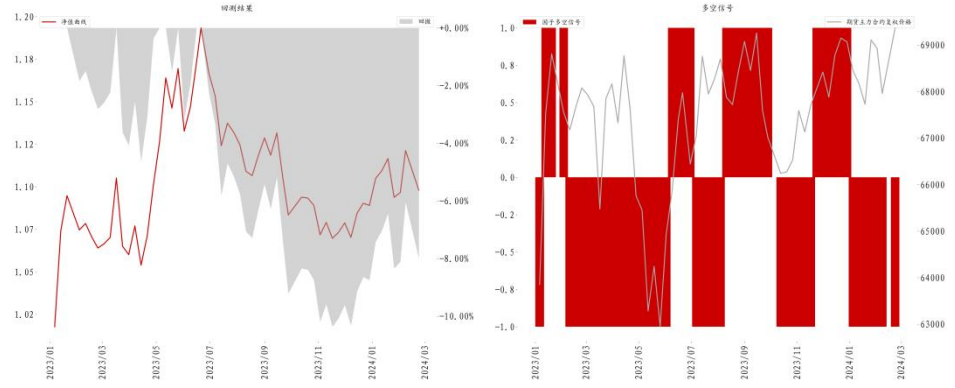
图表 32：回测指标

总收益	130.23%
年化收益	12.94%
年化波动	17.68%
夏普值	0.65
最大回撤	-21.89%
收益风险比	0.57
胜率	0.55
盈亏比	1.11
sortino 比率	0.97
平均持仓时间	20.29

资料来源：东证衍生品研究院

预测结果显示，胜率为 55%，年化收益 12.94%，最大回撤为 21.89%，策略在收益端效果明显，然而在风险控制层面表现并不尽如人意。我们推测模型在剥离基本面数据趋势数据的过程中，同时也舍去了部分细节信息，导致模型对短期内基本面数据的变化反应速度较慢，进而产生较幅度的回撤。

图表 33：回测表现——样本外



资料来源：东证衍生品研究院

图表 34：回测指标

总收益	9.80%
年化收益	8.62%
年化波动	12.54%
夏普值	0.61
最大回撤	-10.36%
收益风险比	0.83
胜率	0.54
盈亏比	1.08
sortino 比率	0.94
平均持仓时间	35.60

资料来源：东证衍生品研究院

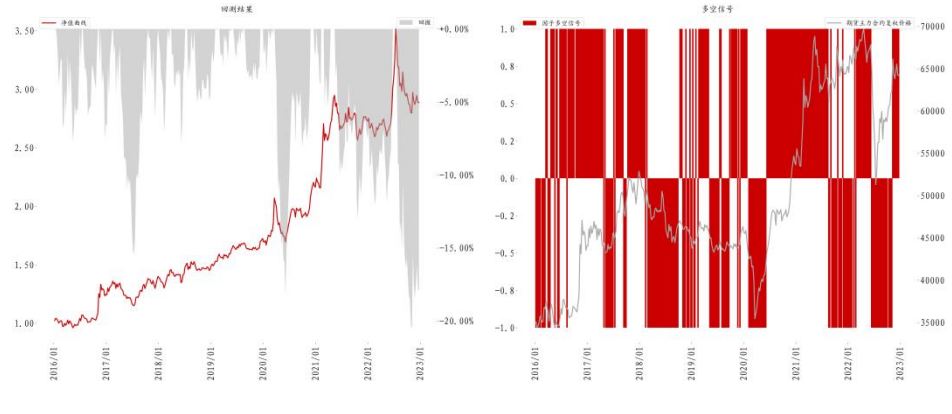
样本外的回测结果显示如上，同样表现出策略较大的回撤现象，在 2023 年 1 月到 2024 年 2 月的时间段，策略年化收益 8.62%，最大回撤 10.36%，结合信号图不难发现，在 2023 年 7 月之后，模型给出信号方向与价格走势出现较长时间的背离，其背后原因与上文的

推测一致，只保留基本面的趋势信息无法快速对价格走势的变化做出反应。

8.2、基于收益率序列进行 Prophet 自预测

在之前的基本面量化框架中，我们均以期货的收益率序列而非价格序列作为基本面数据回归的目标，因为收益率序列相对价格序列是一个更加平稳的时间序列，然而在 Prophet 预测模型中，由于模型自身会对数据进行周期性分解，我们本不必继续以收益率序列作为输入，但此处为了作为效果对比，依然采用收益率序列的方法进行结果展示。

