



Research and
Development Center

深度学习揭秘系列之三： 用 DeepSeek 优化价量因子

2025 年 03 月 21 日

于明明 金融工程与金融产品首席分析师
执业编号：S1500521070001
联系电话：+86 18616021459
邮 箱：yumingming@cindasc.com

周金铭 金融工程与金融产品分析师
执业编号：S1500523050003
联系电话：+86 18511558803
邮 箱：zhoujinming@cindasc.com

证券研究报告

金工研究

金工专题报告

于明明 金融工程与金融产品

首席分析师

执业编号：S1500521070001

联系电话：+86 18616021459

邮箱：yumming@cindasc.com

周金铭 金融工程与金融产品

分析师

执业编号：S1500523050003

联系电话：+86 18511558803

邮箱：zhoujinming@cindasc.com

信达证券股份有限公司

CINDA SECURITIES CO., LTD

北京市西城区宣武门西大街甲 127 号金隅

大厦B座

邮编：100031

深度学习揭秘系列之三：用 DeepSeek 优化 价量因子

2025 年 03 月 21 日

- **本文是深度学习揭秘系列之三。**聚焦于借助 DeepSeek 模型对选股因子进行生成与改进，在量化投资领域展开深入探索。研究以 Qlib 集成的 Alpha158 量价因子为基础，该因子集涵盖日内、波动、价、量及量价相关性五类因子。通过设定统一测算口径，运用特定的 Prompt Engineering 和 AI 交互流程，借助 DeepSeek 对原始因子进行优化，以及生成相关性较低的新因子，取得显著成果。
- **在因子优化方面，多数因子经 DeepSeek 优化后预测能力提升显著。**测试的 Alpha158 因子集中，75%的因子 RankIC 均值提升，50%的因子 RankIC 均值达 1.2 倍提升，35%的因子 RankIC 均值有 1.5 倍提升；ICIR 指标同样向好，众多因子 ICIR 提升且多倍提升的因子数量可观。从不同窗口期数据看，优化后的因子表达式普适性强。以波动率因子 std20 为例，多次改进中引入平均真实波幅 ATR 概念、成交量加权机制、EMA 双重平滑及四维波动极值捕捉等，虽部分改进使 RankIC 均值有波动，但整体提升了因子选股效果。不过，部分因子如 Beta20、min20 等受限于原始计算逻辑，在优化过程中未呈现显著提升，揭示传统因子改进存在理论天花板。
- **在因子生成上，从零生成因子较难达到理想效果，然而，站在成功案例基础上生成因子，效果显著。**若从零开始，20 次迭代生成的因子具备一定逻辑，但预测效果一般，ICIR 未达预期。若给予 DeepSeek 部分预测效果尚佳的因子表达式作为参考，例如 Alpha158 原始及优化因子表达式与对应的 IC 统计量，则能够在较少的迭代次数内，生成 5 个 ICIR 在 0.8 以上且与样例因子相关性低的新因子。其中第一个有效因子通过捕捉量价协同增强效应，第二个聚焦量价共振强度维度，展现出较好的选股能力。
- **组合维度对比发现，将优化后的 Alpha158 因子及新生成因子线性结合，使用 Lasso 模型合成因子，能提升选股能力。**全 A 数据中，原始因子、增强因子、原始 + 生成因子及最终合成因子的 RankIC 均值和多头超额收益逐次提升。落地到中证 800 指增组合，原始因子叠加优化与新生成因子得到的复合因子，月频 RankIC 均值从 9.01%提升至 10%，ICIR 从 0.93 提升至 1.01，多头超额年化收益从 7.05%提升至 7.92%，年化信息比从 1.63 增强至 1.89。
- **综上，本文引入 DeepSeek 模型，通过构建“优化 - 验证 - 再迭代”框架，**贡献了因子优化与因子生成的新思路，实现了大语言 AI 模型对量化研究的赋能。

风险因素：结论基于历史数据，在市场环境转变时模型存在失效的风险。



目录

深度学习揭秘系列之三：用 DeepSeek 优化价量因子	4
一、Alpha158 因子体系解析与优化框架	5
1.1、Qlib 算子与 Alpha158 因子集概述	5
1.2、动态交互框架：基于 Prompt Engineering 的因子优化流程设计	8
1.3、优化效果验证：RankIC 的跨周期普适性与稳定性突破	12
1.4、以波动率因子为例：洞察模型优化方向	16
二、从优化到创造：基于大语言模型的新因子生成范式	21
2.1、从零生成的困境：独立探索下的效率与效果瓶颈	21
2.2、案例驱动的突破：基于先验知识的低相关性因子挖掘	24
三、策略应用与组合绩效验证：复合因子在指增组合中的增益	30
3.1、复合因子的收益增强：中证 800 中 RankIC 均值提升至 11.03%	30
3.2、指增组合的实战检验：月度 800 指增超额提升近 1%	34
风险因素	37

表目录

表 1：Alpha 158 因子列表	5
表 2：Qlib 部分算子	7
表 3：原始因子与改进因子 RankIC 均值和 ICIR	12
表 4：不同因子类型下原始表达式与增强表达式 RankIC 均值	15
表 5：Beta20 原始因子与历次优化因子表达式及 IC 统计量	19
表 6：min20 原始因子与历次优化因子表达式及 IC 统计量	19
表 7：qtl20 原始因子与历次优化因子表达式及 IC 统计量	19
表 8：历史模型生成的因子表达式与 IC 统计量	22
表 9：从 Alpha158 生成的因子表达式与 IC 统计量	25
表 10：新因子与 ICIR 大于 0.7 的样例因子之间的相关性	27
表 11：新因子之间相关性	27
表 12：原始合成因子与增强合成因子 IC 统计量	30
表 13：原始合成因子与增强合成因子多头超额与收益波动比	31
表 14：原始合成因子与原始+生成合成因子 IC 统计量	31
表 15：原始因子与原始+生成因子多头超额与收益波动比	32
表 16：各因子 IC 统计量	32
表 17：各因子多头超额与收益波动比	33
表 18：原始因子与复合因子在中证 800 中 IC 统计量	34
表 19：原始因子多头组合收益波动绩效指标	35
表 20：原始因子组合多头超额收益波动绩效指标	35
表 21：复合因子多头组合收益波动绩效指标	35
表 22：复合因子组合多头超额收益波动绩效指标	35

图目录

图 1：因子优化流程示意图	9
图 2：std20 原始因子与第一次改进因子累计 RankIC	17
图 3：std20 原始因子与第一次改进因子多头超额净值	17
图 4：std20 原始因子与第一二次改进因子累计 RankIC	18
图 5：std20 原始因子与第一二次改进因子多头超额净值	18
图 6：std20 原始因子与第一二三次改进因子累计 RankIC	19
图 7：std20 原始因子与第一二三次改进因子多头超额净值	19
图 8：新生成因子累计 RankIC	28
图 9：新生成因子多头超额净值	28
图 11：原始合成因子与增强合成因子累计 RankIC	30
图 12：原始合成因子与增强合成因子多头超额收益	31
图 13：原始合成因子与原始+生成合成因子累计 RankIC	31
图 14：原始合成因子与原始+生成合成因子多头超额收益	32
图 15：各因子累计 RankIC	32
图 16：各因子多头超额收益	33
图 17：原始因子与复合因子在中证 800 中累计 RankIC	34
图 18：原始因子与复合因子多头超额净值	35

深度学习揭秘系列之三：用 DeepSeek 优化价量因子

近年来，深度学习技术凭借其强大的非线性建模能力，逐步成为量化投资领域的重要工具。在前期系列研究中，我们通过《深度学习揭秘系列之一：基于量价与基本面结合的深度学习选股策略》系统对比了线性模型(Lasso)与基础非线性模型(MLP)在不同因子输入场景下的表现，揭示了量价因子与基本面因子结合的有效性。在《深度学习揭秘系列之二：涵盖量价与基本面因子的多模型结合神经网络》报告中，引入了时序神经网络，并探讨了适用于时序神经网络的因子类型以及时序神经网络的参数敏感性，最后打造了输入端覆盖低频价量、高频价量、基本面因子与行情数据，包含全连接与 GRU 神经网络的多模型组合。

本报告作为《深度学习揭秘》系列的第三篇章，开创性地引入大语言模型 DeepSeek，构建“优化 - 验证 - 再迭代”的全流程研究框架，试图为量化投资的因子工程提供新的方法论突破。

本文聚焦于两大核心任务：其一，以 Alpha158 因子为蓝本，通过 Prompt Engineering 与 DeepSeek 的交互优化，提升因子选股能力；其二，突破传统因子库限制，探索基于大语言模型的新因子生成范式，其中：

- **因子优化流程的重构：**设计“优化-验证-再迭代”的动态框架，通过多轮次的因子表达式改进，显著提升 RankIC 与 ICIR 指标。例如，针对波动率因子 std20，DeepSeek 通过引入 ATR 概念、成交量加权机制与 EMA 双重平滑，使其 RankIC 均值从 4.03% 提升至 7.8%，ICIR 从 0.31 跃升至 0.55；
- **新因子生成范式的突破：**在零基础生成效果有限的情况下，通过引入成功案例（如优化后的 Alpha158 因子）作为先验知识，DeepSeek 成功生成 5 个 ICIR 超 0.8 且低相关性的新因子。例如，首个有效因子通过捕捉量价协同增强效应，构建“上涨日量价动量-下跌日量价惩罚”的双向驱动逻辑，ICIR 达 0.83；
- **800 指增组合实证：**将原始因子、优化因子与新生成因子输入 Lasso 模型进行线性合成，中证 800 成分股范围内复合因子的周频 RankIC 均值达 11.41%，较原始 Alpha158 合成因子提升 0.81pct。中证 800 指增组合测试显示，复合因子年化超额收益从 7.05% 提升至 7.92%，信息比从 1.63 优化至 1.89，验证了方法论的有效性。

后续章节将围绕上述核心问题展开论证，通过数据回溯、案例解读与组合测试，揭示以 DeepSeek 为代表的大语言模型与传统多因子体系的融合潜力。

一、Alpha158 因子体系解析与优化框架

1.1、Qlib 算子与 Alpha158 因子集概述

在前期的系列研究报告中，我们借助 Github 上的开源项目 Qlib，采用其所集成的 Alpha158 量价因子开展了模型层面的基础研究。Alpha158 量价因子是一个丰富且多样化的因子集合，它涵盖了基于价格和成交量等多维度信息构建的多种因子。从因子构成来看，若不考虑窗口期，该因子集包含 42 个基础因子。而当纳入窗口期考量时，部分因子综合考虑了 5 日、10 日、20 日、30 日以及 60 日等不同长度的交易日窗口期，由此，Alpha158 因子集扩展为共计 158 个因子。为了更系统地理解和分析这些因子，我们依据其特征与计算逻辑，将 Alpha158 因子细致地划分为以下 5 大类别：

日内因子：这些因子仅使用当天的开盘价、收盘价、最高价、最低价以及均价数据，共包含 13 个因子。这些因子捕捉了市场在单个交易日内的波动和变化特征。

波动因子：波动因子主要衡量股票价格的波动性，共包含 5 个因子。波动性因子可以帮助我们理解股票价格的变动幅度和频率，从而更好地评估风险和收益。

价因子：价因子是基于股票的价格信息计算得出的，共包含 100 个因子。这些因子涉及多种价格计算方法和统计指标，能够反映出股票价格的长期和短期趋势。

量因子：量因子基于成交量数据，共包含 30 个因子。成交量是市场交易活动的直接反映，量因子能够提供关于市场流动性和投资者行为的重要信息。

量价相关性因子：这些因子同时考虑了成交量和价格的关系，共包含 10 个因子。通过分析量价关系，可以更深入地了解市场的供需动态和价格变动的内在驱动力。

表 1：Alpha 158 因子列表

因子大类	因子名称	窗口参数	因子释义	算式
单 日	HIGH0		最高价除以收盘价	High/Close
	KLEN		K 线长度	$(\text{High}-\text{Low})/\text{Open}$
	KLOW1		下影线长度相对开盘价的比例	$(\text{Less}(\text{Open}, \text{Close})-\text{Low})/\text{Open}$
	KLOW2		下影线长度相对 K 线整体的比例	$(\text{Less}(\text{Open}, \text{Close})-\text{Low})/(\text{High}-\text{Low}+1e-12)$
	KMID1		K 线实体长度相对开盘价的比例	$(\text{Close}-\text{Open})/\text{Open}$
	KMID2		K 线实体长度相对 K 线整体的比例	$(\text{Close}-\text{Open})/(\text{High}-\text{Low}+1e-12)$
	KSFT1		收盘价在整个价格区间（最高价到最低价）中的位置，相对开盘价的比例	$(2*\text{Close}-\text{High}-\text{Low})/\text{Open}$
	KSFT2		收盘价在整个价格区间（最高价到最低价）中的位置，相对 K 线整体的比例	$(2*\text{Close}-\text{High}-\text{Low})/(\text{High}-\text{Low}+1e-12)$
	KUP1		上影线长度相对开盘价的比例	$(\text{High}-\text{Greater}(\text{Open}, \text{Close}))/\text{Open}$
	KUP2		上影线长度相对 K 线整体的比例	$(\text{High}-\text{Greater}(\text{Open}, \text{Close}))/(\text{High}-\text{Low}+1e-12)$
	LOW0		最低价除以收盘价	Low/Close



	OPEN0		开盘价除以收盘价	$\$open/\$close$
	VWAP0		均价除以收盘价	$\$vwap/\$close$
波动	STD	[5, 10, 20, 30, 60]	过去 d 天的收盘价标准差, 除以最新的收盘价以去除单位。	$Std(\$close, \%d)/\$close$
	BETA	[5, 10, 20, 30, 60]	过去 d 天的收盘价变化率, 除以最新的收盘价以去除单位。例如, 过去 d 天每天价格上涨 10 美元, 那么斜率将是 10。	$Slope(\$close, \%d)/\$close$
	CNTD	[5, 10, 20, 30, 60]	过去上涨天数与过去下跌天数之间的差异。	$Mean(\$close > Ref(\$close, 1), \%d) - Mean(\$close < Ref(\$close, 1), \%d)$
	CNTN	[5, 10, 20, 30, 60]	过去 d 天内价格下跌的天数百分比。	$Mean(\$close < Ref(\$close, 1), \%d)$
	CNTP	[5, 10, 20, 30, 60]	过去 d 天内价格上涨的天数百分比。	$Mean(\$close > Ref(\$close, 1), \%d)$
	IMAX	[5, 10, 20, 30, 60]	当前日期与之前最高价日期之间的天数, 属于 Aroon 指标的一部分。该指标衡量一段时间内高点之间和低点之间的时间。强劲的上升趋势会定期看到新的高点, 而强劲的下降趋势会定期看到新的低点。	$IdxMax(\$high, \%d)/\%d$
	IMIN	[5, 10, 20, 30, 60]	当前日期与之前最低价日期之间的天数, 属于 Aroon 指标的一部分。	$IdxMin(\$low, \%d)/\%d$
	IMXD	[5, 10, 20, 30, 60]	之前最低价日期出现在最高价日期之后的时间段。较大的值表示向下的动量。	$(IdxMax(\$high, \%d) - IdxMin(\$low, \%d))/\%d$
	MA	[5, 10, 20, 30, 60]	简单移动平均线, 过去 d 天的简单移动平均线, 除以最新的收盘价以去除单位。	$Mean(\$close, \%d)/\$close$
	MAX	[5, 10, 20, 30, 60]	过去 d 天的最高价, 除以最新的收盘价以去除单位。	$Max(\$high, \%d)/\$close$
	MIN	[5, 10, 20, 30, 60]	过去 d 天的最低价, 除以最新的收盘价以去除单位。	$Min(\$low, \%d)/\$close$
	QTLN	[5, 10, 20, 30, 60]	过去 d 天收盘价的 20% 分位数, 除以最新的收盘价以去除单位。	$Quantile(\$close, \%d, 0.2)/\$close$
	QTLU	[5, 10, 20, 30, 60]	过去 d 天收盘价的 80% 分位数, 除以最新的收盘价以去除单位。	$Quantile(\$close, \%d, 0.8)/\$close$
	RANK	[5, 10, 20, 30, 60]	当前收盘价在过去 d 天收盘价中的百分位数, 表示当前价格水平相对于过去 d 天的比较, 提供额外的信息给移动平均线。	$Rank(\$close, \%d)$
	RESI	[5, 10, 20, 30, 60]	过去 d 天线性回归的残差, 表示过去 d 天的趋势线性程度。	$Resi(\$close, \%d)/\$close$
	ROC	[5, 10, 20, 30, 60]	变化率, 过去 d 天的价格变化, 除以最新的收盘价以去除单位。	$Ref(\$close, \%d)/\$close$
	RSQR	[5, 10, 20, 30, 60]	过去 d 天线性回归的 R 平方值, 表示趋势的线性程度。	$Rsquare(\$close, \%d)$
	RSV	[5, 10, 20, 30, 60]	表示过去 d 天内价格在上下阻力价格之间的位置。	$(\$close - Min(\$low, \%d)) / (Max(\$high, \%d) - Min(\$low, \%d) + 1e-12)$
	SUMD	[5, 10, 20, 30, 60]	总收益与总损失之间的差异比率, 类似于 RSI 指标。	$(Sum(Greater(\$close - Ref(\$close, 1), 0), \%d) - Sum(Greater(Ref(\$close, 1) - \$close, 0), \%d)) / (Sum(Abs(\$close - Ref(\$close, 1)), \%d) + 1e-12)$

量	SUMN	[5, 10, 20, 30, 60]	总损失与绝对总价格变化的比率，可以通过 SUMN = 1 - SUMP 得到，类似于 RSI 指标。	$\text{Sum}(\text{Greater}(\text{Ref}(\$close, 1) - \$close, 0), \%d) / (\text{Sum}(\text{Abs}(\$close - \text{Ref}(\$close, 1)), \%d) + 1e-12)$
	SUMP	[5, 10, 20, 30, 60]	总收益与绝对总价格变化的比率，类似于 RSI 指标。	$\text{Sum}(\text{Greater}(\$close - \text{Ref}(\$close, 1), 0), \%d) / (\text{Sum}(\text{Abs}(\$close - \text{Ref}(\$close, 1)), \%d) + 1e-12)$
	WVMA	[5, 10, 20, 30, 60]	交易量加权的价格变化波动率。	$\text{Std}(\text{Abs}(\$close / \text{Ref}(\$close, 1) - 1) * \$volume, \%d) / (\text{Mean}(\text{Abs}(\$close / \text{Ref}(\$close, 1) - 1) * \$volume, \%d) + 1e-12)$
	VMA	[5, 10, 20, 30, 60]	简单交易量移动平均线。	$\text{Mean}(\$volume, \%d) / (\$volume + 1e-12)$
	VSTD	[5, 10, 20, 30, 60]	过去 d 天的交易量标准差。	$\text{Std}(\$volume, \%d) / (\$volume + 1e-12)$
	VSUMD	[5, 10, 20, 30, 60]	总交易量增加与总交易量减少之间的差异比率，类似于交易量的 RSI 指标。	$\text{Sum}(\text{Greater}(\text{Ref}(\$volume, 1) - \$volume, 0), \%d) / (\text{Sum}(\text{Abs}(\$volume - \text{Ref}(\$volume, 1)), \%d) + 1e-12)$
	VSUMP	[5, 10, 20, 30, 60]	总交易量增加与绝对总交易量变化的比率。	$\text{Sum}(\text{Greater}(\$volume - \text{Ref}(\$volume, 1), 0), \%d) / (\text{Sum}(\text{Abs}(\$volume - \text{Ref}(\$volume, 1)), \%d) + 1e-12)$
	VSUMN	[5, 10, 20, 30, 60]	总交易量减少与绝对总交易量变化的比率，可以通过 VSUMN = 1 - VSUMP 得到。	$\text{Sum}(\text{Greater}(\text{Ref}(\$volume, 1) - \$volume, 0), \%d) / (\text{Sum}(\text{Abs}(\$volume - \text{Ref}(\$volume, 1)), \%d) + 1e-12)$
	CORD	[5, 10, 20, 30, 60]	价格变化率与交易量变化率之间的相关性。	$\text{Corr}(\$close / \text{Ref}(\$close, 1), \text{Log}(\$volume / \text{Ref}(\$volume, 1) + 1), \%d)$
	CORR	[5, 10, 20, 30, 60]	绝对收盘价与对数交易量之间的相关性。	$\text{Corr}(\$close, \text{Log}(\$volume + 1), \%d)$

资料来源: Qlib、信达证券研发中心

备注: 算式中的 “%d” 代表回看窗口期

表 2: Qlib 部分算子

算子	算子释义
Ref	变量检索, N=0, 检索第一个数据; N>0, 检索 N 个周期前的数据; N<0, 检索未来数据
Max	单变量滚动 N 个窗口期内最大值
Min	单变量滚动 N 个窗口期内最小值
Sum	单变量滚动 N 个窗口期求和
Mean	单变量滚动 N 个窗口期均值
Std	单变量滚动 N 个窗口期标准差
Var	单变量滚动 N 个窗口期方差
Skew	单变量滚动 N 个窗口期偏度
Kurt	单变量滚动 N 个窗口期峰度
Med	单变量滚动 N 个窗口期中位数
Mad	单变量滚动 N 个窗口期内和均值偏离的绝对值
Slope	单变量与 T(1,2,3...) 的滚动回归的回归系数项
Rsquare	单变量与 T(1,2,3...) 的滚动回归的 R 方
Resi	单变量与 T(1,2,3...) 的滚动回归的残差
Rank	单变量滚动 N 个窗口期排名
Quantile	单变量滚动 N 个窗口期百分位
Count	单变量滚动 N 个窗口期内非空数值
EMA	单变量滚动 N 个窗口期的指数移动平均
WMA	单变量滚动 N 个窗口期的加权移动平均

Corr	两个变量在滚动 N 个窗口期的相关性
Cov	两个变量在滚动 N 个窗口期的协方差
Delta	单变量在滚动 N 个窗口期的最后值减开始值
Abs	单变量的绝对值
Sign	单变量大于 0 的值置为 1，小于 0 的值置为-1
Log	单变量的自然对数
Power	单变量的指数幂
Greater	比较两个变量，返回最大值
Less	比较两个变量，返回最小值
IdxMax	单变量滚动 N 个窗口期内最大值的索引
IdxMin	单变量滚动 N 个窗口期内最小值的索引
If	条件判断

资料来源：Qlib、信达证券研发中心

Qlib 中集成了大量基于 Cython 的算子，这些算子与高开低收均价成交量数据巧妙结合，共同构成了因子表达式。将因子表达式输入 Qlib 框架，即可高效地对因子进行高性能计算。在此背景下，如何巧妙运用 DeepSeek 对因子表达式进行改进，进而实现对因子选股能力的显著增强，成为本章着重研究的核心内容。

1.2、动态交互框架：基于 Prompt Engineering 的因子优化流程设计

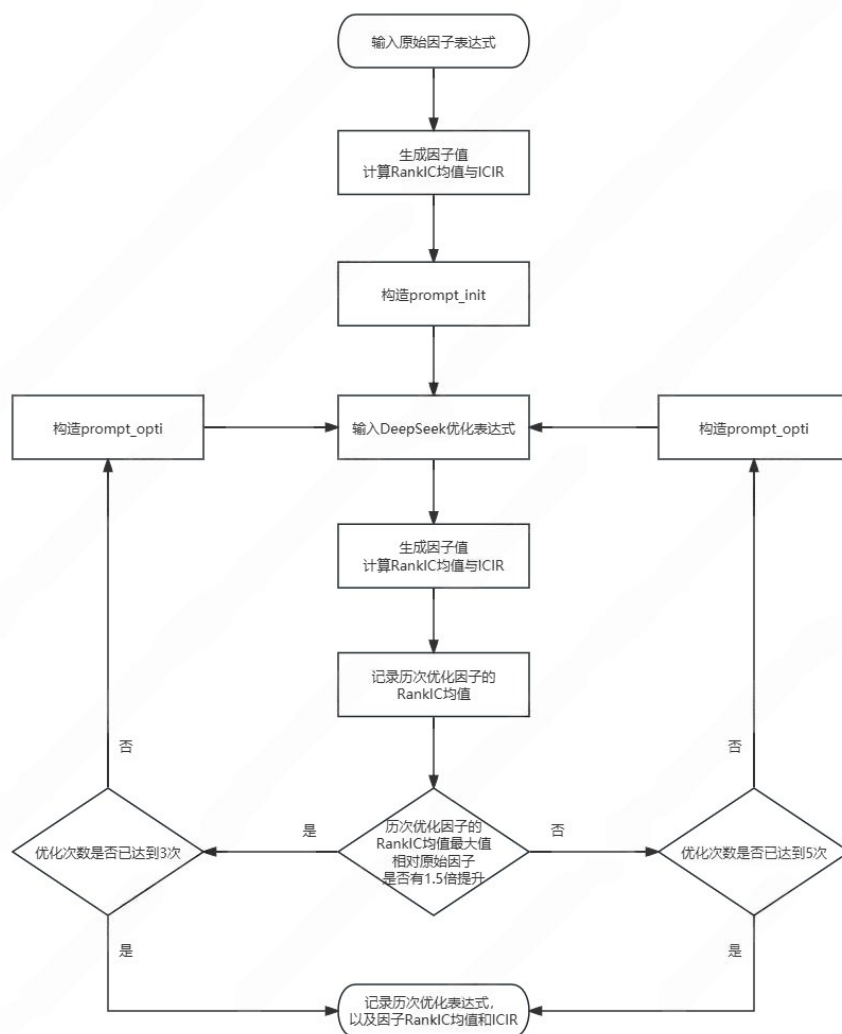
由于 Alpha158 量价因子支持多种窗口期的灵活计算，为确保研究结果的一致性与可比性，在测算因子 RankIC 均值与 ICIR 时，我们统一将窗口期参数设定为 20 个交易日。同时，明确以下的回测细节：

- 回测区间：2013 年 12 月 31 日至 2024 年 12 月 31 日。
- 剔除：剔除上市不满 365 个自然日的新股，剔除 ST 股。
- 中性化：对因子进行市值行业中性化
- 交易频率：周频调仓，以下周第一个交易日的 VWAP 价格成交，计算 VWAP 收益率的 RankIC 均值与 ICIR。
- 方向调整：根据 RankIC 均值的正负，对因子方向进行调整，使得 RankIC 为正，便于比较。

因子优化流程如下：

- 1) 对于每个因子，至少进行 3 次深度优化。在这 3 次优化过程中，让 DeepSeek 挖掘因子改进的潜力。若在 3 次优化后，优化因子的 RankIC 均值最大值达到原始因子的 1.5 倍，这意味着该因子已实现显著优化，此时将直接终止优化流程，并输出历次优化结果，以保留最佳优化路径与成果。
- 2) 若 3 次优化未能达到上述目标，即优化因子的 RankIC 均值未提升至原始因子的 1.5 倍，则继续进行优化。在后续优化过程中，最多尝试 5 次。这是为了在尽可能挖掘因子潜力的同时，避免过度优化导致的复杂性增加与收益递减，以及无谓的 token 消耗。若经过 5 次尝试后仍未达标，则终止优化并输出历次优化结果。

图 1：因子优化流程示意图



资料来源：信达证券研发中心

如何巧妙编写 prompt 是实现与大语言模型高效交互的核心要点。在与 DeepSeek 模型的交互中，我们需要精准地告知模型任务的具体内容、可用的关键信息以及一些不容忽视的重要注意事项，以下是 prompt_init 的主干内容：

假如你是一位资深的量化选股因子专家，你将根据现有的截面日频量化选股因子的相关信息，对以下因子进行改进以提升其 RankIC 均值：

[{factor_algo}]

这个因子是{direction}因子。

可用变量：

\$open：开盘价；\$close：收盘价；\$high：最高价；\$low：最低价；\$vwap：均价；\$volume：成交量。



可用算子代码:

[{code_content}]

表达式支持不同的窗口期,若窗口期为 20 个交易日,调整因子方向并进行市值行业中性化后,因子 2014 年以来周度 RankIC 均值为{rankic}, ICIR 为{icir}。

以提升因子的 RankIC 均值为目标,对这个因子的表达式进行改进,先列出至少 5 个改进方案,比较每个方案可能对因子 IC 提升的潜力,返回你认为最好的因子表达式及其优化逻辑,因子表达式格式参考提供的样例。

注意以下几点:

1. 只使用提供的算子,且保证调用方法正确,可用算子见代码中 OpsList 变量。
2. 对因子进行正确的去量纲操作,转换成比例的形式,使得不同股票间可比。
3. 除了 Ref 外,每个算子的窗口期相同,在表达式中仍以%d 表示。
4. 从逻辑出发进行改进,不需要计算因子的 RankIC 或 ICIR 等指标。
5. 因子根本逻辑不能被改变,例如波动类因子不能被改成动量类因子。

按照以下格式返回结果,其中改进后因子表达式写在中括号内部,不要换行,表达式内部括号只能用小括号,优化逻辑要分点罗列,并对改进后的因子表达式进行解释:

"

改进后因子表达式: [表达式]

因子优化逻辑: [优化逻辑]