图表 35：回测表现



资料来源：东证衍生品研究院

图表 36：回测指标

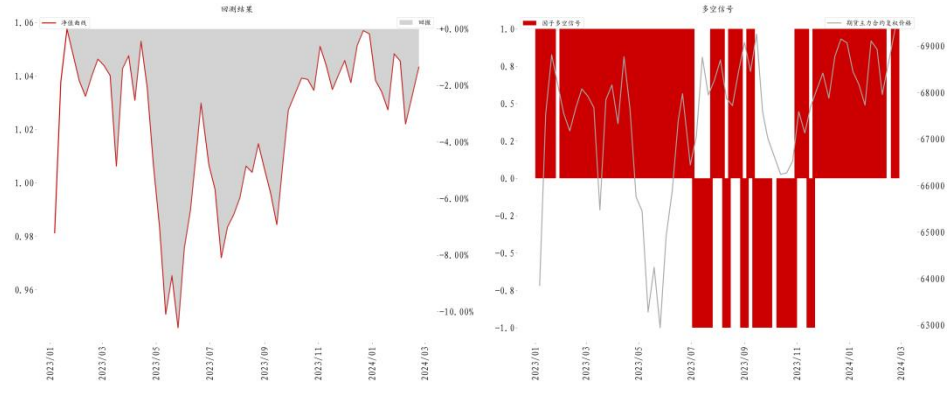
总收益	188.64%
年化收益	16.14%
年化波动	17.75%
夏普值	0.85
最大回撤	-20.46%
收益风险比	0.79
胜率	0.52
盈亏比	1.35
sortino 比率	1.64
平均持仓时间	33.70

资料来源：东证衍生品研究院

样本内表现如上，收益端表现相对优异，年化收益达 16.14%，而在风险控制的表现不佳，最大回撤达到 20.46%，使得其收益风险比只有 0.79%

而在样本外表现衰减幅度较大，具体结果如下，由于模型是根据趋势来进行预测的，而若数据是收益率序列时，会导致会在长期出现相同信号的风险，故无法真实反应模型的预测结果，使得模型在价格震荡幅度较大的行情下预测正确率显著下降。

图表 37：回测表现——样本外



资料来源：东证衍生品研究院

图表 38：回测指标

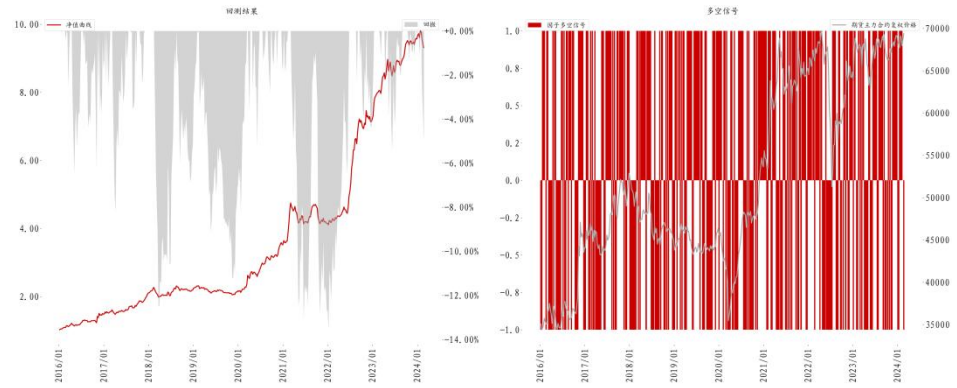
总收益	4.33%
年化收益	3.82%
年化波动	12.71%
夏普值	0.22
最大回撤	-10.57%
收益风险比	0.36
胜率	0.47
盈亏比	1.26
sortino 比率	0.21
平均持仓时间	28.5

资料来源：东证衍生品研究院

8.3、基于价格序列进行 Prophet 自预测

利用基本面信号趋势数据进行预测时，舍去其周期性影响时，也舍弃了部分的额外信息。在这一部分，我们继续尝试使用 Prophet 模型，对价格序列本身进行自预测。同前述的建模逻辑一样，Prophet 模型可以有效剔除价格时间序列在周期上的影响，只对价格序列的趋势进行预测。

图表 39：回测表现



资料来源：东证衍生品研究院

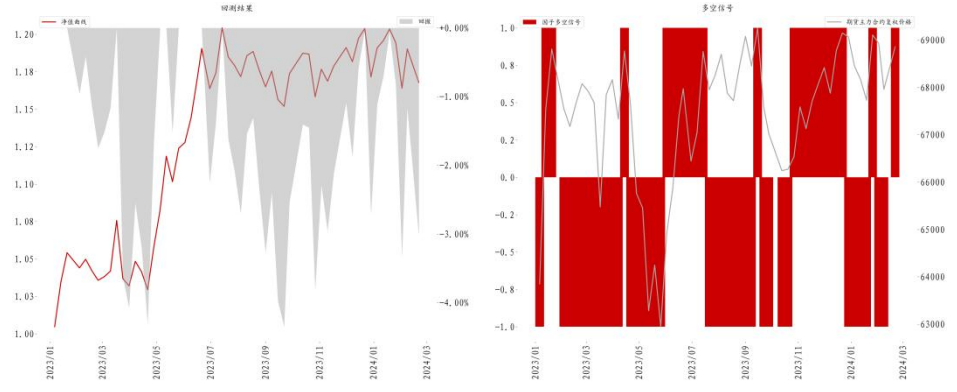
图表 40：回测指标

总收益	329.10%
年化收益	22.83%
年化波动	16.99%
夏普值	1.29
最大回撤	-13.64%
收益风险比	1.67
胜率	0.54
盈亏比	1.42
sortino 比率	2.89
平均持仓时间	20.05

资料来源：东证衍生品研究院

预测结果显示，胜率为 54%，同时年化收益达到 22.83%，最大回撤为 13.64%，整体表现优于基于基本面数据趋势进行预测。优于模型只纳入了价格数据本身的信息，能够更为直接地对价格数据的趋势进行反馈，相较于第一部分基于基本面数据趋势进行预测，直接基于价量信号本身进行自预测逻辑更为直接。基于上述的结果比较，我们在第六章节第二部分模式识别下的 Prophet 趋势因子优化中，选择采用基于价量信号本身进行自预测的方式对量价信号生成机制进行优化。

图表 41：回测表现——样本外



资料来源：东证衍生品研究院

图表 42：回测指标

总收益	16.79%
年化收益	14.71%
年化波动	11.32%
夏普值	1.21
最大回撤	-4.35%
收益风险比	3.38
胜率	0.58
盈亏比	1.12
sortino 比率	2.06
平均持仓时间	21.9

资料来源：东证衍生品研究院

样本外模型表现存在一定程度衰减，但在可接受的范围内，夏普值从 1.29 下降至 1.21，

仍然存在较强的盈利能力，表明模型的预测结果具有样本内外的一致性。

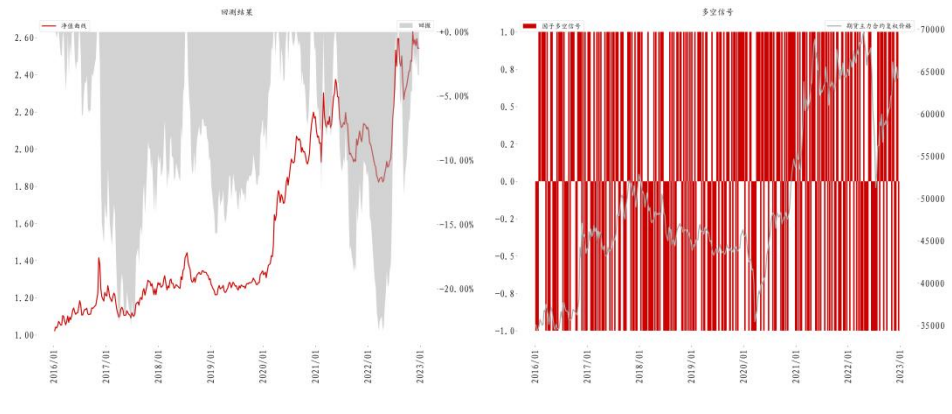
9、模式识别

在我们过往对于模型效果的分析中发现，存在某一时段，在该时段会出现期货价格走势与其基本面数据走势背离的情况，而往往这样的时段会维持相当长的一段时间。这种情况可能是由政策性时间所影响，也有可能是被市场情绪所主导。总之，不管是出于什么样的原因，此时市场行情无法由基本面数据所解释。此类行情的出现往往会对基本面量化模型造成相对长期且幅度较大的回撤，为了避免该风险，本报告引入模式识别的机制。在每一个进行预测时，我们首先用近期的数据进行拟合，来判断当前商品的价格趋势是否已经偏离基本面走势，若检测到偏离现象发生，则由基本面量化模型的信号生成机制（基本面模式）转化到量价因子的信号生成机制（量价模式）。关于量价模式下的信号生成机制，在接下来的部分我们作深入讨论。