因子解释: [因子解释]

"

资料来源:信达证券研发中心

以上 prompt 主要在于实现以下目的:

- **任务与信息传达:** 我们首先将 AI 带入资深量化选股因子专家的角色,简洁明了地阐述任务目标,即根据现有的截面日频量化选股因子相关信息,对特定因子进行改进,以显著提升其 RankIC 均值。同时,通过 [{factor_algo}]明确原始因子表达式,使 AI 清晰理解原始因子的量化逻辑;借助{direction}变量说明因子方向,帮助 AI 进一步区分不同类型因子,如反转因子与动量因子,从而为因子改进提供精准方向。
- **数据与算子说明:** 详细告知 AI 可用变量,包括 open (开盘价)、open (开盘价)、close (收盘价)、high (最高价)、high (最高价)、low (最低价)、vwap (均价)、vwap (均价)、volume (成交量),这些变量作为因子计算的基础数据,为模型改进因子表达式提供丰富素材。同时,将 Qlib 项目中的 ops.py 文件代码输入进大语言模型,以[{code_content}]的形式让 AI 清楚了解有哪些算子可用,以及每个算子的具体使用方法,为



因子表达式的创新改进提供技术支持。

- **测试条件与评估告知：**向 AI 明确后文中将以 20 个交易日为窗口期进行测试，并且在测试过程中会对因子进行方向调整、市值行业中性化等预处理操作。同时，告知 AI 原始因子在特定条件下的 RankIC 均值与 ICIR，使其对原始因子的选股效果形成初步评估，从而在改进过程中有针对性地提升因子表现。
- **任务要求与思路引导：**再次强调以提升因子的 RankIC 均值为核心目标，要求对因子表达式进行改进。先列出至少 5 个改进方案，并对每个方案可能提升因子 IC 的潜力进行比较，最终返回认为最好的因子表达式及其优化逻辑，因子表达式格式参考提供的样例。这不仅为模型提供了清晰的任务要求，还通过推荐的初始思路引导 Deepseek 在思维链中进行系统、全面的初步思考。
- **注意事项明确：**为确保模型改进的准确性与有效性，我们还明确了一系列注意事项。例如，限定只使用提供的算子，且保证调用方法正确，防止大模型幻觉导致 AI 使用不存在的算子；提示对因子进行正确的去量纲操作，转换成比例形式，以保证不同股票间可比；统一除 Ref 外每个算子的窗口期以 %d 表示，避免 AI 将某些算子的窗口期设置为其他值；要求从逻辑出发进行改进，避免 AI 因大模型幻觉“猜测”优化后的因子 IC 统计量；强调因子根本逻辑不能被改变，防止为追求高 RankIC 而改变因子本质类型。
- **返回格式约束：**按照特定格式要求 AI 返回结果，其中改进后因子表达式写在中括号内部，不得换行，表达式内部括号统一使用小括号。优化逻辑要分点罗列，并对改进后的因子表达式进行详细解释。这种严格的格式约束，方便了 Python 代码对返回内容进行准确解析。

基于上述针对 prompt_init 的设计思路，我们进一步构建了 prompt_opti。prompt_opti 主要用于收集模型历次优化结果，并据此告知模型继续优化。

以提升因子的 RankIC 均值为目标，比较之前表达式得到的 RankIC 均值与 ICIR，继续对这个因子的表达式进行改进，如果有必要的话可以推翻过去的方案重新思考。先列出至少 5 个改进方案，比较每个方案可能对因子 IC 提升的潜力，返回你认为最好的因子表达式及其优化逻辑，改进后的表达式需要不同于之前的表达式，因子表达式格式参考提供的样例。

按照以下格式返回结果，其中改进后因子表达式写在中括号内部，不要换行，表达式内部括号只能用小括号，优化逻辑要分点罗列，并对改进后的因子表达式进行解释：

"

改进后因子表达式：[表达式]

因子优化逻辑：[优化逻辑]

因子解释：[因子解释]

"

原始因子表达式为：{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}；

第 1 次改进后因子表达式为{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}；

第 2 次改进后因子表达式为{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}；

第 3 次改进后因子表达式为{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}；

.....

资料来源：信达证券研发中心

其主干内容设计思路为：首先以提升因子的 RankIC 均值为目标，引导模型比较之前表达式得到的 RankIC 均值与 ICIR，在此基础上继续对因子表达式进行改进。特别强调在必要情况下，可以推翻过去的方案重新思考，避免优化结果陷入局部最优解，确保模型能够持续探索更优的因子表达式。同时，通过收集原始因子和历次优化因子的表达式与预测效果，以详细罗列的方式（如原始因子表达式为：{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}；第 1 次改进后因子表达式为{}，RankIC 均值为{}，ICIR 为{}，因子方向为{}.....）让 AI 全面、深入地分析过去的优化方向与效果，从而有针对性地进一步改进或重新思考因子表达式。通过 prompt_init 与 prompt_opti 的协同配合，我们构建了一个完整、高效的与大语言模型交互的体系，形成因子优化的底层框架。

1.3、优化效果验证：RankIC 的跨周期普适性与稳定性突破

经过对 DeepSeek 改进后因子的测试，在设定的 5 次迭代范围内，29 个窗口期因子的表现呈现出积极态势。具体而言，共有 22 个因子的 RankIC 有所提升，其中 15 个因子的 RankIC 均值提升至 1.2 倍以上，10 个因子的 RankIC 均值更是提升至 1.5 倍以上，这些因子在选股能力的提升幅度上较为突出。

在衡量因子稳定性及预测能力的 ICIR 指标方面，同样有 23 个因子实现提升。其中 14 个因子的 ICIR 提升至 1.2 倍以上，10 个因子的 ICIR 提升至 1.5 倍以上。这一系列数据体现出 DeepSeek 对多数因子的优化卓有成效，切实增强了因子在选股策略中的有效性与可靠性。

表 3：原始因子与改进因子 RankIC 均值和 ICIR

因子类型	原始因子表达式	原始 RankIC 均值	原始 ICIR	改进次数	改进 RankIC 均值	改进 ICIR	RankIC 提升幅度	与原始因子相关性
std20	Std(\$close, %d)/\$close	4.03%	0.31	1	6.18%	0.42	2.14pct	69.30%
				2	7.80%	0.55	3.76pct	70.15%
				3	7.01%	0.79	2.98pct	6.21%
beta20	Slope(\$close, %d)/\$close	4.99%	0.45	1	4.49%	0.39	-0.49pct	72.75%
				2	3.66%	0.35	-1.32pct	-2.10%
				3	3.78%	0.38	-1.20pct	79.64%
				4	3.90%	0.38	-1.08pct	87.55%
				5	3.68%	0.36	-1.31pct	79.18%
cntd20	Mean(\$close>Ref(\$close, 1), %d)- Mean(\$close<Ref(\$close, 1), %d)	2.25%	0.29	1	3.99%	0.49	1.74pct	80.56%
				2	7.02%	0.70	4.77pct	40.43%
				3	1.23%	0.21	-1.02pct	-3.22%
cntn20	Mean(\$close<Ref(\$close, 1), %d)	2.07%	0.27	1	2.38%	0.39	0.31pct	9.99%
				2	4.59%	0.52	2.52pct	31.58%
				3	5.01%	0.54	2.94pct	24.05%
cntp20	Mean(\$close>Ref(\$close, 1), %d)	2.13%	0.29	1	3.30%	0.44	1.17pct	89.12%
				2	5.11%	0.56	2.98pct	55.19%
				3	4.20%	0.51	2.07pct	36.29%



imax20	IdxMax(\$high, %d)/%d	3.37%	0.44	1	1.48%	0.18	-1.89pct	36.99%
				2	2.04%	0.22	-1.33pct	48.64%
				3	1.18%	0.14	-2.20pct	-6.46%
				4	3.80%	0.51	0.43pct	48.55%
				5	3.16%	0.46	-0.22pct	41.35%
imin20	IdxMin(\$low, %d)/%d	2.59%	0.35	1	4.08%	0.53	1.50pct	-25.36%
				2	1.33%	0.17	-1.25pct	-39.09%
				3	5.11%	0.68	2.53pct	-5.69%
imxd20	(IdxMax(\$high, %d)-IdxMin(\$low, %d))/%d	3.86%	0.46	1	5.36%	0.53	1.50pct	75.29%
				2	1.67%	0.24	-2.19pct	-4.33%
				3	4.26%	0.53	0.40pct	76.22%
				4	2.85%	0.40	-1.02pct	65.02%
				5	5.59%	0.55	1.73pct	51.10%
ma20	Mean(\$close, %d)/\$close	6.89%	0.56	1	7.32%	0.57	0.43pct	97.36%
				2	4.60%	0.50	-2.30pct	46.88%
				3	4.62%	0.49	-2.27pct	50.04%
				4	6.04%	0.58	-0.85pct	42.40%
				5	1.89%	0.22	-5.00pct	21.53%
max20	Max(\$high, %d)/\$close	1.07%	0.09	1	1.89%	0.23	0.82pct	48.85%
				2	1.80%	0.23	0.73pct	11.34%
				3	1.88%	0.23	0.81pct	48.42%
min20	Min(\$low, %d)/\$close	7.48%	0.72	1	4.63%	0.38	-2.85pct	78.38%
				2	3.62%	0.29	-3.86pct	62.27%
				3	3.63%	0.49	-3.85pct	21.83%
				4	3.40%	0.38	-4.08pct	72.10%
				5	0.14%	0.03	-7.34pct	4.55%
qtl20	Quantile(\$close, %d, 0.2)/\$close	7.34%	0.66	1	7.37%	0.65	0.03pct	99.56%
				2	7.31%	0.57	-0.03pct	87.47%
				3	0.62%	0.11	-6.72pct	-0.08%
				4	7.31%	0.57	-0.03pct	87.14%
				5	7.08%	0.54	-0.26pct	77.67%
qtlu20	Quantile(\$close, %d, 0.8)/\$close	5.73%	0.47	1	3.87%	0.45	-1.85pct	27.59%
				2	3.93%	0.56	-1.80pct	-4.98%
				3	1.24%	0.28	-4.48pct	4.75%
				4	3.80%	0.54	-1.93pct	-5.65%
				5	2.81%	0.49	-2.92pct	1.26%
rank20	Rank(\$close, %d)	5.64%	0.55	1	2.71%	0.33	-2.93pct	61.99%
				2	2.59%	0.32	-3.05pct	21.88%
				3	2.24%	0.34	-3.40pct	15.43%
				4	2.66%	0.32	-2.99pct	21.36%
				5	3.92%	0.44	-1.73pct	68.29%
resi20	Resi(\$close, %d)/\$close	3.68%	0.37	1	3.88%	0.40	0.19pct	81.93%
				2	3.87%	0.40	0.19pct	85.69%
				3	3.46%	0.42	-0.22pct	71.67%
				4	1.90%	0.25	-1.78pct	14.12%



				5	3.87%	0.41	0.19pct	88.68%
				1	1.07%	0.09	-5.61pct	52.04%
				2	6.89%	0.56	0.21pct	82.54%
roc20	Ref(\$close, %d)/\$close	6.68%	0.56	3	7.32%	0.57	0.63pct	83.12%
				4	3.68%	0.37	-3.00pct	-4.80%
				5	6.94%	0.57	0.26pct	83.06%
rsqr20	Rsquare(\$close, %d)	0.15%	0.02	1	0.04%	0.01	-0.12pct	-13.30%
				2	0.68%	0.12	0.53pct	-18.55%
				3	0.46%	0.09	0.31pct	-17.22%
				1	5.54%	0.51	1.03pct	91.54%
rsv20	(\$close-Min(\$low, %d))/(Max(\$high, %d)-Min(\$low, %d)+1e-12)	4.51%	0.42	2	5.62%	0.48	1.11pct	83.88%
				3	3.93%	0.40	-0.58pct	42.00%
				4	5.95%	0.51	1.44pct	80.91%
				5	5.82%	0.50	1.31pct	82.33%
				1	6.18%	0.69	0.94pct	31.87%
sumd20	(Sum(Greater(\$close-Ref(\$close, 1), 0), %d)-Sum(Greater(Ref(\$close, 1)-\$close, 0), %d))/(Sum(Abs(\$close-Ref(\$close, 1)), %d)+1e-12)	5.24%	0.49	2	4.42%	0.62	-0.81pct	15.96%
				3	3.31%	0.43	-1.92pct	39.40%
				4	5.26%	0.62	0.03pct	10.02%
				5	2.99%	0.56	-2.25pct	11.82%
				1	1.23%	0.17	-4.01pct	-13.20%
sumn20	Sum(Greater(Ref(\$close, 1)-\$close, 0), %d)/(Sum(Abs(\$close-Ref(\$close, 1)), %d)+1e-12)	5.24%	0.49	2	1.39%	0.18	-3.85pct	-14.97%
				3	0.91%	0.13	-4.32pct	-14.91%
				4	1.05%	0.15	-4.19pct	-13.00%
				5	0.91%	0.13	-4.32pct	-14.91%
				1	5.94%	0.55	0.70pct	40.16%
sump20	Sum(Greater(\$close-Ref(\$close, 1), 0), %d)/(Sum(Abs(\$close-Ref(\$close, 1)), %d)+1e-12)	5.24%	0.49	2	6.29%	0.54	1.06pct	86.48%
				3	3.56%	0.47	-1.68pct	27.56%
				4	3.10%	0.40	-2.13pct	-3.94%
				5	5.73%	0.59	0.50pct	60.27%
				1	3.64%	0.52	0.03pct	94.00%
				2	3.27%	0.47	-0.34pct	-23.54%
vma20	Mean(\$volume, %d)/(\$volume+1e-12)	3.61%	0.51	3	0.29%	0.06	-3.33pct	52.82%
				4	2.85%	0.47	-0.76pct	88.07%
				5	6.25%	0.80	2.64pct	3.93%
				1	4.61%	0.77	4.02pct	-40.40%
vstd20	Std(\$volume, %d)/(\$volume+1e-12)	0.59%	0.11	2	4.09%	0.72	3.50pct	-56.15%
				3	4.35%	0.75	3.76pct	-17.28%
				1	4.17%	0.62	-0.11pct	36.50%
				2	2.42%	0.52	-1.86pct	-5.40%
vsumd20	(Sum(Greater(\$volume-Ref(\$volume, 1), 0), %d)-Sum(Greater(Ref(\$volume, 1)-\$volume, 0), %d))/(Sum(Abs(\$volume-Ref(\$volume, 1)), %d)+1e-12)	4.28%	0.67	3	1.53%	0.31	-2.75pct	40.46%
				4	4.25%	0.70	-0.03pct	91.27%
				5	2.13%	0.39	-2.15pct	59.45%
				1	4.28%	0.67	0.00pct	99.99%
vsumn20	Sum(Greater(Ref(\$volume, 1)-\$volume, 0), %d)/(Sum(Abs(\$volume-Ref(\$volume, 1)), %d)+1e-12)	4.28%	0.67	2	5.59%	0.80	1.31pct	54.73%
				3	2.32%	0.54	-1.96pct	56.72%