9.1、动量趋势信号和 Prophet 趋势信号的对比

在引入模式识别机制之前，我们首先需要对比两种量价信号生成模型的表现。第一种方式直接采用短期动量趋势作为量价信号，第二种参考 8.3 章节中的方式，基于价量信号价格序列进行自预测。第二种方式的回测表现见 8.3，其样本内表现为，年化收益达到 22.83%，最大回撤为 13.64%，胜率为 54%，夏普值 1.29；样本外表现为，年化收益达到 14.71%，最大回撤为 4.35%，胜率为 58%，夏普值 1.21。而以短期动量趋势作为量价信号的表现如下样本内表现为，年化收益达到 14.08%，最大回撤为 23.27%，胜率为 52%，夏普值 0.74；样本外表现为，年化收益达到 12.17%，最大回撤为 11.30%，胜率为 61%，夏普值 0.89。作为商品时序单因子而言，短期趋势因子具有一定的预测能力，然而其表现相较于上文的基于价量信号价格序列进行 Prophet 自预测的表现仍然存在一定劣势。

图表 43：回测表现



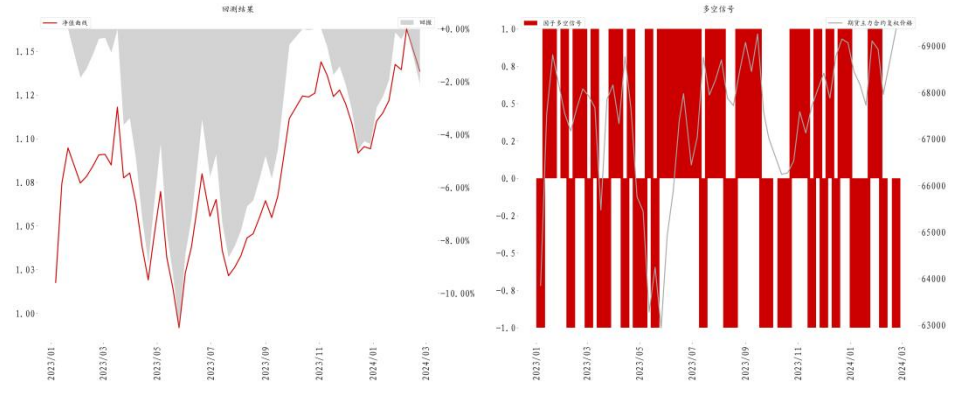
资料来源：东证衍生品研究院

图表 44：回测指标

总收益	154.30%
年化收益	14.08%
年化波动	17.78%
夏普值	0.74
最大回撤	-23.27%
收益风险比	0.61
胜率	0.52
盈亏比	1.29
sortino 比率	1.37
平均持仓时间	9.39

资料来源：东证衍生品研究院

图表 45：回测表现——样本外



资料来源：东证衍生品研究院

图表 46：回测指标

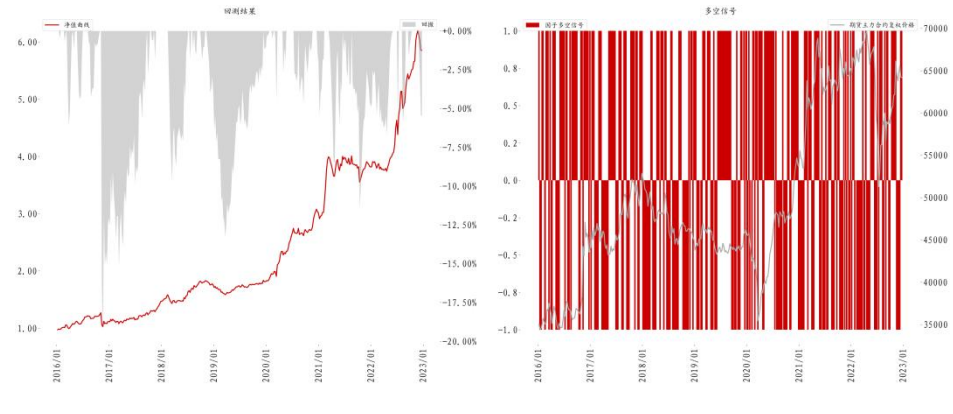
总收益	13.87%
年化收益	12.17%
年化波动	12.53%
夏普值	0.89
最大回撤	-11.30%
收益风险比	1.08
胜率	0.61
盈亏比	0.89
sortino 比率	1.43
平均持仓时间	10.2