					4	3.61%	0.52	-0.67pct	38.54%
					5	5.57%	0.64	1.30pct	49.60%
					1	5.22%	0.81	0.94pct	12.81%
					2	5.30%	0.81	1.02pct	12.94%
					3	5.67%	0.86	1.39pct	12.32%
vsump20	Sum(Greater(\$volume-Ref(\$volume, 1), 0), %d)/(Sum(Abs(\$volume-Ref(\$volume, 1)), %d)+1e-12)	4.28%	0.67		4	3.03%	0.43	-1.25pct	7.33%
					5	3.77%	0.58	-0.51pct	5.11%
					1	3.42%	0.60	1.65pct	72.01%
					2	3.31%	0.60	1.54pct	69.17%
					3	3.07%	0.54	1.30pct	75.16%
wvma20	Std(Abs(\$close/Ref(\$close, 1)-1)*\$volume, %d)/(Mean(Abs(\$close/Ref(\$close, 1)-1)*\$volume, %d)+1e-12)	1.77%	0.32		1	1.76%	0.40	-3.33pct	9.64%
					2	0.23%	0.05	-4.86pct	-3.55%
					3	1.00%	0.21	-4.09pct	-0.10%
					4	2.57%	0.65	-2.52pct	14.77%
					5	3.99%	0.59	-1.10pct	70.53%
cord20	Corr(\$close/Ref(\$close, 1), Log(\$volume/Ref(\$volume, 1)+1), %d)	5.09%	0.76		1	4.51%	0.62	-0.07pct	70.62%
					2	3.93%	0.75	-0.65pct	32.57%
					3	5.17%	0.77	0.59pct	50.06%
					4	5.25%	0.79	0.67pct	49.30%
					5	4.46%	0.64	-0.12pct	63.78%
corr20	Corr(\$close, Log(\$volume+1), %d)	4.58%	0.68						

资料来源: Qlib、信达证券研发中心

数据日期: 2013 年12 月31 日至2024 年12 月31 日

以上测算基于周频调仓,并将表达式中的窗口期%d 设置为 20。为进一步探究 DeepSeek 优化效果的稳定性,我们思考若将窗口期设置为其他参数,优化出来的因子表达式是否仍能保持提升态势?为此,本文选取 Alpha158 中优化后 RankIC 均值有提升的因子,对原始表达式与改进后 20 日 RankIC 均值最高的表达式,在 5 / 10 / 20 / 30 / 60 个交易日的窗口期下测算其 RankIC 均值。

表 4: 不同因子类型下原始表达式与增强表达式 RankIC 均值

因子类型	表达式状态	5	10	20	30	60
std	原始	5.55%	5.19%	4.03%	3.42%	2.33%
	增强	8.43%	8.37%	7.79%	7.23%	5.95%
cntd	原始	3.04%	2.42%	2.25%	2.04%	1.55%
	增强	6.33%	6.40%	7.02%	6.92%	6.36%
cntn	原始	2.78%	2.17%	2.07%	1.82%	1.21%
	增强	3.86%	4.43%	5.01%	4.86%	4.26%
cntp	原始	3.01%	2.40%	2.13%	1.88%	1.42%
	增强	4.77%	4.62%	5.11%	5.04%	4.57%
imax	原始	3.47%	3.11%	3.37%	3.47%	2.67%
	增强	4.00%	3.68%	3.80%	3.91%	3.32%
imin	原始	2.83%	2.51%	2.59%	2.62%	1.76%
	增强	6.63%	6.16%	5.11%	4.62%	4.29%
imxd	原始	3.96%	3.63%	3.86%	3.89%	2.83%
	增强	5.48%	5.30%	5.59%	5.27%	3.60%
ma	原始	5.63%	6.17%	6.89%	7.19%	7.03%

	增强	5.95%	6.72%	7.32%	7.47%	7.31%
	原始	0.14%	0.14%	1.07%	1.64%	2.14%
max	增强	2.52%	2.35%	1.89%	2.30%	3.08%
	原始	6.53%	6.98%	7.34%	7.57%	7.31%
qtd	增强	6.61%	6.98%	7.37%	7.59%	7.31%
	原始	2.21%	3.51%	3.68%	4.06%	5.34%
resi	增强	2.63%	3.97%	3.88%	3.86%	4.69%
	原始	5.86%	5.99%	6.68%	6.60%	5.68%
roc	增强	5.95%	6.72%	7.32%	7.47%	7.31%
	原始	0.54%	0.47%	0.15%	0.34%	0.00%
rsqr	增强	0.84%	0.77%	0.68%	0.49%	0.07%
	原始	4.39%	4.27%	4.51%	4.66%	4.63%
rsv	增强	5.43%	5.78%	5.95%	5.82%	5.36%
	原始	4.66%	4.66%	5.24%	5.23%	4.75%
sumd	增强	5.33%	5.84%	6.18%	6.19%	5.94%
	原始	4.66%	4.66%	5.24%	5.23%	4.75%
sump	增强	5.40%	5.87%	6.29%	6.42%	6.31%
	原始	2.23%	2.84%	3.61%	4.14%	5.01%
vma	增强	6.21%	6.32%	6.25%	6.15%	5.93%
	原始	1.68%	0.65%	0.59%	1.48%	3.15%
vstd	增强	3.66%	4.26%	4.61%	4.43%	3.81%
	原始	2.69%	3.10%	4.28%	4.47%	4.97%
vsumn	增强	4.66%	4.95%	5.59%	5.42%	5.09%
	原始	2.69%q	3.10%	4.28%	4.47%	4.97%
vsump	增强	6.57%	6.27%	5.68%	5.31%	4.50%
	原始	0.77%	1.21%	1.77%	1.74%	1.35%
wvma	增强	2.79%	3.31%	3.42%	3.17%	2.35%
	原始	4.83%	5.25%	4.58%	4.01%	2.60%
corr	增强	4.33%	4.92%	5.25%	4.86%	4.10%

资料来源：Qlib、信达证券研发中心

数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

从表 4 数据可见，绝大部分原始因子表达式经 DeepSeek 优化后，不仅在 20 日窗口期下 RankIC 均值显著提升，在其他窗口期下同样展现出增强态势。这充分证明 AI 所优化的因子表达式在不同时间窗口下具备普适性，能够在多种市场时间尺度下，有效提升因子对股票收益的预测能力。

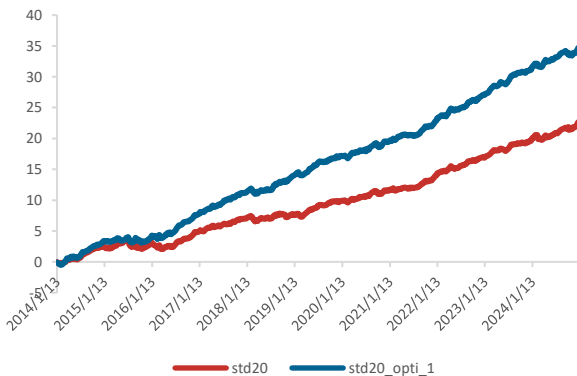
1.4、以波动率因子为例：洞察模型优化方向

我们选取波动率因子 std20 展开深入剖析，以此探究 DeepSeek 对因子的改进方式，并明晰因子选股效果的提升究竟源于大语言模型的内在实力还是偶然因素。

std20 因子的原始表达式为 “Std(\$close,%d)/\$close”，其含义明确，旨在计算过去 20 个交易日收盘价的标准差，并通过除以收盘价实现去量纲处理，以此衡量价格的波动程度。

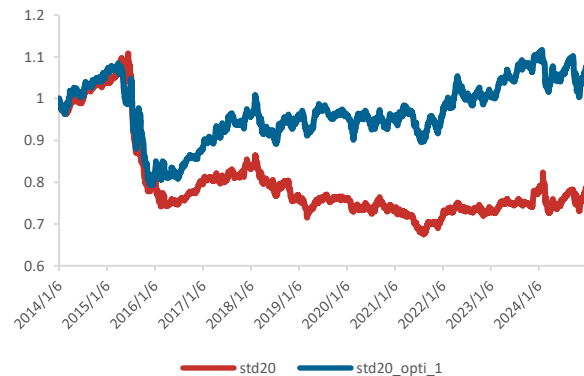
在第一次改进中，DeepSeek 将因子表达式修改为“ $\text{Mean}(\text{Greater}(\text{Shigh}-\text{Slow}, \text{Greater}(\text{Abs}(\text{Shigh}-\text{Ref}(\$close,1)), \text{Abs}(\text{Slow}-\text{Ref}(\$close,1))))), \%d)/\$close$ ”，该表达式在原始基础上，对分子进行了创新调整，引入了平均真实波幅 ATR 的概念。此 ATR 捕捉了价格波动中的日内波动、向上跳空和向下跳空三种模式，相较于单纯依赖收盘价标准差，能更为敏锐地识别价格剧烈波动的股票。分母依旧维持除以最新收盘价的形式，保持去量纲的操作。引入 ATR 计算波动率后，因子的 RankIC 均值从 4.03% 提升至 6.18%，ICIR 从 0.31 提升至 0.42，年化多头超额从 -2.23% 提升至 0.71%，年化多空收益从 7.33% 大幅提升至 17.42%。

图 2：std20 原始因子与第一次改进因子累计 RankIC



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

图 3：std20 原始因子与第一次改进因子多头超额净值



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

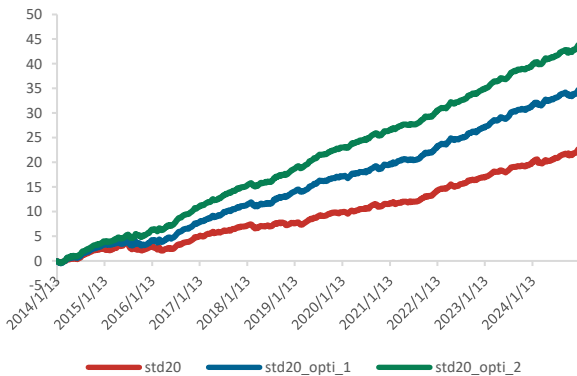
在第二次改进中，DeepSeek 将因子表达式修改为“ $\text{EMA}(\text{Greater}(\text{Shigh}-\text{Slow}, \text{Greater}(\text{Abs}(\text{Shigh}-\text{Ref}(\$close,1)), \text{Abs}(\text{Slow}-\text{Ref}(\$close,1)))) * \$volume, \%d) / \text{EMA}(\$volume, \%d) / \$close$ ”，该表达式在第一次优化表达式的基础上，主要实施了两点关键改进：

- 1) 引入成交量加权机制：将真实波动幅度(TR)与当日成交量相乘，强化量价共振效应，高成交量伴随的波动更具信息含量，通过 $\text{EMA}(\$volume, \%d)$ 消除成交量绝对值影响，构建单位成交量波动比率。
- 2) 采用 EMA 双重平滑：对分子分母同时进行指数加权平均，既保留成交量加权特性又加强近期数据的权重。

DeepSeek 对该因子的逻辑解释为：“该因子通过成交量加权的指数移动平均真实波动率，捕捉资金流动驱动的价格不稳定性。相比简单平均 TR，成交量加权能识别主力资金参与的异常波动，EMA 处理强化了近期市场情绪的敏感性，双重 EMA 标准化有效剥离了量价量纲差异。该设计同时满足波动测量的全面性、量价关系的协同性以及市场记忆衰减的特性，预期能更精准识别高投机风险个股。”，可以看出模型对该因子的理解基本无误。

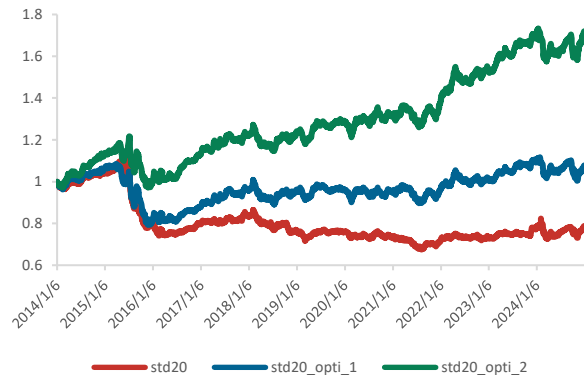
经此改进，因子的 RankIC 均值由第一次优化的 6.18% 提升至 7.8%，ICIR 由 0.42 提升至 0.55，年化多头超额从 0.71% 提升至 5.24%，年化多空收益由 17.42% 提升至 33.81%。从相关对比图中能够明显观察到，该波动率因子在 2015 年的回撤得到显著改善。

图 4：std20 原始因子与第一二次改进因子累计 RankIC



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

图 5：std20 原始因子与第一二次改进因子多头超额净值



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

在第三次改进中，DeepSeek 将因子表达式优化为“EMA(Greater(Greater(\$high-\$low, Greater(Abs(\$high-Ref(\$close,1)), Abs(\$low-Ref(\$close,1))))), Abs(\$open-Ref(\$close,1))) * Sign(\$close-Ref(\$close,1)) + 1e-5 * \$volume, %d) / EMA(\$volume, %d) / \$close”。此表达式在第二次优化的基础上，又增添了两点改进：

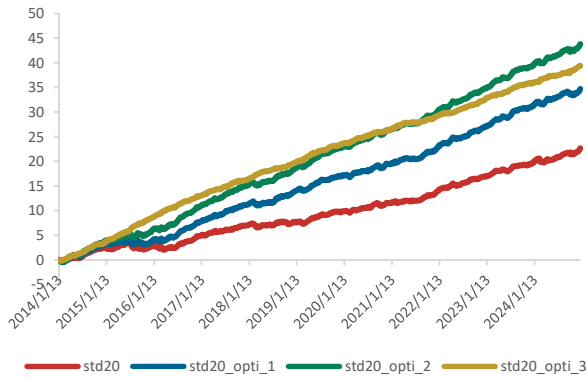
- 1) 四维波动极值捕捉：在原有三重比较基础上新增开盘跳空缺口（\$open 与前收盘价差），形成 high-low、high-prev_close、low-prev_close、open-prev_close 四维波动极值筛选。
- 2) 引入方向敏感系数：通过 Sign(\$close-Ref(\$close,1)) 捕捉收盘方向，下跌日波动赋予负权重强化空头波动信号。