资料来源：东证衍生品研究院

9.2、模式识别下的趋势量价信号生成

在每一次进行信号生成前，先以最近一周的日度价格数据与预测值进行预测精度分析，采用 R 方作为度量。若 R 方较高，说明基本面量化模型拟合结果相对精确，则判断为基本面模式，继续采用基本面模型生成相应信号；若 R 方偏低，表明基本面量化模型拟合结果效果不佳，此时切换为量价模式。在量价模式下，我们首先尝试利用趋势因子进行信号生成。由于本报告中的预测周期为周度预测，周期偏短，趋势因子相对有效，故做此选择。

图表 47：回测表现



资料来源：东证衍生品研究院

图表 48：回测指标

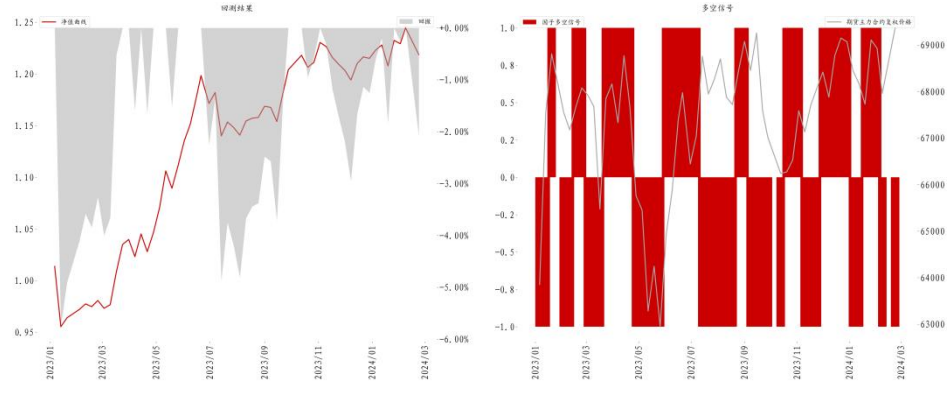
总收益	485.31%
年化收益	28.33%
年化波动	17.01%
夏普值	1.61
最大回撤	-19.25%
收益风险比	1.47
胜率	0.61
盈亏比	1.19
sortino 比率	2.75
平均持仓时间	13.3

资料来源：东证衍生品研究院

根据样本内表现分析可得，在采用模式识别下的趋势量价信号生成的方式后，模型整体表现并无提升，样本内表现为，年化收益 28.33%，夏普值为 1.61，最大回撤为 19.25%。此现象背后的原因可能有两个，第一，在量价模型下，采用趋势因子作为信号的表现不

如直接以基本面因子回归出来的结果好；第二，模式识别的机制并不准确，即无法简单通过判断回归方差的 R 方大小来说明当前处于基本面模式还是量价模式。为了验证上述的两个推测，我们在 9.3 尝试模式识别下的 Prophet 趋势因子优化，通过观察其表现的好坏，可以判断出上述两个推测是否正确。

图表 49：回测表现——样本外



资料来源：东证衍生品研究院

图表 50：回测指标

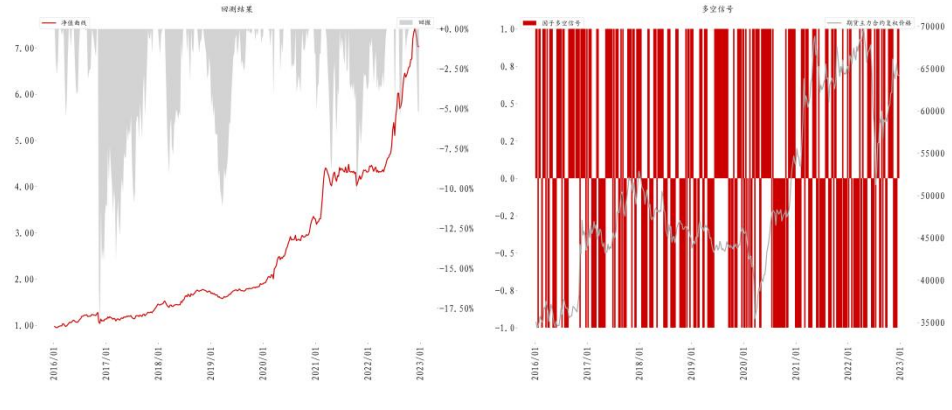
总收益	21.87%
年化收益	19.11%
年化波动	11.67%
夏普值	1.55
最大回撤	-5.76%
收益风险比	3.32
胜率	0.61
盈亏比	1.09
sortino 比率	2.35
平均持仓时间	17.8

资料来源：东证衍生品研究院

9.3、模式识别下的 Prophet 趋势因子优化

除了利用量价因子进行量价模型下的信号生成之外，基于我们在第 8 部分的讨论，我们尝试在量价模型下，直接利用 Prophet 模型生成趋势信号。

图表 51：回测表现



资料来源：东证衍生品研究院

图表 52：回测指标

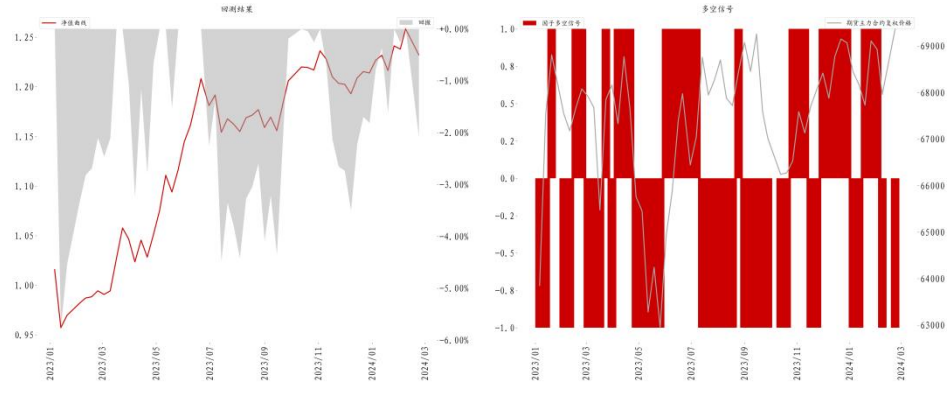
总收益	602.78%
年化收益	31.69%
年化波动	17.13%
夏普值	1.79
最大回撤	-18.71%
收益风险比	1.69
胜率	0.60
盈亏比	1.32
sortino 比率	3.18
平均持仓时间	14.9

资料来源：东证衍生品研究院

结果显示，量价模式下利用 Prophet 预测模型进行修正之后，模型的回测结果表现小幅提升，夏普值达到 1.79，最大回撤控制在 18.71%。经过比较，我们认为在进行周度预测任务时，Prophet 预测模型的胜率相对传统趋势因子小幅提升，且运算速度相对较快，

可作为传统量价因子的补充或替换。这一结果也解答了我们在 9.2 部分的两个推测，数据表明在模式识别下的 Prophet 趋势因子优化能够相较于原模型进行一定程度的提升，也即说明模式识别的机制是具有一定的有效性的。

图表 53：回测表现——样本外



资料来源：东证衍生品研究院

图表 54：回测指标

总收益	23.23%
年化收益	20.28%
年化波动	12.07%
夏普值	1.60
最大回撤	-5.76%
收益风险比	3.53
胜率	0.60
盈亏比	1.18
sortino 比率	2.5
平均持仓时间	15.85

资料来源：东证衍生品研究院

为了进一步验证，同样进行样本外的回测，回测结果显示，模型在样本外的表现为，年化收益率为 20.28%，夏普值为 1.60，最大回撤 5.76%，胜率达 60%，与模型在样本内的表现具有一致性。

10、模型应用与展望

本报告在先前基本面量化模型的基础进行了多个层面的优化：

首先，对但品种的商品基本面因子进行精简，确保模型具有较强可解释性且已被复制及跟踪；

其次，尝试利用 Prophet 模型剔除基本面数据的周期性影响，同时也可基于价格数据对价格进行自预测；

最后，引入模式识别机制，在价格走势与基本面数据走势背离时，采用量价因子进行信号生成。

本报告不仅可以为量化客户提供新的建模思路，同时为产业客户提供了单品种价格预测的参考。

优化模型（指 9.3 部分的模式识别下的 Prophet 趋势因子优化）能够相较于原模型提供收益端的提升，然而在风险控制端，优化模型相较于原模型并吴明显提升。由于本报告采用的模式识别机制相对简单的粗糙，故无法做到对模式的完美识别，所以模型目前还不能有效规避基本面与价格走势偏离的行情，这是目前该优化模型存在的不足之处。