DeepSeek 对该因子的逻辑解释为：“该因子通过整合日内波动、隔夜跳空和方向性量价信息，构建多维波动冲击指标。在捕捉最大价格变动的基础上，通过收盘方向符号强化下跌波动的负面效应，结合成交量加权突出资金驱动的不稳定性。相比前两版改进，该设计更精准识别空头力量主导的异常波动，预期可提升对风险溢价补偿要求的定价能力。”

第三次优化后，因子的 RankIC 均值由第二次优化的 7.8% 略微降低至 7.01%，但 ICIR 由 0.55 提升至 0.79，年化多头超额由 5.24% 提升至 6.29%，年化多空收益由 33.81% 大幅提升至 51.10%。从测算效果来看，虽然第三次优化的 RankIC 均值有所下降，但 ICIR、年化多头超额收益与年化多空超额收益均有显著提升。然而，从人为主观理解因子的角度，我们认为 DeepSeek 对因子的理解存在一定偏差。

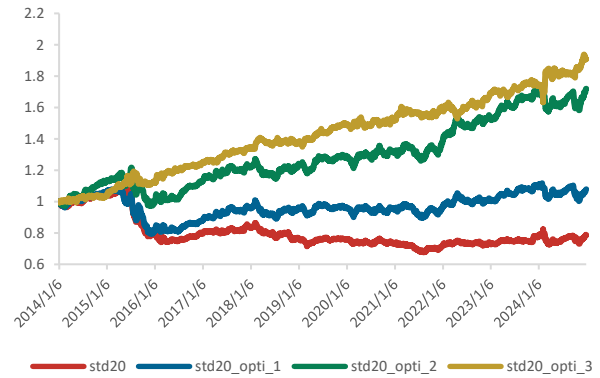
原始波动因子与历次优化因子均为负向因子，意味着因子值越大，后续可能跑输市场整体；因子值越小，后续可能战胜市场整体。第三次的因子表达式中“Greater(Greater(\$high-\$low, Greater(Abs(\$high-Ref(\$close,1)), Abs(\$low-Ref(\$close,1))))), Abs(\$open-Ref(\$close,1)))”部分会返回一个正数，而表达式中“Sign(\$close-Ref(\$close,1))”部分则根据当天的涨跌调整前者的方向（涨为 1，跌为 -1）。因此，若股票长期下跌，分子端可能为负数，因子值较小；若长期盘整，分子端趋近于 0；若呈上涨趋势，分子端为正数，因子值较大。所以，我们理解该因子在原有波动率因子的基础上，叠加了反转因子，从而使因子 ICIR 及收益得到一定提升。

图 6：std20 原始因子与第一二三次改进因子累计 RankIC



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

图 7：std20 原始因子与第一二三次改进因子多头超额净值



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

除优化成功的案例外，我们也对优化无明显提升的案例进行了总结与分析。以 Beta20 因子为例，DeepSeek 在该因子上尝试了 5 次优化，但 RankIC 均值和 ICIR 均未得到提升。Beta20 原始因子构造逻辑为过去 N 个交易日收盘价的斜率除以收盘价，因子方向为负向，本质仍为反转因子。由于受到 prompt 中“因子根本逻辑不能被改变”的限制，DeepSeek 主要对分母端进行改进尝试，但效果不佳。类似的情况还出现在 min20 因子与 qtlu20 因子上，同样因分子端逻辑限制，导致改进效果不显著。

表 5：Beta20 原始因子与历次优化因子表达式及 IC 统计量

因子名称	因子表达式	RankIC 均值	ICIR
beta20	$\text{Slope}(\$/\text{close}, \%d) / \$/\text{close}$	4.99%	0.45
beta20_opti_1	$\text{Slope}(\text{EMA}(\$/\text{close}, \%d), \%d) / \text{EMA}(\$/\text{close}, \%d)$	4.49%	0.39
beta20_opti_2	$\text{Slope}(\text{WMA}(\$/\text{close}, \%d), \%d) / \text{WMA}(\$/\text{close}, \%d)$	3.66%	0.35
beta20_opti_3	$\text{Slope}(\$/\text{close}, \%d) / \text{Std}(\$/\text{close}, \%d)$	3.78%	0.38
beta20_opti_4	$\text{Slope}(\$/\text{vwap}, \%d) / \text{Mean}(\text{Greater}(\$/\text{high} - \$/\text{low}, \text{Abs}(\$/\text{vwap} - \text{Ref}(\$/\text{vwap}, 1))), \%d)$	3.90%	0.38
beta20_opti_5	$\text{Slope}(\$/\text{close}, \%d) / \text{Std}(\$/\text{high} - \$/\text{low}, \%d)$	3.68%	0.36

资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

表 6：min20 原始因子与历次优化因子表达式及 IC 统计量

因子名称	因子表达式	RankIC 均值	ICIR
min20	$\text{Min}(\$/\text{low}, \%d) / \$/\text{close}$	7.48%	0.72
min20_opti_1	$\text{Min}(\$/\text{low}, \%d) / \text{EMA}(\$/\text{close}, \%d)$	4.63%	0.38
min20_opti_2	$\text{Min}(\$/\text{low}, \%d) / \text{Ref}(\text{Mean}(\text{Greater}(\$/\text{close}, \$/\text{low}), \%d), 1)$	3.62%	0.29
min20_opti_3	$\text{Min}(\$/\text{low}, \%d) / (\text{WMA}(\$/\text{close}, \%d) * \text{Std}(\$/\text{close}, \%d))$	3.63%	0.49
min20_opti_4	$\text{Min}(\$/\text{low}, \%d) / (\text{EMA}(\$/\text{close}, \%d) - \text{Std}(\$/\text{close}, \%d))$	3.40%	0.38
min20_opti_5	$\text{Min}(\$/\text{low}, \%d) * \text{Mean}(\text{Greater}(\$/\text{close}, \text{Ref}(\$/\text{low}, 1)), \%d) / \$/\text{close}$	0.14%	0.03

资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

表 7：qtlu20 原始因子与历次优化因子表达式及 IC 统计量

因子名称	因子表达式	RankIC 均值	ICIR
------	-------	-----------	------

qtlu20	$\text{Quantile}(\$close, \%d, 0.8)/\$close$	5.73%	0.47
qtlu20_opti_1	$\text{Quantile}(\$close, \%d, 0.9)*\text{Slope}(\$close, \%d)/\$close$	3.87%	0.45
qtlu20_opti_2	$\text{EMA}(\text{Quantile}(\text{Greater}(\$close, \text{EMA}(\$close, \%d)), \%d, 0.8), \%d)/(\text{Std}(\$close, \%d)*\$close)$	3.93%	0.56
qtlu20_opti_3	$\text{Resi}(\text{Quantile}(\text{Greater}(\$close*\$volume, \text{Mean}(\$close*\$volume, \%d)), \%d, 0.85), \%d)/(\$close*\text{Std}(\$close, \%d))$	1.24%	0.28
qtlu20_opti_4	$\text{Greater}(\text{Quantile}(\$close, \%d, 0.8), \text{EMA}(\$close, \%d))/(\text{Std}(\$close, \%d)*\$close)$	3.80%	0.54
qtlu20_opti_5	$\text{Quantile}(\text{Greater}(\$close, \text{EMA}(\$close, \%d)), \%d, 0.75)*\text{Rsquare}(\$close, \%d)/(\$close*\text{Std}(\$close, \%d))$	2.81%	0.49

资料来源：万得、信达证券研发中心

数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

二、从优化到创造：基于大语言模型的新因子生成范式

2.1、从零生成的困境：独立探索下的效率与效果瓶颈

在第一章里，我们以 Alpha158 原始因子表达式为依托，借助 DeepSeek 开展因子改进工作，取得了一定成果，但也察觉到部分因子受限于原有逻辑，难以进一步优化。由此引发思考：若给予 DeepSeek 更自由的发挥空间，使其从因子改进转向从零进行因子生成的任务，它能否巧妙组合各类算子与可用数据，挖掘出具备出色选股能力的因子？这正是本节着力探讨的核心问题。

我们依旧从 prompt 入手，清晰告知模型所需执行的任务。因子生成任务的 prompt 由 prompt_init 与 prompt_opti 构成，以下详述 prompt_init 的主干内容：

假如你是一位资深的量化选股因子专家，你将根据现有的算子，挖掘周度 ICIR 的绝对值在 0.8 以上新因子，你返回因子表达式之后，我会调整因子方向并进行市值行业中性化，自行测算因子 IC 和 ICIR。

可用变量：

\$open：开盘价；\$close：收盘价；\$high：最高价；\$low：最低价；\$vwap：均价；\$volume：成交量。

可用算子代码：

[{code_content}]

注意以下几点：

1. 只使用提供的算子，且保证调用方法正确，可用算子见代码中 OpsList 变量。
2. 对因子进行正确的去量纲操作，转换成比例的形式，使得不同股票间可比。
3. 除了 Ref 外，每个算子的窗口期相同，且都用 %d 表示，不要出现算子窗口期为固定常数的情况，之后在测算过程中会将 %d 设置为 20。
4. 从逻辑出发进行挖掘，不需要计算因子的 RankIC 或 ICIR 等指标。

按照以下格式返回结果，其中挖掘的因子表达式写在中括号内部，不要换行，表达式内部括号只能用小括号，并对挖掘的因子表达式进行解释：

"

改进后因子表达式：[表达式]

因子解释：[因子解释]

"

资料来源：信达证券研发中心

prompt_opti 内容为：

第 1 次尝试因子表达式为 {}，RankIC 均值为 {}，ICIR 为 {}，因子方向为 {}。

第 2 次尝试因子表达式为 {}，RankIC 均值为 {}，ICIR 为 {}，因子方向为 {}。

第 3 次尝试因子表达式为 {}, RankIC 均值为 {}, ICIR 为 {}, 因子方向为 {}。

.....

在此基础上继续改进或者重新思考, 如果表达式越来越复杂且效果越来越差, 则可能需要对表达式进行简化或者完全重新思考新的方案。

资料来源: 信达证券研发中心

因子生成部分的 prompt_init 与因子优化部分的 prompt_init 大致相似, 只是去除了对 Alpha158 表达式和 IC 统计量的描述, 同时为模型设定了生成周度 ICIR 在 0.8 以上因子的明确目标。prompt_opti 部分则着重将历次生成的表达式与 IC 统计量反馈给模型, 促使模型在前述基础上不断改进。在实际操作过程中, 我们发现模型可能会将表达式修改得过于复杂, 而实际效果却不尽人意, 所以在 prompt_opti 中特别提示模型, 可适时简化表达式或者重新构思全新方案。

表 8: 历史模型生成的因子表达式与 IC 统计量

迭代次数	因子表达式	方向	RankIC 均值	ICIR
1	Log(Greater(\$close,Mean(\$high,%d))/Less(\$close,Mean(\$low,%d))*Sign(Slope(\$volume,%d))	负向	3.74%	0.46
2	Slope((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d)),%d) * Sign(Resi(\$volume,%d))	负向	1.09%	0.21
3	Sign(Slope(\$close/%d,%d)) * (Mean(\$volume,%d)/Ref(\$volume,%d) - 1)	负向	2.33%	0.46
4	Slope(\$close/%d,%d) * Slope(Log(\$volume,%d))	负向	2.37%	0.41
5	Slope(\$close,%d) * (Ref(\$volume,%d)/Mean(\$volume,%d) - 1)/Std(\$volume,%d)	正向	0.91%	0.21
6	Slope((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d)),%d) * ((\$volume - Mean(\$volume,%d))/Std(\$volume,%d))	负向	1.73%	0.26
7	Slope((\$close/Mean(\$close,%d)),%d) * (Rank(\$volume,%d) - Ref(Rank(\$volume,%d),%d))	负向	1.97%	0.32
8	Log((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d))) * (\$volume/Mean(\$volume,%d) - 1)	负向	3.46%	0.52
9	(\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d)) * Sign((\$volume/Ref(\$volume,%d) - 1) - Mean((\$volume/Ref(\$volume,%d) - 1),%d))	负向	1.55%	0.24
10	(\$close/Mean(\$close,%d) - 1)/Std(\$close,%d) * (\$volume - Mean(\$volume,%d))/Std(\$volume,%d)	负向	1.92%	0.32
11	Log((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d))) * (Slope(\$volume,%d)/Std(\$volume,%d))	负向	3.35%	0.49
12	Log(\$close/Mean(\$high,%d)) * (Ref(\$volume,%d)/Mean(\$volume,%d) - 1)	正向	2.13%	0.35
13	Log(\$close/Mean(\$high,%d)) * ((\$volume - Mean(\$volume,%d))/Std(\$volume,%d))	负向	2.18%	0.27
14	Log(\$close/Mean(\$close,%d)) * Slope(\$volume,%d)	负向	2.99%	0.46
15	Log(\$close/Mean(\$close,%d)) * Sign(Slope(\$volume,%d))	负向	2.70%	0.35
16	Log((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d))) * (\$volume/Mean(\$volume,%d) - 1)	负向	3.46%	0.52
17	(((\$close - Min(\$low,%d))/(Max(\$high,%d) - Min(\$low,%d))) * ((\$volume - Mean(\$volume,%d))/Std(\$volume,%d))	负向	3.88%	0.59
18	Log((\$close - Min(\$low,%d))/(Max(\$high,%d) - Min(\$low,%d))) * (Slope(\$volume,%d)/Std(\$volume,%d))	正向	2.28%	0.33
19	Log((\$close - Min(\$low,%d))/(Max(\$high,%d) - Min(\$low,%d))) * ((\$volume - Min(\$volume,%d))/(Max(\$volume,%d) - Min(\$volume,%d)))	负向	0.68%	0.09
20	Log((\$close - Mean(\$low,%d))/(Mean(\$high,%d) - Mean(\$low,%d))) * Sign(Slope(\$volume,%d))	负向	2.80%	0.44

资料来源: 万得、信达证券研发中心

数据日期: 2013 年 12 月 31 日至 2024 年 12 月 31 日

从上表可知, 我们让 DeepSeek 对因子表达式进行了 20 次迭代, 然而生成的因子预测效果普遍一般。其中,

RankIC 均值最大值为 3.88%，最小值仅 0.68%；ICIR 最大值为 0.59，与 prompt 中要求 0.8 的 ICIR 仍存在一定差距。

尽管在这 20 次迭代中，ICIR 未达预设目标，但 DeepSeek 给出的因子表达式并非毫无逻辑。比如第一次生成的因子表达式为：

$$\text{Log}(\text{Greater}(\$close, \text{Mean}(\$high, \%d)) / \text{Less}(\$close, \text{Mean}(\$low, \%d))) * \text{Sign}(\text{Slope}(\$volume, \%d))$$