10.1、模型的应用

在我们早先的研究服务过程中，不仅协助多家量化投资机构搭建相关模型，同时也帮助

多家非量化的产业客户提供全套的建模支持，在服务过程中必然会遇到一系列问题，比如数据可得性和模型精度不够，我们也因此在跟踪模型表现的同时不断进行修正，本篇报告可视为对早前基本面量化模型的一次大规模修正，在此后的服务过程中，我们会在新的建模框架的基础上，进行多维度的策略研发。不仅有多品种策略化的模型搭建，也会深耕单品种长周期的趋势性预测。进一步的，根据我们之前服务实体产业的经验，相关产业客户无论是处于套保的需求还是投机的需求，对于单品种的价格趋势预测具有较大的需求，报告中提到的框架可以很好地满足这方面的需求。

10.2、多品种策略的研发

在单品种基本面量化建模框架的基础上，我们同时也会进行多品种策略的研发。我们对多品种策略的要求是高夏普、低回撤、可跟踪、可复现。先前的全品种基本面量化策略我们已经跟踪了近两年的表现，策略整体净值曲线相对稳健，我们希望在此基础上，对模型表现进一步优化。

期货走势评级体系（以收盘价的变动幅度为判断标准）

走势评级	短期（1-3 个月）	中期（3-6 个月）	长期（6-12 个月）
强烈看涨	上涨 15%以上	上涨 15%以上	上涨 15%以上
看涨	上涨 5-15%	上涨 5-15%	上涨 5-15%
震荡	振幅-5%-+5%	振幅-5%-+5%	振幅-5%-+5%
看跌	下跌 5-15%	下跌 5-15%	下跌 5-15%
强烈看跌	下跌 15%以上	下跌 15%以上	下跌 15%以上

上海东证期货有限公司

上海东证期货有限公司成立于 2008 年,是一家经中国证券监督管理委员会批准的经营期货业务的综合性公司。东证期货是东方证券股份有限公司全资子公司。公司主要从事商品期货经纪、金融期货经纪、期货交易咨询、资产管理、基金销售等业务，拥有上海期货交易所、大连商品交易所、郑州商品交易所、上海国际能源交易中心和广州期货交易所会员资格，是中国金融期货交易所全面结算会员。公司拥有东证润和资本管理有限公司，上海东祺投资管理有限公司和东证期货国际（新加坡）私人有限公司三家全资子公司。

自成立以来，东证期货秉承稳健经营、创新发展的宗旨，坚持以金融科技助力衍生品发展为主线，通过大数据、云计算、人工智能、区块链等金融科技手段打造研究和技术两大核心竞争力，坚持市场化、国际化、集团化发展方向，朝着建设一流衍生品服务商的目标继续前行。

免责声明

本报告由上海东证期货有限公司（以下简称“本公司”）制作及发布。

本公司已取得期货投资咨询业务资格，投资咨询业务资格：证监许可【2011】1454号。

本研究报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本研究报告是基于本公司认为可靠的且目前已公开的信息撰写，本公司力求但不保证该信息的准确性和完整性，客户也不应该认为该信息是准确和完整的。同时，本公司不保证文中观点或陈述不会发生任何变更，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。本公司会适时更新我们的研究，但可能会因某些规定而无法做到。除了一些定期出版的报告之外，绝大多数研究报告是在分析师认为适当的时候不定期地发布。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，若有必要应寻求专家意见。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买投资标的的邀请或向人作出邀请。

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引致的任何损失负任何责任，投资者需自行承担风险。

本报告主要以电子版形式分发，间或也会辅以印刷品形式分发，所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。

如需引用、刊发或转载本报告，需注明出处为东证衍生品研究院，且不得对本报告进行任何有悖原意的引用、删节和修改。

东证衍生品研究院

地 址：上海市中山南路 318 号东方国际金融广场 2 号楼 21 楼
联系人：梁爽
电 话：8621-63325888-1592
传 真：8621-33315862
网 址：www.orientfutures.com
Email：research@orientfutures.com