可以看出 DeepSeek 试图通过该因子捕捉量价协同效应，对放量上涨或缩量下跌的股票赋予因子值权重，而长期盘整的股票因子值则相对较小。

第二次生成的因子表达式为：

$$\text{Slope}((\$close - \text{Mean}(\$low, \%d)) / (\text{Mean}(\$high, \%d) - \text{Mean}(\$low, \%d)), \%d) * \text{Sign}(\text{Resi}(\$volume, \%d))$$

是在第一次表达式基础上的升级。DeepSeek 对该表达式的解释为：

- 1) 价格标准化：(收盘价-最低价均值)/(最高价均值-最低价均值)形成 0-1 区间波动，消除价格绝对水平影响；
- 2) 价格趋势：对标准化价格序列计算斜率捕捉趋势强度；
- 3) 量能异常：用成交量残差的符号表征非趋势性放量。该设计同时捕捉价格突破强度与异常量能信号，且标准化过程使跨股票可比性增强。

不过我们也发现，DeepSeek 对该表达式的理解存在一定误区，例如模型提到“(收盘价 - 最低价均值)/(最高价均值 - 最低价均值)形成 0-1 区间波动”，实际上对于趋势下跌的股票，最新收盘价可能小于过去 N 个交易日最低价均值，波动区间可能低于 0。但同时也能看出，DeepSeek 有一些创新构造方式，比如用“Sign(Resi(\$volume,%d))”捕捉异常的放量与缩量，在高频场景下，这种方式对收益预测确有一定帮助。

2.2、案例驱动的突破：基于先验知识的低相关性因子挖掘

在上一节中，我们让 DeepSeek 从零开始生成因子，历经 20 次迭代，挖掘出的因子 ICIR 仍未达预期。那么，若为模型提供一些成功案例，使其能总结这些案例的优点，是否能实现更出色的因子挖掘效果呢？基于此，本节我们将第一章中的 Alpha158 原始因子表达式、优化后的表达式以及对应的 IC 统计量作为“成功案例”输入模型，探究在此基础上能否取得更好成果。

我们依旧从 prompt 着手，清晰告知模型需要执行的任务。因子生成任务的 prompt 仍由 prompt_init 与 prompt_opti 构成。以下是 prompt_init 的主干内容：

假如你是一位资深的量化选股因子专家，你将根据现有的截面日频量化选股因子的相关信息，根据样例因子表达式，挖掘周度 ICIR 的绝对值在 0.8 以上新因子：

以下是通过算子计算得到的因子与对应的 RankIC 均值及 ICIR，名称中含_opti 的因子是由原始因子(不含 opti)尝试改进后得到，RankIC 均值与 ICIR 是假设表达式中 %d 为 20 个交易日，调整因子方向并进行市值行业中性化后计算得到的：

[

因子名称：std20；因子表达式：Std(\$close,%d)/\$close；RankIC 均值：0.0403；ICIR：0.3105；因子方向：负向；

因子名称：std20_opti_1；因子表达式：Mean(Greater(\$high-\$low, Greater(Abs(\$high-Ref(\$close,1)), Abs(\$low-Ref(\$close,1)))), %d)/\$close；RankIC 均值：0.0618；ICIR：0.4152；因子方向：负向；

因子名称：std20_opti_2；因子表达式：EMA(Greater(\$high-\$low, Greater(Abs(\$high-Ref(\$close,1)), Abs(\$low-Ref(\$close,1))))*\$volume,%d)/EMA(\$volume,%d)/\$close；RankIC 均值：0.0780；ICIR：0.5520；因子方向：负向；

.....

]

可用变量：

\$open：开盘价；\$close：收盘价；\$high：最高价；\$low：最低价；\$vwap：均价；\$volume：成交量。

可用算子代码：

[{code_content}]

注意以下几点：

1. 只使用提供的算子，且保证调用方法正确，可用算子见代码中 OpsList 变量。
2. 对因子进行正确的去量纲操作，转换成比例的形式，使得不同股票间可比。
3. 除了 Ref 外，每个算子的窗口期相同，且都用 %d 表示，不要出现算子窗口期为固定常数的情况。

4. 从逻辑出发进行挖掘，不需要计算因子的 RankIC 或 ICIR 等指标。
5. 生成的新因子与样例因子不能相同，相关性尽可能低。

按照以下格式返回结果，其中挖掘的因子表达式写在中括号内部，不要换行，表达式内部括号只能用小括号，并对挖掘的因子表达式进行解释：

"

改进后因子表达式：[表达式]

因子解释：[因子解释]

"

资料来源：信达证券研发中心

prompt_opti 内容为：

第 1 次尝试因子表达式为 {}，RankIC 均值为 {}，ICIR 为 {}，因子方向为 {}。
 第 2 次尝试因子表达式为 {}，RankIC 均值为 {}，ICIR 为 {}，因子方向为 {}。
 第 3 次尝试因子表达式为 {}，RankIC 均值为 {}，ICIR 为 {}，因子方向为 {}。

在此基础上继续改进或者重新思考，如果表达式越来越复杂且效果越来越差，则可能需要对表达式进行简化或者完全重新思考新的方案。

资料来源：信达证券研发中心

本节的 prompt 与从零生成因子的 prompt 相比，新增了将 Alpha158 原始与优化表达式及其 IC 统计量输入给模型这一内容，prompt_opti 则与前文保持一致。

表 9：从 Alpha158 生成的因子表达式与 IC 统计量

因子 序号	迭代 次数	因子表达式	方向	RankIC 均值	ICIR
1	1	EMA(Greater(\$close-Ref(\$close,1),0)*Power(\$close/Ref(\$close,1)-1,2)*\$volume,%d)/EMA(\$volume,%d)-EMA(Greater(Ref(\$close,1)-\$close,0)*Power(Ref(\$close,1)/\$close-1,2)*\$volume,%d)/EMA(\$volume,%d)	负向	6.74%	0.83
	1	EMA(Greater(Greater(\$high-Ref(\$close,1),Abs(\$open-Ref(\$close,1))),\$low-Ref(\$close,1))*Sign(\$volume-EMA(\$volume,%d))*Power(\$close/Ref(\$close,1),2,%d)/EMA(Power(\$close/Ref(\$close,1),2)*\$volume,%d)	负向	3.89%	0.65
2	2	EMA((\$close/Ref(\$close,1)-1)*Sign(\$volume-EMA(\$volume,%d))*Log(\$volume+1,%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	4.67%	0.77
	3	EMA((((\$close/Ref(\$close,1)-1)*Sign(\$volume-Ref(\$volume,1)) - Mean((\$close/Ref(\$close,1)-1)*Sign(\$volume-Ref(\$volume,1)),%d)),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	3.52%	0.64
	4	EMA((\$close/Ref(\$close,1)-1)*Sign(\$volume-EMA(\$volume,%d))*Abs(\$close/Ref(\$close,1)-1,%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	6.56%	0.82



3	1	EMA(Greater(\$high-Ref(\$close,1),\$close-Ref(\$low,1))*(\$close/Ref(\$close,1)-1)*Sign(\$volume-EMA(\$volume,%d)),%d)/(Std(\$close,%d)*EMA(\$volume,%d))	负向	2.56%	0.40
	2	EMA((Sign(Slope(\$close,%d))*(\$close/Ref(\$close,1)-1)*(\$volume/EMA(\$volume,%d))),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	2.36%	0.29
	3	EMA((\$close/Ref(\$close,1)-1)*Power(\$volume/EMA(\$volume,%d),2),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	6.59%	0.72
	4	EMA((\$close/Ref(\$close,1)-1)*Power(\$volume/EMA(\$volume,%d),2)*Sign(Slope(\$volume,%d)),%d)/Std(Resi(\$close,%d),%d)	负向	2.48%	0.47
	5	EMA((\$close/Ref(\$close,1)-1)*Power(\$volume/EMA(\$volume,%d),2),%d)/Std(Resi(\$close,%d),%d)	负向	5.17%	0.65
	6	EMA((\$close/Ref(\$close,1)-1)*Power(\$volume/EMA(\$volume,%d),2),%d)/(Std(Resi(\$close,%d),%d)*Std(\$volume/EMA(\$volume,%d),%d))	负向	4.60%	0.59
	7	EMA((\$close/Ref(\$close,1)-1)*Power(\$volume/EMA(\$volume,%d),1.5),%d)/Std(Resi(\$close,%d),%d)	负向	5.38%	0.64
	8	EMA(Greater(\$high-Ref(\$close,1),Abs(\$close-Ref(\$low,1)))*(\$close/Ref(\$close,1)-1)*(\$volume/EMA(\$volume,%d)),%d)/Std(Resi(\$close,%d),%d)	负向	8.76%	0.87
4	1	EMA(Greater(\$high-Ref(\$close,1),Abs(\$close-Ref(\$low,1)))*Sign(\$close-EMA(\$close,%d))*Power(\$volume/EMA(\$volume,%d),0.5),%d)/Std(Resi(\$close,%d),%d)	负向	5.49%	0.54
	2	EMA(Greater(\$high-Ref(\$close,1),Abs(\$close-Ref(\$low,1)))*Sign(\$close-EMA(\$close,%d))*(\$volume/EMA(\$volume,%d)),%d)/Std(Resi(\$close,%d),%d)	负向	5.91%	0.59
	3	EMA((Greater(\$high-Ref(\$close,1),Abs(\$close-Ref(\$low,1)))*Sign(\$close-Ref(\$close,1))*(\$volume/EMA(\$volume,%d)),%d)/Std(Greater(\$high-\$low,Abs(\$high-Ref(\$close,1))),%d)	负向	5.24%	0.59
	4	EMA((Slope(\$close,%d)*Sign(\$volume-EMA(\$volume,%d))),%d)/Std(Resi(\$volume,%d),%d)	正向	0.09%	0.02
	5	EMA((Slope(\$close,%d)*Sign(\$volume/Ref(\$volume,1)-1)),%d)/Std(\$close/Ref(\$close,1)-1,%d)	正向	0.60%	0.13
	6	EMA((\$close/Ref(\$close,1)-1)*(\$volume/Ref(\$volume,1)-1),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	6.20%	0.87
5	1	EMA(Greater(Greater(\$high-\$low,Abs(\$high-Ref(\$close,1))),Abs(\$low-Ref(\$close,1)))*Sign(\$close-Ref(\$close,1)+1e-5)*\$volume,%d)/EMA(\$volume,%d)/Std(\$close,%d)	负向	5.12%	0.65
	2	EMA((\$high-\$low)*(\$close/Ref(\$close,1)-1)*(\$volume/EMA(\$volume,%d)),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	7.27%	0.87
	3	EMA(Resi(\$close,%d)*Sign(\$volume-EMA(\$volume,%d))*(Max(\$high,%d)/\$close-1),%d)/Std(\$close,%d)	负向	2.90%	0.43
	4	EMA((\$close/Ref(\$close,1)-1)*Sign(\$volume/Ref(\$volume,1)-1),%d)/Std(\$close/Ref(\$close,1)-1,%d)	负向	5.87%	0.90

资料来源：万得、信达证券研发中心

数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

经过有限测试，最终生成了 5 个 ICIR 在 0.8 以上的因子。其中，第一次尝试经一次迭代便得到满足条件的因子表达式，有时则需历经 8 次迭代才得到较好的因子。

尽管我们在 prompt 中要求新因子与参考样例因子相关性尽可能低，但 DeepSeek 作为大语言模型自身不具备

计算能力。因此，我们测算了新因子与样例因子中 ICIR 大于 0.7 的因子之间的相关性，以及新因子之间的相关性。

表 10：新因子与 ICIR 大于 0.7 的样例因子之间的相关性

	Factor_1	Factor_2	Factor_3	Factor_4	Factor_5
std20_opti_3	31.85%	16.81%	42.47%	21.85%	20.32%
cntd20_opti_2	29.35%	22.24%	44.47%	24.76%	22.70%
min20	-24.01%	-16.18%	-34.03%	-15.51%	-15.98%
vma20_opti_5	6.01%	-2.90%	3.51%	-0.31%	-1.07%
vstd20_opti_1	10.58%	9.81%	22.56%	13.39%	10.74%
vstd20_opti_2	7.28%	6.08%	15.79%	9.72%	7.36%
vstd20_opti_3	6.09%	9.02%	19.50%	14.73%	9.47%
vsumn20_opti_2	19.01%	13.67%	26.94%	16.71%	21.44%
vsump20_opti_1	44.38%	3.83%	8.01%	4.52%	4.12%
vsump20_opti_2	45.28%	3.97%	8.29%	4.65%	4.23%
vsump20_opti_3	43.98%	7.72%	18.73%	12.38%	10.62%
cord20	14.60%	8.25%	14.39%	7.54%	15.56%
corr20_opti_2	7.29%	24.52%	24.25%	17.13%	71.36%
corr20_opti_3	14.22%	8.19%	14.28%	7.34%	15.94%
corr20_opti_4	12.19%	8.92%	9.00%	4.18%	13.68%

资料来源：万得、信达证券研发中心

数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

表 11：新因子之间相关性

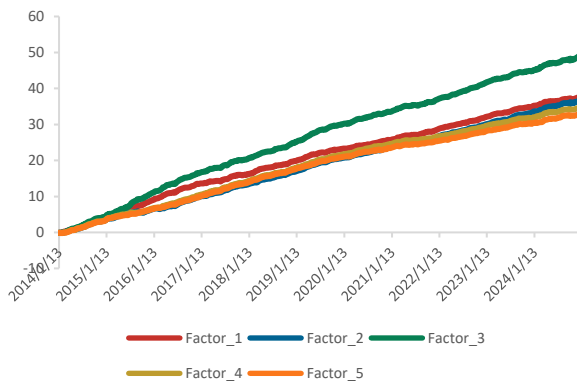
	Factor_1	Factor_2	Factor_3	Factor_4	Factor_5
Factor_1		8.90%	20.70%	11.61%	9.40%
Factor_2	8.92%		11.22%	33.16%	48.19%
Factor_3	20.91%	11.17%		41.24%	29.23%
Factor_4	11.64%	33.16%	41.20%		46.75%
Factor_5	9.42%	48.19%	29.32%	46.75%	

资料来源：万得、信达证券研发中心

数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

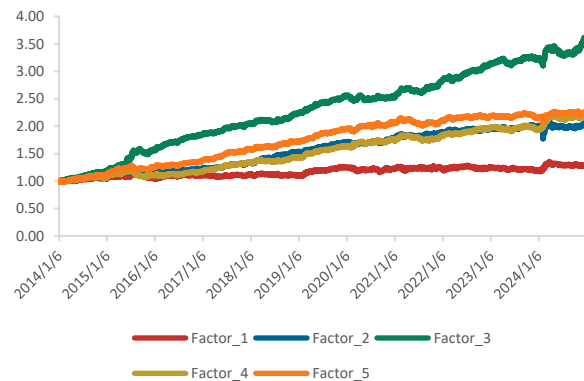
从表中可见，新生成的因子与样例因子中效果较好的因子相关性较低，仅 Factor_5 与 corr20_opti_2 之间的相关性达到 71.36%，其他因子之间的相关性均在 50% 以下，其中 Factor_2 与所有出色样例因子相关性均在 25% 以下。新因子之间的相关性也在可接受范围内，所有因子之间相关性最大值为 48.19%，均值为 26.07%，其中 Factor_1 与其它 4 个因子的相关性均在 21% 以下。

图 8：新生成因子累计 RankIC



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

图 9：新生成因子多头超额净值



资料来源：万得、信达证券研发中心
数据日期：2013 年 12 月 31 日至 2024 年 12 月 31 日

以新生成的前两个因子为例，分析 DeepSeek 所生成因子的底层逻辑。

第一个有效因子表达式为：

$$\text{EMA}(\text{Greater}(\$close - \text{Ref}(\$close, 1), 0) * \text{Power}(\$close / \text{Ref}(\$close, 1) - 1, 2) * \$volume, \%d) / \text{EMA}(\$volume, \%d) - \text{EMA}(\text{Greater}(\text{Ref}(\$close, 1) - \$close, 0) * \text{Power}(\text{Ref}(\$close, 1) / \$close - 1, 2) * \$volume, \%d) / \text{EMA}(\$volume, \%d)$$

AI 对该因子的逻辑解释为：“该因子捕捉量价协同增强效应。分子端计算上涨日价格涨幅平方与成交量的加权均值（EMA），突出强势股的量能驱动特征；分母端进行成交量标准化消除规模影响。通过对称计算下跌日的量价惩罚项并做差值，有效区分多头与空头动能。价格涨幅平方处理放大突破行情的区分度，EMA 平滑避免短期噪声，最终形成兼具趋势强度和量能验证的双向驱动因子。”

可见 DeepSeek 对因子表达式的理解基本正确。通过学习样例表达式，AI 在第一次迭代中便能够使用“ $\$volume / \text{EMA}(\$volume, \%d)$ ”表达成交量加权。在创新性上，AI 使用“上涨日价格变动 \times 涨幅平方 - 下跌日价格变动 \times 跌幅平方”区分多头与空头动能，使得因子 ICIR 达到 0.83。

第二个有效因子表达式为：

$$\text{EMA}((\$close / \text{Ref}(\$close, 1) - 1) * \text{Sign}(\$volume - \text{EMA}(\$volume, \%d)) * \text{Abs}(\$close / \text{Ref}(\$close, 1) - 1, \%d) / \text{Std}(\$close / \text{Ref}(\$close, 1) - 1, \%d))$$

AI 对该因子的逻辑解释为：“该因子聚焦量价共振的强度维度：1）价格收益率与成交量突破 EMA 的方向符号相乘，捕捉量价同向动量；2）乘以收益率绝对值强化大波动日的信号权重；3）EMA 平滑保留趋势持续性特征；4）用收益率波动率标准化消除市场整体波动影响。通过“方向 \times 幅度 \times 量能验证”三维度构建纯净动量信号。”

人为拆分该因子，表达式分子由 3 个部分相乘组成“ $(\$close / \text{Ref}(\$close, 1) - 1)$ ”、“ $\text{Sign}(\$volume -$

EMA(\$volume,%d))”和“Abs(\$close/Ref(\$close,1)-1)”。不难发现，对于连续放量大涨，或者连续缩量大跌，又或者处于盘整过程中涨时放量跌时缩量的股票，在该表达式中的分子端值会偏大，反之则偏小。而在分母端除以近期的波动率，则进一步放大了近期连续放量大涨或缩量大跌的因子值。该因子周频 ICIR 达到 0.82，且与第一个因子的相关性仅为 8.92%。

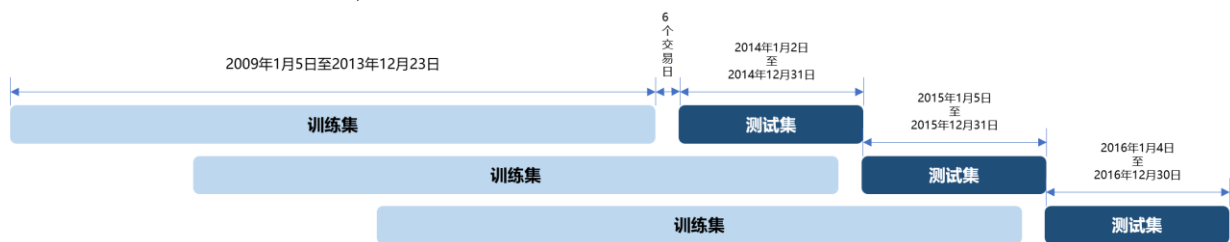
三、策略应用与组合绩效验证：复合因子在指增组合中的增益

在前两章中，我们借助 DeepSeek 成功实现了对原始 Alpha158 因子的优化，并在此基础上挖掘出 5 个相关性较低的新因子。从单因子测试结果来看，优化因子与新因子均展现出出色的选股预测能力。那么，若采用线性模型将这些因子进行复合，效果究竟如何？落实到指增组合上，又能带来多少提升？这是本节重点探讨的内容。

3.1、复合因子的收益增强：中证 800 中 RankIC 均值提升至 11.03%

我们以量价因子作为输入特征，在全 A 股票范围内，运用惩罚系数为 0.001 的 Lasso 模型对因子进行合成。具体而言，训练集长度设定为过去 5 个自然年，每年年初重新估算模型参数，这些参数将用于接下来一年的因子合成以及对未来 5 个交易日的收益预测，数据划分示意图如下。

图 10：预测未来 5 个交易日的均价收益率，数据划分示意图



资料来源：万得、信达证券研发中心

若使用 Alpha158 增强因子替代原始因子，代入 Lasso 模型进行合成，结果显示增强合成因子的最终选股能力略优于原始合成因子。增强合成因子的 RankIC 均值为 11.03%，超越了原始因子的 10.6%。从各年份数据来看，仅 2015、2021 与 2025 年 RankIC 均值低于原始合成因子，在其他年份均有不同程度的提升。同时，增强合成因子的年化多头超额收益为 23.5%，收益波动比为 4.66，这表明增强合成因子在获取收益方面具有一定优势，且收益的稳定性相对较好。

图 11：原始合成因子与增强合成因子累计 RankIC

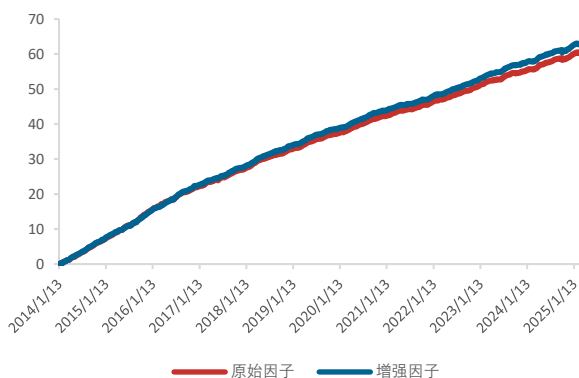
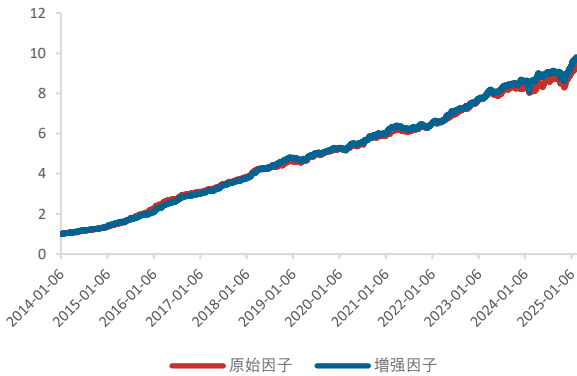


表 12：原始合成因子与增强合成因子 IC 统计量

	RankIC 均值		ICIR	
	原始因子	增强因子	原始因子	增强因子
2014	13.66%	13.94%	1.99	1.85
2015	16.01%	15.46%	1.51	1.50
2016	13.38%	14.37%	1.52	1.27
2017	9.91%	10.43%	1.08	0.97
2018	11.32%	12.05%	1.50	1.23
2019	8.87%	9.61%	1.15	1.16
2020	9.30%	9.66%	1.06	1.09
2021	7.67%	7.64%	1.04	1.03
2022	9.23%	9.89%	1.35	1.26
2023	8.79%	9.75%	1.14	1.02
2024	8.98%	9.28%	0.60	0.60
2025	7.88%	7.14%	0.47	0.63

资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

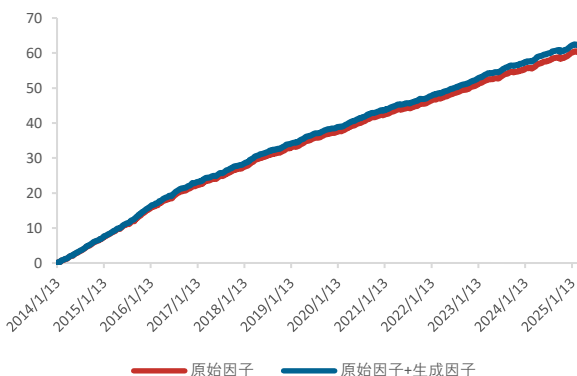
图 12：原始合成因子与增强合成因子多头超额收益



资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

进一步探讨第二章中生成的因子相对于原始 Alpha158 因子是否具有增量信息。我们将原始因子与生成因子一同代入 Lasso 模型进行合成，并比较“原始 + 生成”合成因子与原始合成因子的选股能力。结果表明，“原始 + 生成”合成因子的 RankIC 均值为 10.94%，高于原始因子的 10.6%。从各年份数据来看，仅 2025 年 RankIC 均值低于原始因子，在其他年份加入新生成因子后均有稳定提升。

图 13：原始合成因子与原始+生成合成因子累计 RankIC



资料来源：万得、信达证券研发中心

全区间 10.60% 11.03% 1.13 1.08
资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

表 13：原始合成因子与增强合成因子多头超额与收益波动比

	多头超额		收益波动比	
	原始因子	增强因子	原始因子	增强因子
2014	33.78%	36.40%	7.67	8.14
2015	67.24%	52.22%	10.33	7.50
2016	37.58%	44.32%	6.21	8.07
2017	23.81%	25.59%	6.09	6.82
2018	21.55%	27.14%	4.32	6.35
2019	13.18%	10.39%	3.77	3.05
2020	12.41%	14.00%	2.73	3.06
2021	10.10%	8.01%	2.27	1.85
2022	18.06%	18.72%	4.09	4.50
2023	8.06%	10.81%	1.97	2.77
2024	8.59%	9.39%	1.12	1.22
2025	3.97%	3.44%	0.78	0.65
全区间	23.10%	23.50%	4.48	4.66

资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

表 14：原始合成因子与原始+生成合成因子 IC 统计量

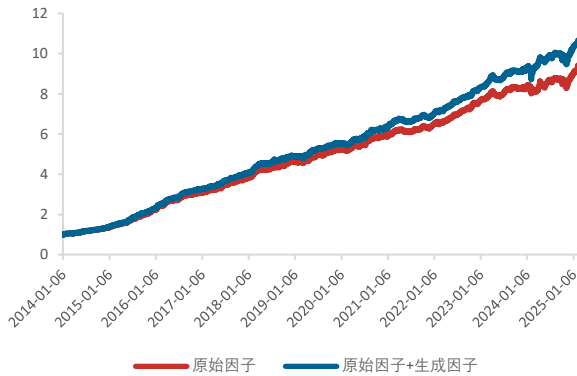
	RankIC 均值		ICIR	
	原始因子	原始+生成因子	原始因子	原始+生成因子
2014	13.66%	13.86%	1.99	1.93
2015	16.01%	16.80%	1.51	1.65
2016	13.38%	14.16%	1.52	1.36
2017	9.91%	10.05%	1.08	1.00
2018	11.32%	11.69%	1.50	1.31
2019	8.87%	9.09%	1.15	1.19
2020	9.30%	9.38%	1.06	1.12
2021	7.67%	7.80%	1.04	1.03
2022	9.23%	9.71%	1.35	1.40
2023	8.79%	9.37%	1.14	1.11
2024	8.98%	9.21%	0.60	0.60
2025	7.88%	6.51%	0.47	0.50
全区间	10.60%	10.94%	1.13	1.11

资料来源：万得、信达证券研发中心

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

图 14：原始合成因子与原始+生成合成因子多头超额收益



资料来源：万得、信达证券研发中心

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

表 15：原始因子与原始+生成因子多头超额与收益波动比

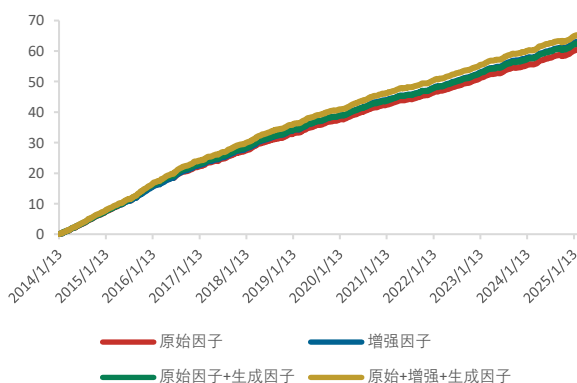
	多头超额		收益波动比	
	原始因子	原始+生成因子	原始因子	原始+生成因子
2014	33.78%	36.65%	7.67	8.41
2015	67.24%	68.82%	10.33	10.34
2016	37.58%	41.23%	6.21	7.24
2017	23.81%	24.56%	6.09	6.54
2018	21.55%	20.94%	4.32	4.30
2019	13.18%	12.93%	3.77	3.68
2020	12.41%	13.72%	2.73	3.14
2021	10.10%	11.08%	2.27	2.64
2022	18.06%	18.65%	4.09	4.48
2023	8.06%	10.65%	1.97	2.66
2024	8.59%	12.09%	1.12	1.50
2025	3.97%	2.33%	0.78	0.58
全区间	23.10%	24.49%	4.48	4.80

资料来源：万得、信达证券研发中心

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

最后，我们将所有因子，即原始因子、增强因子与生成因子一起代入 Lasso 模型，得到最终的合成因子。该合成因子的 RankIC 均值为 11.41%，明显优于原始因子、增强因子以及叠加生成因子后的原始因子。在多头超额收益方面，合成因子达到 24.64%，同样优于原始因子组合的 23.1%。

图 15：各因子累计 RankIC



资料来源：万得、信达证券研发中心

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

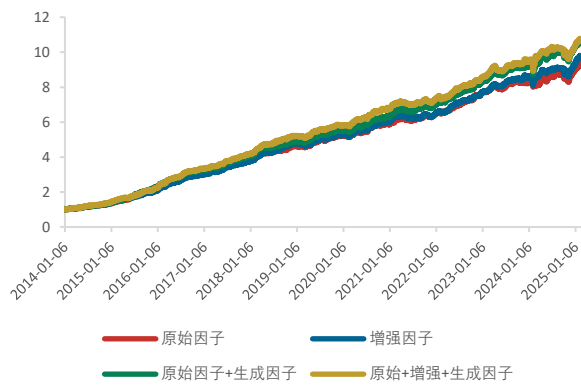
表 16：各因子 IC 统计量

	RankIC 均值		ICIR	
	原始因子	合成因子	原始因子	合成因子
2014	13.66%	14.55%	1.99	2.08
2015	16.01%	16.72%	1.51	1.62
2016	13.38%	15.52%	1.52	1.57
2017	9.91%	11.05%	1.08	1.10
2018	11.32%	12.39%	1.50	1.50
2019	8.87%	9.53%	1.15	1.18
2020	9.30%	10.18%	1.06	1.14
2021	7.67%	7.94%	1.04	1.09
2022	9.23%	9.84%	1.35	1.33
2023	8.79%	9.59%	1.14	1.13
2024	8.98%	9.09%	0.60	0.59
2025	7.88%	6.92%	0.47	0.47
全区间	10.60%	11.41%	1.13	1.16

资料来源：万得、信达证券研发中心

数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

图 16：各因子多头超额收益



资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

表 17：各因子多头超额与收益波动比

	多头超额		收益波动比	
	原始因子	合成因子	原始因子	合成因子
2014	33.78%	41.62%	7.67	9.20
2015	67.24%	58.60%	10.33	8.37
2016	37.58%	48.10%	6.21	7.65
2017	23.81%	25.38%	6.09	6.95
2018	21.55%	25.13%	4.32	5.41
2019	13.18%	11.59%	3.77	3.31
2020	12.41%	16.10%	2.73	3.64
2021	10.10%	8.58%	2.27	1.99
2022	18.06%	16.77%	4.09	3.70
2023	8.06%	10.56%	1.97	2.60
2024	8.59%	9.42%	1.12	1.15
2025	3.97%	3.01%	0.78	0.75
全区间	23.10%	24.64%	4.48	4.70

资料来源：万得、信达证券研发中心
数据日期：2014 年 1 月 6 日至 2025 年 2 月 28 日

3.2、指增组合的实战检验：月度 800 指增超额提升近 1%

我们将 Lasso 模型合成的因子值应用到 800 指增策略中，以此深入探究大语言模型优化与生成的因子对指增组合的增量贡献。约束条件与交易方式如下：

约束条件：

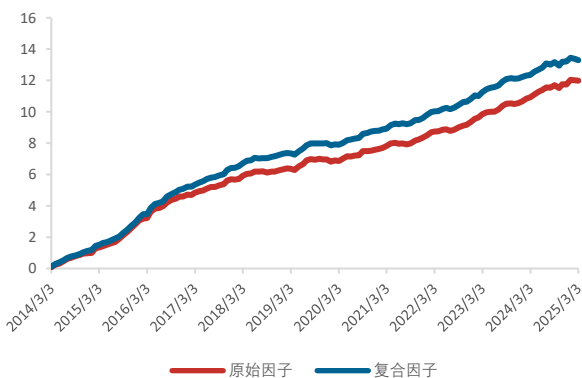
- ✓ 100%指数成分股内选股。
- ✓ 个股权重最大偏离 0.8%。
- ✓ 中信一级行业最大偏离 3%。
- ✓ 市值风格最大偏离 0.01，其它 CNE5 风格最大偏离 0.3。
- ✓ 每次调仓单边换手率 40%以下。

交易方式：

- 月度调仓，以每月初第一个交易日的 VWAP 价格成交。
- 一字涨停不能买入，一字跌停不能卖出，停牌不能交易。
- 手续费：单边千分之一。

在中证 800 指数成分股中，将原始因子与 DeepSeek 优化因子及新生成因子叠加后得到复合因子。与原始因子相比，复合因子展现出显著优势。月频 RankIC 均值从 9.01%提升至 10%，ICIR 从 0.93 提升至 1.01。

图 17：原始因子与复合因子在中证 800 中累计 RankIC



资料来源：万得、信达证券研发中心
数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

表 18：原始因子与复合因子在中证 800 中 IC 统计量

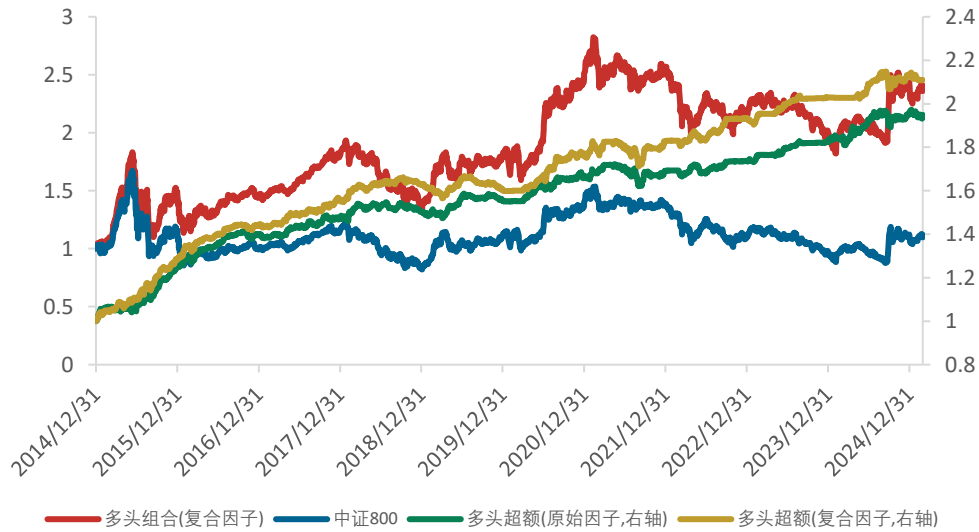
	RankIC 均值		ICIR	
	原始因子	复合因子	原始因子	复合因子
2014	9.78%	11.11%	2.37	3.70
2015	15.17%	15.05%	1.62	2.03
2016	14.89%	17.99%	1.33	1.46
2017	9.23%	11.17%	1.37	1.70
2018	4.81%	6.91%	0.63	0.98
2019	5.80%	6.34%	0.55	0.66
2020	5.10%	6.37%	0.47	0.63
2021	6.51%	6.95%	0.79	0.83
2022	8.17%	9.95%	0.97	1.11
2023	10.16%	11.05%	1.12	1.01
2024	9.91%	9.00%	0.80	0.66
2025	7.74%	3.20%	0.41	0.18
全区间	9.01%	10.00%	0.93	1.01

资料来源：万得、信达证券研发中心
数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

当落实到中证 800 指增组合层面，复合因子的优势进一步凸显。复合因子多头超额年化收益达到 7.92%，相较于原始因子组合 7.05%的年化超额收益，实现了显著提升。同时，复合因子组合年化信息比为 1.89，相较于原始组合 1.63 的年化信息比也有所增强。这表明复合因子不仅能够获取更高的超额收益，而且在风险调整后收益

的稳定性方面表现更佳。

图 18：原始因子与复合因子多头超额净值



资料来源：万得、信达证券研发中心

数据日期：2014 年 12 月 31 日至 2025 年 2 月 28 日

表 19：原始因子多头组合收益波动绩效指标

年份	年化 收益	年化 波动	最大 回撤	年化 信息比	收益 波动比	收益 回撤比
2015	43.86%	39.74%	-41.35%	3.73	1.10	1.06
2016	-2.72%	23.90%	-24.13%	3.07	-0.11	-0.11
2017	21.48%	9.15%	-4.99%	1.56	2.35	4.30
2018	-25.98%	22.01%	-31.57%	0.54	-1.18	-0.82
2019	38.47%	19.55%	-12.92%	0.97	1.97	2.98
2020	35.08%	22.61%	-15.51%	1.89	1.55	2.26
2021	0.72%	17.51%	-15.02%	0.29	0.04	0.05
2022	-18.95%	19.58%	-24.95%	1.05	-0.97	-0.76
2023	-5.85%	12.79%	-17.29%	2.74	-0.46	-0.34
2024	21.41%	20.42%	-11.29%	1.51	1.05	1.90
2025	-0.91%	16.16%	-5.62%	-0.86	-0.06	-0.16
全区间	8.27%	22.04%	-41.35%	1.63	0.38	0.20

资料来源：万得、信达证券研发中心

数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

表 20：原始因子组合多头超额收益波动绩效指标

年份	年化 收益	年化 波动	最大 回撤	收益 波动比	收益 回撤比
2015	24.99%	6.24%	-2.45%	4.01	10.21
2016	12.00%	3.84%	-2.50%	3.13	4.79
2017	5.32%	3.46%	-1.46%	1.54	3.64
2018	1.97%	3.89%	-2.81%	0.51	0.70
2019	3.38%	3.61%	-2.69%	0.94	1.26
2020	7.21%	3.85%	-1.85%	1.87	3.90
2021	1.41%	5.61%	-5.85%	0.25	0.24
2022	2.85%	2.83%	-2.49%	1.01	1.15
2023	5.00%	1.86%	-0.87%	2.69	5.71
2024	7.56%	5.11%	-3.89%	1.48	1.94
2025	-0.71%	5.64%	-2.10%	-0.13	-0.34
全区间	7.05%	4.25%	-5.85%	1.66	1.21

资料来源：万得、信达证券研发中心

数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

表 21：复合因子多头组合收益波动绩效指标

年份	年化 收益	年化 波动	最大 回撤	年化 信息比	收益 波动比	收益 回撤比
2015	47.92%	39.53%	-40.15%	3.98	1.21	1.19
2016	-2.19%	23.63%	-23.17%	3.03	-0.09	-0.09
2017	24.59%	9.38%	-4.96%	2.39	2.62	4.96
2018	-23.97%	20.60%	-29.54%	1.30	-1.16	-0.81
2019	32.00%	19.32%	-12.58%	-0.48	1.66	2.54

表 22：复合因子组合多头超额收益波动绩效指标

年份	年化 收益	年化 波动	最大 回撤	收益 波动比	收益 回撤比
2015	28.39%	6.54%	-3.02%	4.34	9.41
2016	12.52%	4.05%	-2.23%	3.09	5.62
2017	8.03%	3.36%	-1.06%	2.39	7.57
2018	4.45%	3.51%	-1.96%	1.27	2.26
2019	-1.48%	3.11%	-4.04%	-0.48	-0.37

2020	39.04%	22.55%	-15.51%	2.53	1.73	2.52
2021	2.20%	18.11%	-16.32%	0.59	0.12	0.14
2022	-16.49%	19.53%	-23.41%	2.02	-0.84	-0.70
2023	-5.88%	12.77%	-16.91%	2.59	-0.46	-0.35
2024	18.61%	21.32%	-10.80%	1.49	0.87	1.72
2025	-1.58%	17.27%	-6.00%	-2.28	-0.09	-0.26
全区间	9.16%	21.96%	-40.15%	1.89	0.42	0.23

资料来源：万得、信达证券研发中心

数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

2020	10.34%	4.06%	-1.79%	2.54	5.77
2021	3.02%	5.49%	-6.34%	0.55	0.48
2022	5.95%	2.99%	-2.34%	1.99	2.54
2023	4.96%	1.96%	-1.60%	2.54	3.11
2024	5.33%	3.68%	-3.80%	1.45	1.40
2025	-1.35%	4.13%	-1.66%	-0.33	-0.81
全区间	7.92%	4.09%	-6.34%	1.94	1.25

资料来源：万得、信达证券研发中心

数据日期：2014 年 3 月 3 日至 2025 年 2 月 28 日

风险因素

结论基于历史数据，在市场环境转变时模型存在失效的风险。

分析师声明

负责本报告全部或部分内容的每一位分析师在此申明，本人具有证券投资咨询执业资格，并在中国证券业协会注册登记为证券分析师，以勤勉的职业态度，独立、客观地出具本报告；本报告所表述的所有观点准确反映了分析师本人的研究观点；本人薪酬的任何组成部分不曾与，不与，也将不会与本报告中的具体分析意见或观点直接或间接相关。

免责声明

信达证券股份有限公司(以下简称“信达证券”)具有中国证监会批复的证券投资咨询业务资格。本报告由信达证券制作并发布。

本报告是针对与信达证券签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本报告仅提供给上述特定客户，并不面向公众发布。信达证券不会因接收人收到本报告而视其为本公司的当然客户。客户应当认识到有关本报告的电话、短信、邮件提示仅为研究观点的简要沟通，对本报告的参考使用须以本报告的完整版本为准。

本报告是基于信达证券认为可靠的已公开信息编制，但信达证券不保证所载信息的准确性和完整性。本报告所载的意见、评估及预测仅为本报告最初出具日的观点和判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会出现不同程度的波动，涉及证券或投资标的的历史表现不应作为日后表现的保证。在不同时期，或因使用不同假设和标准，采用不同观点和分析方法，致使信达证券发出与本报告所载意见、评估及预测不一致的研究报告，对此信达证券可不发出特别通知。

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、意见及推测仅供参考，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人做出邀请。

在法律允许的情况下，信达证券或其关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能会为这些公司正在提供或争取提供投资银行业务服务。

本报告版权仅为信达证券所有。未经信达证券书面同意，任何机构和个人不得以任何形式翻版、复制、发布、转发或引用本报告的任何部分。若信达证券以外的机构向其客户发放本报告，则由该机构独自为此发送行为负责，信达证券对此等行为不承担任何责任。本报告同时不构成信达证券向发送本报告的机构之客户提供的投资建议。

如未经信达证券授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。信达证券将保留随时追究其法律责任的权利。

评级说明

投资建议的比较标准	股票投资评级	行业投资评级
本报告采用的基准指数：沪深300指数（以下简称基准）； 时间段：报告发布之日起6个月内。	买入： 股价相对强于基准15%以上；	看好： 行业指数超越基准；
	增持： 股价相对强于基准5%~15%；	中性： 行业指数与基准基本持平；
	持有： 股价相对基准波动在±5%之间；	看淡： 行业指数弱于基准。
	卖出： 股价相对弱于基准5%以下。	

风险提示

证券市场是一个风险无时不在的市场。投资者在进行证券交易时存在赢利的可能，也存在亏损的风险。建议投资者应当充分深入地了解证券市场蕴含的各项风险并谨慎行事。

本报告中所述证券不一定能在所有的国家和地区向所有类型的投资者销售，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专业顾问的意见。在任何情况下，信达证券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。